Summer 2016

Analysis off Dependent Discrete Choices Using Gaussian Copula

Arjun Poddar
Old Dominion University, vickyarjun33@gmail.com

ANALYSIS OF DEPENDENT DISCRETE CHOICES USING GAUSSIAN COPULA

by

Arjun Poddar
B.Sc. July 2007, University of Calcutta, India
M.Sc. May 2009, Indian Institute of Technology, Kharagpur, India
M.S. December 2014, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTATIONAL APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
August 2016

Approved by:

N. Rao Chaganty (Director)

Norou Diawara (Member)

Kayoung Park (Member)

Rajesh Paleti (Member)

ABSTRACT

ANALYSIS OF DEPENDENT DISCRETE CHOICES USING GAUSSIAN COPULA

Arjun Poddar
Old Dominion University, 2016
Director: Dr. N. Rao Chaganty

A popular tool for analyzing product choices of consumers is the well-known conditional logit discrete choice model. Originally publicized by McFadden (1974), this model assumes that the random components of the underlying latent utility functions of the consumers follow independent Gumbel distributions. However, in practice the independence assumption may be violated and a more reasonable model should account for the dependence of the utilities. In this dissertation we use the Gaussian copula with compound symmetric and autoregressive of order one correlation matrices to construct a general multivariate model for the joint distribution of the utilities. The induced correlations on the utilities and the choice probabilities are studied using analytic expressions and simulations. For regression with consumer and product specific covariates, we derive expressions for the likelihood function and the score functions. We use numerical methods and computer code to obtain the maximum likelihood estimates of the regression and correlation parameters. The standard errors of the estimates were obtained using bootstrap. Comparison of our model with other competing methods and practical applicability is illustrated using both real world consumer preference and simulated data.

Dedicated to my parents *Pradip Poddar* and *Shyamali Chandra*,
and to the memories of my grandmother *Kalpana Poddar*.

ACKNOWLEDGEMENTS

I sincerely thank my advisor Dr. N. Rao Chaganty, Professor in the Department of Mathematics and Statistics for his guidance, valuable suggestions and support throughout the course of this dissertation. He has been instrumental in teaching different important aspects of research to me and I am grateful to him for encouraging me every step of the way.

My heartfelt thanks and appreciation go to the members of my dissertation committee- Dr. Norou Diawara, Dr. Kayoung Park and Dr. Rajesh Paleti, who have been very gracious in helping me write and edit this dissertation.

I would like to thank the faculty and the staff at the Department of Mathematics and Statistics for providing such an effective, serene and productive environment for students to grow and excel in their careers. I acknowledge the Chesapeake Bay Program at Old Dominion University and the Leroy T. Canoles Jr. Cancer Research Center at Eastern Virginia Medical School in Norfolk, Virginia for arranging the financial assistantships during my graduate studies at this university.

Words do not do justice to my gratitude for my friend *Eeshita Manna* for her relentless encouragement, overwhelming support, and unwavering confidence in me, so I shall keep it short. Thank you.

This dissertation, much like anything else in my life, would not have been possible without the love and care of, and most importantly the sacrifices made by my family members. My two sisters, *Satabdi* and *Samapti*, have been persistent in believing in me and motivating me. I share my joy and love with them. I also want to thank my niece *Taneesha* and my nephews *Taran*, *Rooney*, and *Rivan* for bringing so much happiness and positivity to the world, and for keeping me sane. Lastly, I thank my parents from the bottom of my heart and send my eternal gratitude and love their way. Their affection, love and sacrifices are the only reasons I could ever do anything my life. This dissertation belongs to them as much as it belong to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Discrete choice models are statistical models that are used when an individual has to make a choice from a list of available (discrete) options. These models are based on the fundamental idea that an option is chosen only if the person choosing it reckons that it has the highest value amongst all available options. Examples may include buying a car, selecting a school for one's child, deciding on range of expenditure on weekly groceries etc..

1.1 CHOICE SET

In the context of a discrete choice model, the first task is to define a choice set. A choice set is the collection of alternatives that is presented to all the consumers in a particular situation. There are three properties that a choice set should adhere to. First, the elements (choices/alternatives) in the choice set presented to each consumer in the study should be mutually exclusive. This means that choosing one option automatically eliminates all the other options from being chosen by the consumer. This assumption is necessary to ensure that every consumer chooses one and only one alternative. Second, a choice set should be exhaustive which means that all possible choices/alternatives are included. This allows any consumer in the study to choose one option. Third, a choice set should be countably finite, in that if someone starts to count the number of choices in the set, he/she can actually finish the counting process.

We assume that there are n consumers and each of them face c choices. Throughout this dissertation the expressions "consumer", "customer", "decision maker" are synonymous, as do "choice", "alternative", "option" and "product".

The response variables Y_{ij}'s are indicator variables taking the value one if the ith consumer chooses the jth product. That is,

$$Y_{ij} = \begin{cases} 1 & \text{if } i\text{th consumer chooses } j\text{th alternative} \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, c$. In Table 1, we display all the choice variables Y_{ij}'s for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, c$.

Table 1: Layout of the responses

Consumer	Choice Alternatives					
	1	2	$\ldots$	j	$\ldots$	c
1	Y_{11}	Y_{12}	$\ldots$	Y_{1j}	$\ldots$	Y_{1c}
2	Y_{21}	Y_{22}	$\ldots$	Y_{2j}	$\ldots$	Y_{2c}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	Y_{i1}	Y_{i2}	$\ldots$	Y_{ij}	$\ldots$	Y_{ic}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	Y_{n1}	Y_{n2}	$\ldots$	Y_{nj}	$\ldots$	Y_{nc}

Each row in the table is associated with one consumer. The values in the table are either a 1 or a 0. The property of mutual exclusiveness of the choice set dictates that only one value in a row can be 1 and the rest should be 0's.

1.2 RANDOM UTILITY MODEL

In the random utility model we assume that there is a random variable U_{ij} which reflects the utility for the ith consumer and the jth alternative. Therefore, the ith consumer has a set of utility values $\{U_{i1}, U_{i2}, \ldots, U_{ic}\}$ for the c elements in the choice set. The underlying assumption imposed on the consumers in discrete choice model is based on utility maximization. The term "utility" carries the same connotation here as it does in any other parlance. It signifies the usefulness a choice (product/option) carries to a consumer. This idea was first introduced by Thurstone (1927) in the context of psychometrics. He described that the effect of a stimuli can lead to different judgements in different subjects and the difference can be measured. Based on this, Marschak (1960) first introduced the random utility model interpreting stimuli as utility of a choice. The model states that a consumer will choose the alternative which has highest utility in his or her mind. That is,

$$Y_{ij} = \begin{cases} 1 & \text{if} \quad U_{ik} < U_{ij}, \text{ for } k = 1, 2, \ldots, c, k \neq j, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

For example, when we walk into a store to buy a shirt we might be presented with a choice set of several shirts. We will attach, in our minds, an utility value to each shirt and buy the one which has the highest utility for us.

For most of the random utility models, the utility is linearly broken down into a deterministic and a probabilistic component. The first one, similar to any other statistical methodology, shall be referred to as the mean of the utility and the second component shall be called the error or the random part. We write

$$U_{ij} = \mu_{ij} + Z_{ij}. \tag{3}$$

As we shall see shortly that μ_{ij}, the mean, is the quantity that brings the covariates, the variables or factors or features that influence the consumer's choice, into the analysis. On the other hand, different assumptions on the random component (Z_{ij}) lead to different discrete choice models. The mean can be thought of as the part of utility which can be explained in terms of the covariates and the random component is the unexplained part.

1.3 CHOICE PROBABILITY

Choice probability is the probability with which an alternative in the choice set can be chosen by a customer. We denote it by P_{ij}, the probability that the ith consumer chooses the jth alternative. Therefore,

$$\begin{aligned} P_{ij} &= \Pr(Y_{ij} = 1) \\ &= \Pr(U_{ik} < U_{ij}, \; \forall \; k \neq j) \\ &= \Pr(U_{ik} - U_{ij} < 0; \; \forall \; k \neq j) \end{aligned} \tag{4}$$

The value of Y_{ij} depends on whether U_{ij} is the maximum among $U_{i1}, U_{i2}, \ldots, U_{ic}$. It does not depend on the amount by which U_{ij} exceeds the rest of the utilities for the ith consumer. On the other hand, the choice probability, P_{ij}, depends on the margins or the differences between U_{ij} and the other utilities. In the next section, we shall formalize this property of the choice probability. Note that (4) can be written as

$$\begin{aligned} P_{ij} &= \Pr(\mu_{ik} + Z_{ik} - \mu_{ij} - Z_{ij} < 0;\ \forall\ k \neq j) \\ &= \Pr(Z_{ik} < (\mu_{ij} - \mu_{ik}) + Z_{ij};\ \forall\ k \neq j). \end{aligned} \quad (5)$$

1.4 PROPERTIES

Similar to any other probability measure, choice probabilities should be real numbers between 0 and 1, and they should sum to 1 for any consumer over all the choices. In mathematical terms, $0 \leq P_{ij} \leq 1$ and $\sum_{j=1}^{c} P_{ij} = 1,$ for all $i = 1, 2, \ldots, n$. The second property follows from the assumption that the choice set is exhaustive.

1.5 COVARIATES

Recall that the customer/consumer picks an option from the choice set based on the comparative utility of the option with respect to other options. It is to be noted that the utilities (U_{ij}'s) are latent variable and unobserved. They are the sum of two parts - one deterministic and the other probabilistic. The probabilistic part is the random component, Z_{ij} and it is normally specified by a probability distribution. The deterministic part is the mean of utility (μ_{ij}) as shown in (3). It is deterministic because it can be measured by observing other variables associated with the customer and/or the choice. These other variables are known as covariates in the statistical literature. We will assume the deterministic part is a linear function of the covariates.

1.5.1 CHOICE SPECIFIC COVARIATES

In many discrete choice scenarios, data is available on different covariates based on each customer and each choice. For example, when choosing a mode of transportation, each passenger may have different values for some covariates such as total time and total money required, traveling-group size etc. In case of buying a real estate property from a choice set of various types of properties, covariates such as price of the property, population density of the area will be different for different customers and different properties. These type of covariates are choice specific covariates which are different not only for different choices but also for different customers.

Table 2: Choice Specific Covariates

			Covariates			
Consumer	Alternative	Choice	1	2	...	p
1	1	Y_{11}	X_{111}	X_{112}	...	X_{11p}
	2	Y_{12}	X_{121}	X_{122}	...	X_{12p}
	⋮	⋮	⋮	⋮	⋮	⋮
	c	Y_{1c}	X_{1c1}	X_{1c2}	...	X_{1cp}
2	1	Y_{21}	X_{211}	X_{212}	...	X_{21p}
	2	Y_{22}	X_{221}	X_{222}	...	X_{22p}
	⋮	⋮	⋮	⋮	⋮	⋮
	c	Y_{2c}	X_{2c1}	X_{2c2}	...	X_{2cp}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	1	Y_{n1}	X_{n11}	X_{n12}	...	X_{n1p}
	2	Y_{n2}	X_{n21}	X_{n22}	...	X_{n2p}
	⋮	⋮	⋮	⋮	⋮	⋮
	c	Y_{nc}	X_{nc1}	X_{nc2}	...	X_{ncp}

In Table 2, we showcase p choice specific covariates for each customer and each choice. The mean μ_{ij} is modeled as a linear function of the covariates and a regression parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$, that is,

$$\mu_{ij} = \boldsymbol{X}'_{ij}\boldsymbol{\beta} = \sum_{m=1}^{p} X_{ijm}\beta_m, \tag{6}$$

where $\boldsymbol{X}'_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijp})$, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, c$. Based on this, the choice probability (5) involving the choice specific covariates can be written as

$$P_{ij} \quad = \quad \Pr(Z_{ik} < (\boldsymbol{X}_{ij} - \boldsymbol{X}_{ik})'\boldsymbol{\beta} + Z_{ij};\ \forall\ k \neq j). \tag{7}$$

1.5.2 INDIVIDUAL SPECIFIC COVARIATES

Individual specific covariates are those covariates that are unique to the customer/consumer and they do not change from one choice to another. In the discrete choice scenario of which insurance plan to buy, a person's age, medical history of a disease, income are fixed- they do not change from one insurance plan to another. Similarly, when deciding to which school to send their child to, covariates such as parents' income, number of children are important covariates and convey information about the subjects of the study and not the choices or options. These are some examples of individual specific covariates. For the individual specific covariates we have $\boldsymbol{X}_{ij} = \boldsymbol{X}_i$ and thus the mean value of utility is

$$\mu_{ij} = \boldsymbol{X}_i' \boldsymbol{\beta}_j, \text{ where } \boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp})', \tag{8}$$

and the choice probability is

$$P_{ij} = \Pr(Z_{ik} < \boldsymbol{X}_i' (\boldsymbol{\beta}_j - \boldsymbol{\beta}_k) + Z_{ij};\ \forall\ k \neq j). \tag{9}$$

In this dissertation we will not consider the case of individual specific covariates and confine our study only to the choice specific covariates.

1.5.3 TRANSLATION INVARIANCE OF CHOICE PROBABILITY

Suppose $U_{ij}^{'}$ is the translated utility, that is, $U_{ij}^{'} = U_{ij} + \alpha$ for all i and j, where α is a constant real number. If $P_{ij}^{'}$ denotes the choice probability of the ith consumer and the jth product based on $U_{ij}^{'}$, then using (4) we get

$$\begin{aligned} P_{ij}^{'} &= \Pr(U_{ik}^{'} - U_{ij}^{'} < 0;\ \forall\ k \neq j) \\ &= \Pr(U_{ik} - U_{ij} < 0;\ \forall\ k \neq j) \\ &= P_{ij}, \end{aligned}$$

which shows that the choice probabilities are translation invariant, that is they remain the same if we change the utility values by adding a constant to all of them. Looking at (7) and (9), and comparing them with (4) it is to be noted that only those regression coefficients can be estimated which are captured by the differences in the utilities.

1.5.4 SCALE OF UTILITY

As the utility in a discrete choice model is directly related to the random components, the assumptions imposed on the random components influence the estimation of the regression parameters. If one compares two models with different variances for the random components, the results might be misleading if the utilities are not normalized. This in turn would result in faulty comparison of the regression coefficients. It is advisable to normalize the regression coefficients according to the variances of the error components so that they are comparable.

1.6 GOODNESS OF FIT MEASURES

Most of the goodness of fit measures used to judge and compare the performances of discrete choice models are based on the log-likelihood function. If $\boldsymbol{\theta}$ is the vector of parameters (accounting for the covariates and correlations), then the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}) \quad = \quad \log\left(\prod_{i=1}^{n}\prod_{j=1}^{c} P_{ij}^{Y_{ij}}\right) = \sum_{i=1}^{n}\sum_{j=1}^{c} Y_{ij}\log(P_{ij}). \tag{10}$$

since $P_{ij} = P(Y_{ij} = 1)$. If $\widehat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$, then the estimated log-likelihood value is $\ell(\widehat{\boldsymbol{\theta}})$. One measure of goodness of fit is the Akaike information criterion, known as AIC. It is calculated as AIC $= 2\,\kappa - 2\ell(\widehat{\boldsymbol{\theta}})$, where κ is the dimension of the parameter vector $\boldsymbol{\theta}$. This measure penalizes a model for its greater number of parameters. Model with smaller value of AIC is the best according to this criteria. See Akaike (1973) for more information on AIC.

McFadden (1974) introduced a goodness of fit measure which is very similar to the coefficient of determination (R^2) in regression. It is known as McFadden's R^2

and is defined as

$$R_M^2 = 1 - \frac{\ell(\widehat{\boldsymbol{\theta}})}{\ell_0(\widehat{\boldsymbol{\theta}})},$$

where $\ell_0(\widehat{\boldsymbol{\theta}})$ is the log-likelihood value of the intercepts-only model and is treated as the total sum of squares in regression. R_M^2 lies between 0 and 1. High value of R_M^2 is desirable. A good choice model would have its R_M^2 very close to 1 as opposed to a bad model for which R_M^2 will be close to 0. Similar to R_m^2, McFadden's adjusted R^2 is defined as

$$R_{M,Adj}^2 = 1 - \frac{\ell(\widehat{\boldsymbol{\theta}}) - \kappa}{\ell(\widehat{\boldsymbol{\theta}}_0)},$$

κ being the number of parameters in the model. This measure puts a penalty for the number of parameters in the model. There are several other measures of goodness of fit, and most of them mimic the coefficient of determination, R^2. All such measures are called pseudo R^2's.

1.7 POPULAR DISCRETE CHOICE MODELS

In this section, we discuss some of the commonly used discrete choice models. We will, for the most part, describe the choice probabilities in terms of the means μ_{ij}'s of the utility functions.

1.7.1 CONDITIONAL LOGIT MODEL

The most popular discrete choice model is the conditional logit model and is ubiquitous in the literature and in practice. In this model, it is assumed that the choices do not depend on one another and hence the errors in the utility function are independent and follow identical Gumbel distributions. Luce (1959) first introduced the model by defining a theory which is now well-known as Luce's choice axiom and it says that the choice probability of one item compared to another one from a set of multiple items is unaltered by the presence of other items in the set. From this assumption, he laid out the foundation for the choice probability of a logit model as the relative weight of an item. Marschak (1960) showed that a choice model that follows Luce's choice axiom is consistent with random utility maximization. Luce and Suppes (1965) proved that if the random component of the utility function follows an extreme value distribution then the choice probability leads to a logit formula. McFadden (1974) finally completed the proof by illustrating the choice probability is given by the logit formula if and only if the underlying distribution of the error component is Gumbel.

Suppose that the random components Z_{ij}'s follow a Gumbel distribution. The density function of Gumbel is given by

$$f(z_{ij}) = e^{-z_{ij}} e^{-e^{-z_{ij}}},\ -\infty < z_{ij} < \infty,$$

and the cumulative distribution is

$$F(z_{ij}) = e^{-e^{-z_{ij}}},\ -\infty < z_{ij} < \infty.$$

Also assume that for any given i, Z_{ij} is independent of Z_{ik}, for $k \neq j$, that is $\mathrm{corr}(Z_{ij}, Z_{ik}) = 0$.

Using the assumptions of the Z_{ij}'s listed above, we derive the choice probability according to McFadden (1974) as

$$\begin{aligned} P_{ij} &= \Pr(Z_{ik} < (\mu_{ij} - \mu_{ik}) + Z_{ij};\ \forall\, k \neq j) \\ &= \int_{-\infty}^{\infty} \Pr\left(Z_{ik} < (\mu_{ij} - \mu_{ik}) + z | Z_{ik} = z;\ \forall\, k \neq j\right)\ f(z)\, dz \end{aligned}$$

Now, we shall use the fact that Z_{ik} and Z_{ij} are independent $\forall\, k \neq j$. So,

$$\begin{aligned} P_{ij} &= \int_{-\infty}^{\infty} \left(\prod_{k(\neq j)=1}^{c} \exp\bigl(-e^{-(z+\mu_{ij}-\mu_{ik})}\bigr) \right) e^{-z}\ \exp(-e^{-z})\, dz \\ &= \int_{-\infty}^{\infty} \exp\left(-e^{-z} \sum_{k=1}^{c} e^{-(\mu_{ij}-\mu_{ik})} \right) e^{-z}\, dz. \end{aligned}$$

Making a change of variable $t = e^{-z}$ in the integrand, the choice probability P_{ij}

becomes

$$\begin{aligned} P_{ij} &= \int_0^\infty \exp\left(-t\sum_{k=1}^{c} e^{-(\mu_{ij}-\mu_{ik})}\right)\, dt \\ &= \frac{1}{\sum_{k=1}^{c} e^{-(\mu_{ij}-\mu_{ik})}} \\ &= \frac{e^{\mu_{ij}}}{\sum_{k=1}^{c} e^{\mu_{ik}}}. \end{aligned} \tag{11}$$

This shows that the choice probability for the conditional logit model is in a closed form, and is very easy to calculate and does not require evaluation of integrals. These facts account for the widespread use of the model and implementation in various statistical software.

Independence from Irrelevant Alternatives (IIA).

Using the formula (11) we can see that for any two choices j and k in the choice set,

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{\mu_{ij}}}{e^{\mu_{ik}}} = \frac{e^{\boldsymbol{X}'_{ij}\boldsymbol{\beta}}}{e^{\boldsymbol{X}'_{ik}\boldsymbol{\beta}}} = e^{(\boldsymbol{X}_{ij}-\boldsymbol{X}_{ik})'\boldsymbol{\beta}}.$$

This shows that for the ith consumer, the ratio of the probabilities of choosing the jth and the kth choices depends only on the covariates and the coefficients for those two choices only, for all $j \neq k$. That is, even if we change the information on the other choices, these two choice probabilities will change proportionately. As the ratio of any two choice probabilities is independent of all other alternatives, this property is called independence from irrelevant alternatives (IIA).

1.7.2 NESTED LOGIT MODELS

Nested logit models are relevant when the choice set can be partitioned into subsets and the random components of the utilities are Gumbel random variables. These subsets are called nests. For example, parents' choices for their children's schools can be grouped into two nests, namely private school and public school. In the case of individuals purchasing health insurance plans, the choice set can be partitioned into four groups such as vision and dental care included, only vision care included, only dental care included, and both vision and dental care excluded.

The rudimentary idea of nested logit models is that the relative choice probabilities among the choices in one nest always remain fixed while for choices between any two different nests the relative choice probabilities are different.

McFadden (1978) developed the nested logit model. Let us denote the number of nests by N and the nests (disjoint sets) as $S_1, S_2, \ldots, S_N$. If, c_k is the number of choices in nest S_k, then $\sum_{k=1}^{N} c_k = c$. Within nest N_k, the choices are correlated and it is assumed that the CDF of the random components of the utility have the joint CDF

$$F(z_{i1}, z_{i2}, \ldots, z_{ic}) = \exp\left(-\sum_{k=1}^{N}\left(\sum_{j=1}^{c_k} e^{-z_{kj}/\lambda_k}\right)^{\lambda_k}\right).$$

The quantity λ_k is such that $1-\lambda_k$ can be treated as a measure of dependence within S_k and its values are between 0 and 1. The extreme $\lambda_k = 0$ indicates complete dependence among the choices of the nest S_k and when $\lambda_k = 1$, the choices in S_k are independent of each other.

In the special case when $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 0$, the nested logit model transforms to a conditional logit model. The choice probability that the ith consumer chooses the jth product, assuming that it belongs to the nest S_k is

$$P_{ij} = \frac{\exp(\mu_{ij}/\lambda_k)\left\{\sum_{l=1}^{c_k}\exp(\mu_{il}/\lambda_k)\right\}^{\lambda_k-1}}{\sum_{k=1}^{N}\left\{\sum_{l=1}^{c_k}\exp(\mu_{il}/\lambda_k)\right\}^{\lambda_k}},$$

and if the j'th product is in the nest k' then the ratio of the choice probabilities for choices j and j' is

$$\frac{P_{ij}}{P_{ij'}} = \frac{\exp(\mu_{ij}/\lambda_k)\left\{\sum_{l=1}^{c_k}\exp(\mu_{il}/\lambda_k)\right\}^{\lambda_k-1}}{\exp(\mu_{ij'}/\lambda_{k'})\left\{\sum_{l=1}^{c'_k}\exp(\mu_{il}/\lambda_{k'})\right\}^{\lambda_{k'}-1}}.$$

When the two choices j and j' come from the same nest, i.e. $k = k'$,

$$\frac{P_{ij}}{P_{ij'}} = \frac{\exp(\mu_{ij}/\lambda_k)}{\exp(\mu_{ij'}/\lambda_k)}.$$

This illustrates that IIA holds only within each nest. Another interesting observation is that when $k \neq k'$, the ratio of P_{ij} to $P_{ij'}$ is dependent on the covariates of other choices besides the jth and j'th choice, all of those choices are either in nest k or nest k'. This implies that in the nested logit model, relative odds of choosing two choices from two different nests only depend on covariates of choices in those two nests only. This property is referred to as "independence from irrelevant nests", abbreviated as IIN.

1.7.3 PAIRED COMBINATORIAL LOGIT

As the name suggests, the paired combinatorial logit model assumes that each pair of choices constitute a nest and the random components of the utilities are Gumbel variables. For a discrete choice setup with c choices, this model assumes there are $c(c-1)/2$ nests where each choice is represented in $c-1$ nests. Unlike the nested logit model, in paired combinatorial logit, the nests are intersecting. To measure the independence within the nest formed by choices j and k, a quantity λ_{jk} is introduced. The degree of association between choices j and k is given by $1-\lambda_k$. In the case where all λ_{jk}'s are equal to unity, this model reduces to a conditional logit model. The choice probability for this model is of the form

$$P_{ij} = \frac{\sum_{j \neq k} \exp(\mu_{ij}/\lambda_{jk}) \left\{\exp(\mu_{ij}/\lambda_{jk}) + \exp(\mu_{ik}/\lambda_{jk})\right\}^{\lambda_{jk}-1}}{\sum_{l=1}^{c-1} \sum_{l'=l+1}^{c} \left\{\exp(\mu_{il}/\lambda_{ll'}) + \exp(\mu_{il'}/\lambda_{ll'})\right\}^{\lambda_{ll'}}}$$

1.7.4 GENERALIZED NESTED LOGIT

In generalized nested logit model, the choices are grouped into N overlapping nests $S_1, S_2, \ldots, S_N$ and each choice can belong to more than one nest with varying degrees of presence. Simply put, a choice appearing in multiple nests can be more prominent in one nest than others. If c_k is the number of choices in nest S_k then $\sum_{k=1}^{N} c_k \geq c$. An allocation parameter named α_{jk} is included in this model which represents the degree of presence of the jth choice in S_k, $j = 1, 2, \ldots, c$ and $k = 1, 2, \ldots, N$. It is assumed that $\alpha_{jk} \geq 0$ and $\sum_{k=1}^{N} \alpha_{jk} = 1$. Under these assumptions, α_{jk} represents the relative presence of the jth choice in the kth nest as compared to other $N-1$ nests. The probability that the ith consumer selects the jth choice for this model is

$$P_{ij} = \frac{\sum_{k=1}^{G}(\alpha_{jk}\exp(\mu_{ij}))^{1/\lambda_k}\left\{\sum_{l=1}^{c_k}(\alpha_{lk}\exp(\mu_{il}))^{1/\lambda_k}\right\}^{\lambda_k-1}}{\sum_{k=1}^{G}\left\{\sum_{l=1}^{c_k}(\alpha_{lk}\exp(\mu_{il}))^{1/\lambda_k}\right\}^{\lambda_k}}.$$

If all the allocation parameters have degenerate distributions with $\alpha_{jk} = 1$ for exactly one k in $1, 2, \ldots, N$, then this model reduces to the nested logit model. Additionally, if choices in each nest have zero dependency with one another, i.e., $\lambda_k = 1$ for all $k = 1, 2, \ldots, N$ then the generalized nested logit model reduces to the conditional logit model.

1.7.5 GENERALIZED EXTREME VALUE (GEV) MODEL

The GEV setup provides a framework for developing discrete choice models where the choices (hence the random components in the utility functions) do not have to be independent. Based on certain mathematical criteria, this setup facilitates the derivation and computation of the choice probabilities and the dependence parameters. This was originally studied by McFadden (1978, 1981, 1984, 2001).

A GEV model is derived by assuming that the random components follow standard Gumbel distribution $f(z_{ij}) = e^{-z_{ij}}e^{-e^{-z_{ij}}}$, $-\infty < z_{ij} < \infty$ and by using a real-valued function G, defined on the c-dimensional orthant $(w_{i1}, w_{i2}, \ldots, w_{ic}) \geq 0$, where $w_{ij} = \exp(\mu_{ij})$. Furthermore, G satisfies the following four properties:

i. $G(w_1, w_2, \ldots, w_c) \geq 0$ for all $(w_1, w_2, \ldots, w_c) \geq 0$

ii. G is a homogenous function of degree 1, that is, $G(\alpha w_1, \alpha w_2, \ldots, \alpha w_c) = \alpha G(w_1, w_2, \ldots, w_c)$

iii. $\lim_{w_i \to \infty} G(w_1, w_2, \ldots, w_c) = \infty$

iv. If $(i_1, i_2, \ldots, i_k)$ is a k-tuple from $(1, 2, \ldots, c)$, then

$$\left[(-1)^k \frac{\partial^k}{\partial w_{i_1}\partial w_{i_2}\ldots\partial w_{i_k}} G \leq 0\right], \text{ for all } k = 1, 2, \ldots, c.$$

Under the above four assumptions, the utility maximizing choice probability is given by the formula

$$P_{ij} = \frac{w_{ij} G_{ij}}{G}, \tag{12}$$

where $G_{ij} = \partial G / \partial w_{ij}$. It is easy to verify that the conditional logit, nested logit, paired combinatorial logit and generalized nested logit models are special cases of the GEV model. For example, $G = \sum_{j=1}^{c} w_{ij}$ gives us the conditional logit model and $G = \sum_{j=1}^{c-1} \sum_{k=j+1}^{c} (w_{ij}^{1/\lambda_{jk}} + w_{ik}^{1/\lambda_{jk}})^{\lambda_{jk}}$ leads to the paired combinatorial logit model.

Though it is fairly easy to find functions G that satisfy the required four properties and can lead to easy derivation of new choice formulae, Train (2004) argues that this process is motivated by mathematical convenience rather than by scientific intuition.

1.7.6 MULTINOMIAL PROBIT MODEL

In the multinomial probit model, it is assumed that the random components of the utility functions are distributed as normal distributions and depending on the choice set's setup, any correlation structure can be incorporated in this model. This model was first studied by Thurstone (1927) in the case of two choices. Later, Hausman and Wise (1978), and Daganzo (1979) illustrated different properties of this model. In the utility model $U_{ij} = \mu_{ij} + Z_{ij}$, it is assumed that the Z_{ij}'s are normally distributed. In fact,

$$(Z_{i1}, Z_{i2}, \ldots, Z_{ic}) \sim N(\mathbf{0}, \mathbf{\Sigma}).$$

Then (5) for this model becomes

$$P_{ij} = \Pr(Z_{ik} - Z_{ij} < \mu_{ij} - \mu_{ik}; \forall k \neq j),$$

where each $Z_{ik} - Z_{ij}$, $k \neq j$ also follows normal distribution and their joint distribution is also multivariate normal of dimension $c - 1$. Due to this nice property and the availability of numerous simulation techniques for normal probabilities, the multivariate probit model has gained much popularity.

1.7.7 HETEROSCEDASTIC EXTREME VALUE (HEV) MODEL

The heteroscedastic extreme value model assumes that the random components are independent Gumbel variables with different scale parameters. Whereas all the other logit models relaxed the assumption of independence, the HEV model relaxes the assumption of identical distributions of the Z_{ij}'s but allows them to be independent. In mathematical terms,

$$f(z_{ij}) = \frac{1}{\theta_j} e^{-\frac{z_{ij}}{\theta_j}} e^{-e^{\frac{z_{ij}}{\theta_j}}} \text{ and } F(z_{ij}) = \frac{1}{\theta_j} e^{-e^{\frac{z_{ij}}{\theta_j}}}, \theta_j > 0 \; \forall \; j.$$

The above assumptions render the variance of Z_{ij} to be $\pi^2 \theta_j^2/6$.

Bhat (1970) calculated the choice probability for the HEV model as

$$P_{ij} = \int_{-\infty}^{\infty} \Big\{ \prod_{k=1, k\neq j}^{c} e^{-e^{-(\mu_{ij}-\mu_{ik}+\theta_j v)/\theta_j}} \Big\} e^{-v} e^{-e^{-v}} dv,$$

where $v = z_{ij}/\theta_j$.

1.7.8 OVERVIEW OF THE DISSERTATION

The rest of this dissertation is organized as follows. In Chapter 2, we introduce a new choice model. This model assumes that the random components of the utilities are distributed as Gumbel as in McFadden's original conditional logit model. However our model assumes that the random components are dependent and the joint distribution is induced by the Gaussian copula with equicorrelated correlation structure. When the correlation parameter equals zero, our model reduces to the conditional logit model and thus it is a generalization of McFadden's work. We give a brief summary of copulas with special emphasis on the Gaussian copula. We derive analytical expressions for the choice probability and study their behavior as a function of the correlation parameter. The maximum likelihood estimation procedure is discussed for estimating the correlation parameter and the regression parameter for individual specific covariates. We derive simplified expressions for the score equations and develop an R code to solve them. The standard errors for the parameter estimates are obtained using the bootstrap method. We illustrate the practical application of the model using a real life data, and compare the results with the conditional logit model.

In Chapter 3 we consider the case where there is a natural ordering in the choices or in other words, the case where the choices are categorical and ordinal. Equicorrelated correlation structure is not appropriate in that situation and we propose replacing that with an autoregressive of order one (AR(1)) correlation structure. We establish some properties of the multivariate normal distribution with AR(1) correlation structure, in particular, we show that given the present the past is independent of the future. As in Chapter 2 we derive simplified expressions for the choice probabilities and maximum likelihood estimation for the parameters in the model. We illustrate the method on a simulated data consisting of ordered choices.

Finally in the Appendix we state and prove several theorems regarding the multivariate normal distribution that are relevant and useful in this dissertation. We use the R program (R version 2.15.1) and SAS® software (version 9.3 of SAS for Windows). A selection of R code that we developed for this dissertation is also included in the Appendix.

CHAPTER 2

EQUICORRELATED CHOICE MODEL WITH GAUSSIAN COPULA

2.1 INTRODUCTION

In many discrete choice scenarios, the assumption of independent utilities is not realistic. In fact, in most cases two or more choices will be correlated. This association between the choices can be attributed to one or more covariates that affect some choices in one way and others in a different way. For example, a customer's inclination to buy products made in his country may discourage him to buy products made in other countries, which means utilities of the products made within the country will have higher correlation. In the case of choosing a route from a set of viable routes for traveling from point A to point B, a traveler might have higher preferences for routes that offer a more scenic and slow journey as compared to regular fast routes.

As we have discussed in the previous chapter, there are several models that account for dependence among the choices. Our goal is to generalize the conditional logit model proposed by McFadden (1974). Though the GEV model generalizes the conditional logit model, its assumptions are highly mathematical and lack logical intuition. In a way the GEV models work in a backwards approach in that one has to find some mathematical functions that satisfy certain properties and then the formula for choice probability is determined by using the functions. Our goal is to generalize McFadden's original model by using the same assumptions except that of independence among the elements of the choice set.

To start with, we assume that all the choices are correlated to each other with the same correlation. An appropriate structure is the equicorrelation or compound-symmetry structure. The equicorrelated correlation matrix of dimension c with parameter ρ is given by

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}_{c\times c}. \tag{13}$$

The determinant of $\boldsymbol{R}$ is $[1 + (c-1)\rho](1-\rho)^{c-1}$. For $\boldsymbol{R}$ to be positive definite, ρ must lie in $(-\frac{1}{c-1}, 1)$. Clearly, as c increases to ∞ the range of ρ converges to the interval $(0, 1)$. Though theoretically ρ can be in the negative range, for all practical purposes we shall assume that ρ is positive, that is, we assume $0 < \rho < 1$. We will explain the rationale behind this with a contradiction. Consider three random components in the case where the equicorrelation parameter ρ is negative. Then, the first and the second are negatively correlated and so are the first and the third. This implies that the first and the third should be positively correlated which is a contradiction to the assumption of negative equicorrelation.

Let us assume that there are n consumers. Each consumer is presented with an identical set of c items and has to choose exactly one item from that set. We assume that the decision of subject i to choose an item j depends on the utility U_{ij}, and the choice of consumer i does not depend on other consumers. Standard discrete choice models assume that the utility $U_{ij} = \mu_{ij} + Z_{ij}$, is the sum of a deterministic part μ_{ij} and a random component Z_{ij}. For fixed i, $Z_{i1}, Z_{i2}, \dots, Z_{ic}$ are dependent random variables since choosing from the c products are inherently related for any given consumer. In the choice problem a consumer selects item j that has the maximum utility, that is, the consumer selects item j if $U_{ij} > U_{ik}$ for all $k \neq j$. In this situation we would then be interested in computing the choice probability P_{ij} that consumer i chooses product j, which is given by

$$P_{ij} = \Pr\left(U_{ij} > U_{ik}, k = 1, 2, \dots, c, k \neq j\right). \tag{14}$$

Suppose that the random component Z_{ij} is a continuous random variable with distribution function F and probability density function f that does not depend on i

and j. Then (14) can be written as

$$\begin{aligned}
P_{ij} &= \Pr\left(U_{ij} > U_{ik}, k = 1, 2, \ldots, c, k \neq j\right) \\
&= \Pr\left(\mu_{ij} + Z_{ij} > \mu_{ik} + Z_{ik}, k = 1, 2, \ldots, c, k \neq j\right) \\
&= \Pr\left(Z_{ik} < \mu_{ij} - \mu_{ik} + Z_{ij}, k = 1, 2, \ldots, c, k \neq j\right) \\
&= \int_{-\infty}^{\infty} \Pr\left(Z_{ik} < \mu_{ij} - \mu_{ik} + z_{ij}, k = 1, 2, \ldots, c, k \neq j \mid Z_{ij} = z_{ij}\right) \cdot f(z_{ij})\, dz_{ij} \\
&= \int_{-\infty}^{\infty} \Pr\left(Z_{ik} < z^*_{ijk}, k = 1, 2, \ldots, c, k \neq j \mid Z_{ij} = z_{ij}\right) \cdot f(z_{ij})\, dz_{ij}, \qquad (15)
\end{aligned}$$

where $z^*_{ijk} = \mu_{ij} - \mu_{ik} + z_{ij}$. Thus we see the choice probability P_{ij} is a function of the conditional distribution of $(Z_{i1}, \ldots, Z_{i(j-1)}, Z_{i(j+1)}, \ldots, Z_{ic})$ given Z_{ij} which, in turn, is a function of the joint distribution of $(Z_{i1}, \ldots, Z_{i(j-1)}, Z_{ij}, Z_{i(j+1)}, \ldots, Z_{ic})$.

Following the conditional logit model by McFadden (1974), we assume that marginally the errors Z_{ij} are distributed as Gumbel random variables. So, the density for the unobserved utility Z_{ij} for the ith customer choosing the jth item is given by

$$f(z_{ij}) = e^{-z_{ij}} e^{-e^{-z_{ij}}},\ -\infty < z_{ij} < \infty, \qquad (16)$$

and the cumulative distribution is

$$F(z_{ij}) = e^{-e^{-z_{ij}}},\ -\infty < z_{ij} < \infty, \qquad (17)$$

The mean of this distribution is γ, known as Euler's constant. The approximate value of γ is 0.5772. Though the mean is non-zero, it does not affect the choice probability because as can be seen in (15), only differences in the utility appear in the expression. The variance of the Gumbel distribution is $\pi^2/6$. When comparing models with different variances we need to normalize the estimates.

Though the marginal distribution of the Z_{ij}'s have been specified, the conditional distribution of $(Z_{i1}, \ldots, Z_{i(j-1)}, Z_{i(j+1)}, \ldots, Z_{ic})$ given Z_{ij} is unknown. For each $i = 1, 2, \ldots, n$, this conditional distribution depends on the joint distribution of the Z_{ij}'s , $j = 1, 2, \ldots, c$. There are many forms that have been suggested as the joint distribution of multiple random variables the marginal distributions of which are Gumbel distributions. All of them are very complicated, involve implicit functions and not readily interpretable. For more on this, see Kotz et al. (2000). To construct a joint distribution for the errors, we will use what is known in statistics literature as the Gaussian copula. A brief description of copulas is given in the next section.

2.2 COPULAS

Copulas are functions used to describe the unknown multivariate distribution function of a set of random variables with known marginal distributions. Copulas model the interdependence between stochastic variables and thus facilitate modeling and estimation of distributions of random vectors in high dimensional statistical applications. In recent years, copulas have found their use in a variety of fields ranging from engineering to quantitative finance.

Copulas are multivariate distribution functions with uniform marginals. By the inverse transformation method we know that when the known marginal distribution functions are inverted they become uniform random variables on the interval $[0,1]$. These newly formed uniform variables are then used as arguments in a copula with a given dependence (correlation) structure to generate a joint distribution with known marginal distributions. A formal definition of a copula is as follows.

Definition 2.2.1. A c-dimensional copula is a function $C : [0,1]^c \to [0,1]$ with the following properties.

1. $C(1,\ldots,1,u_i,1,\ldots,1) = u_i$ for all $i = 1,2,\ldots,c$ and $u_i \in [0,1]$.
2. $C(u_1,u_2,\ldots,u_c) = 0$ if at least one $u_i = 0$ for $i = 1,2,\ldots,c$.
3. For all $0 < u_{i1} < u_{i2} < 1$, $i = 1,2,\ldots,c$,

$$\sum_{j_1=1}^{2}\sum_{j_2=1}^{2}\cdots\sum_{j_c=1}^{2}(-1)^{j_1+j_2+\cdots+j_c}\, C(u_{1j_1},u_{2j_2},\ldots,u_{cj_c}) \geq 0.$$

4. $C(u_1,u_2,\ldots,u_c)$ is right continuous for $u_i \in [0,1]$ for all $i = 1,2,\ldots,c$

The following famous theorem shows that underlying every multivariate distribution there is a copula that characterizes the dependence within the multivariate distribution.

Sklar's Theorem:

Let $Z_1, Z_2, \ldots, Z_c$ be c random variables with marginal cumulative distributions $F_1, F_2, \ldots, F_c$ respectively. Suppose F is their joint cumulative distribution function.

1. Then there exists a function C such that

   $$F(z_1, z_2, \ldots, z_c) = C(F_1(z_1), F_2(z_2), \ldots, F_c(z_c)),$$

   where $-\infty < z_i < \infty$.

2. If $Z_1, Z_2, \ldots, Z_c$ are continuous random variables then the copula C is unique. If Z_i is a discrete random variable then C is unique on the c-dimensional rectangle $Range(F_1) \times Range(F_2) \times \ldots Range(F_c)$.

2.2.1 EXAMPLES

Some popular and commonly used copulas are given below.

Example 1. The *independence copula* is given by the function

$$C(u_1, u_2, \ldots, u_c) = \prod_{i=1}^{c} u_i$$

Example 2. The *Gaussian copula* is given by the function

$$C(u_1, u_2, \ldots, u_c) = \Phi_c\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_c); \mathbf{0}, \boldsymbol{R}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution and $\Phi_c(\,.\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the cumulative distribution function of a c-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. It is given by

$$\Phi_c(x_1, x_2, \ldots, x_c; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{-\infty}^{x_c} \cdots \int_{-\infty}^{x_1} \frac{1}{(2\pi)^{c/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}-\boldsymbol{\mu})} \, dz_1 \ldots dz_c$$

In the definition of the Gaussian copula, the mean vector $\boldsymbol{\mu}$ is taken to be a zero vector and the covariance matrix $\boldsymbol{\Sigma}$ is assumed to be a correlation matrix $\mathbf{R}$ to ensure the parameter is identifiable.

Example 3. The *Comonotonicity Copula* is given by the function

$$C(u_1, u_2, \ldots, u_c) = \min\{u_1, u_2, \ldots, u_c\}.$$

Example 4. Let M be a univariate distribution function of a positive random variable. Note that $M(0) = 0$. For $x \geq 0$, let

$$h(x) = \int_0^\infty e^{-xz} dM(z), x \geq 0.$$

be the Laplace transform of M. The *Archimedean copula* is defined as

$$C(u_1, u_2, \ldots, u_c) = h\left(\sum_{i=1}^{c} h^{-1}(u_i)\right)$$

In this dissertation we will be dealing only with the Gaussian copula.

2.2.2 COPULA DENSITY FUNCTIONS

Suppose F_i is the marginal cumulative distribution function of Z_i, $i = 1, 2, \ldots, c$. For a copula model, the joint cumulative distribution function for the vector $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_c)$ is given by

$$F(\boldsymbol{z}) = C(F_1(z_1), F_2(z_2), \ldots, F_c(z_c)),$$

where C is a c-dimensional copula. If $\boldsymbol{Z}$ is continuous then its joint density function is

$$f(\boldsymbol{z}) = \prod_{i=1}^{c} f_i(z_i)\; c(f_1(z_1), f_2(z_2), \ldots, f_c(z_c)),$$

where $f_i(z) = \partial F_i(z)/\partial z$ is the marginal density function of Z_i and $c(u_1, u_2, \ldots, u_c)$ is the density of the copula C given by

$$c(u_1, u_2, \ldots, u_c) = \frac{\partial^c C(u_1, u_2, \ldots, u_c)}{\partial u_1 \partial u_2 \ldots \partial u_c}.$$

On the other hand, if $\boldsymbol{Z}$ is a discrete random vector then the c-dimensional joint probability mass function is

$$\Pr(\boldsymbol{Z} = \boldsymbol{z}) = \sum_{j_1=1}^{2} \sum_{j_2=1}^{2} \cdots \sum_{j_c=1}^{2} (-1)^{j_1+j_2+\ldots,+j_c}\, C(u_{1j_1}, u_{2j_2}, \ldots, u_{cj_c}),$$

where $u_{i1}(z_i) = F_i(z_i^-)$ and $u_{i2}(z_i) = F_i(z_i)$, $F_i(z_i^-)$ being the left hand limit of F_i at z_i.

2.3 DISCUSSION ON GAUSSIAN AND GUMBEL DISTRIBUTIONS

As we are planning to use the Gaussian copula in this dissertation for the joint distribution, and Gumbel distribution for the marginals, a discussion of the properties and highlights of the reasons for these choices are in order.

The normal or the Gaussian distribution is the most studied continuous distribution in statistics and its applications are countless. There is a plethora of readily comprehensible properties for both the univariate and multivariate versions of this distribution. The Gaussian copula inherits all of these properties, and so it is a natural and practical choice to model dependence between discrete choices.

Gaussian copulas are constructed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{R}$, where $\boldsymbol{R}$ is a correlation matrix. This renders the univariate components of the copula with an univariate normal distribution with mean 0 and variance 1. The random components in the utility model are marginally distributed as univariate Gumbel distribution. The support of these two distributions is $(-\infty, \infty)$, but they do differ

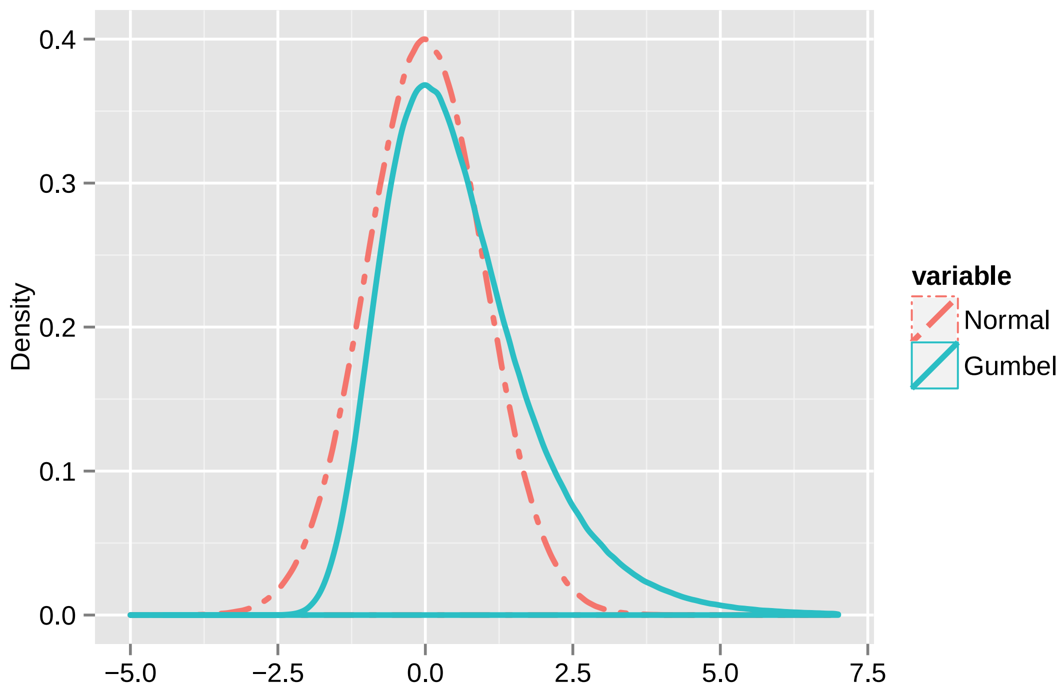

Figure 1: Plot of univariate normal and Gumbel densities

in the moments. The mean of Gumbel distribution is the Euler's constant (γ), the approximate value of which is 0.5772, and it's variance is $\pi^2/6 \approx 1.6449$. Despite this differences in mean and variances, the plot of the densities of these two distributions, as shown in Figure 1, shows that the standard normal distribution can work as a good approximation for the standard Gumbel distribution.

2.4 GAUSSIAN COPULA FOR CHOICE PROBABILITIES

Since the choice probability P_{ij} depends mainly on j and not on i, for notational simplicity we will omit the subscript i and write P_j instead of P_{ij} in further simplified analytical expressions. Thus for the ith customer the choice probability (15) can be re-written as

$$P_j = \int_{-\infty}^{\infty} \Pr\left(Z_k < z_{jk}^*, k = 1, 2, \ldots, c, k \neq j \mid Z_j = z_j\right) \cdot f(z_j)\, dz_j, \tag{18}$$

We will simplify calculation of this probability assuming that the joint cumulative distribution function of $(Z_1, Z_2, \ldots, Z_c)$ is induced by the Gaussian copula. This means,

$$\begin{aligned}
&\Pr\left(Z_1 < z_1, Z_2 < z_2, \ldots, Z_c < z_c\right) \\
&= \Phi_c\left(\Phi^{-1}\left(F(z_1)\right), \Phi^{-1}\left(F(z_2)\right), ..., \Phi^{-1}\left(F(z_c)\right); \mathbf{0}, \mathbf{R}\right), \qquad (19)
\end{aligned}$$

where $\Phi_c(\cdot\ ;\ \mathbf{0},\ \mathbf{R})$ is the cumulative distribution function of a c dimensional multivariate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}$. To make the model identifiable we take $\mathbf{R}$ to be a correlation matrix. Taking partial derivatives of (19) with respect to z_j's we get the probability density function of $(Z_1, Z_2, \ldots, Z_c)$ as

$$\begin{aligned}
&f_{Z_1,Z_2,\ldots,Z_c}(z_1, z_2, \ldots, z_c) \\
&= \frac{\phi_c\left(\Phi^{-1}(F(z_1)), \Phi^{-1}(F(z_2)), \ldots, \Phi^{-1}(F(z_c)); \mathbf{0}, \boldsymbol{R}\right)}{\prod_{k=1}^{c} \phi(\Phi^{-1}(F(z_k)))} \prod_{k=1}^{c} f(z_k), \qquad (20)
\end{aligned}$$

where ϕ_c and ϕ are the probability density functions of multivariate and univariate normal distributions respectively. The conditional pdf of $(Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_c)$ given $Z_j = z_j$ is

$$\begin{aligned}
&f_{Z_1,\ldots,Z_{j-1},Z_{j+1},\ldots,Z_c|Z_j=z_j}(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_c) \\
&= \frac{f_{Z_1,Z_2,\ldots,Z_c}(z_1, z_2, \ldots, z_c)}{f(z_j)} \\
&= \frac{\phi_c\left(\Phi^{-1}(F(z_1)), \Phi^{-1}(F(z_2)), \ldots, \Phi^{-1}(F(z_c)); \mathbf{0}, \boldsymbol{R}\right)}{\prod_{k=1}^{c} \phi(\Phi^{-1}(F(z_k)))} \prod_{k \neq j} f(z_k). \qquad (21)
\end{aligned}$$

Using (21) we can write the conditional cumulative distribution function of $(Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_c)$ given Z_j as

$$Pr(Z_k < z_{jk}^*, k = 1, 2, \ldots, c, k \neq j \mid Z_j = z_j)$$

$$= \int_{-\infty}^{z_{j1}^*} \cdots \int_{-\infty}^{z_{j(j-1)}^*} \int_{-\infty}^{z_{j(j+1)}^*} \cdots \int_{-\infty}^{z_{jc}^*}$$

$$f_{Z_1,\ldots,Z_{j-1},Z_{j+1},\ldots,Z_c|Z_j=z_j}(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_c) \prod_{k\neq j} dz_k$$

$$= \int_{-\infty}^{z_{j1}^*} \cdots \int_{-\infty}^{z_{j(j-1)}^*} \int_{-\infty}^{z_{j(j+1)}^*} \cdots \int_{-\infty}^{z_{jc}^*}$$

$$\frac{\phi_c\left(\Phi^{-1}(F(z_1)), \Phi^{-1}(F(z_2)), \ldots, \Phi^{-1}(F(z_c)); \mathbf{0}, \boldsymbol{R}\right)}{\prod_{k=1}^{c} \phi(\Phi^{-1}(F(z_k)))} \prod_{k\neq j} f(z_k) \prod_{k\neq j} dz_k. \tag{22}$$

Substituting (22) in (15), we get

$$P_j =$$

$$\int_{-\infty}^{\infty} \Bigg\{ \int_{-\infty}^{z_{j1}^*} \cdots \int_{-\infty}^{z_{j(j-1)}^*} \int_{-\infty}^{z_{j(j+1)}^*} \cdots \int_{-\infty}^{z_{jc}^*}$$

$$\frac{\phi_c\left(\Phi^{-1}(F(z_1)), \Phi^{-1}(F(z_2)), \ldots, \Phi^{-1}(F(z_c)); \mathbf{0}, \boldsymbol{R}\right)}{\prod_{k=1}^{c} \phi(\Phi^{-1}(F(z_k)))} \times \prod_{k\neq j} f(z_k) \quad \prod_{k\neq j} dz_k \Bigg\} f(z_j)\, dz_j. \tag{23}$$

Now, we will make a change of variables. Let $v_k = \Phi^{-1}(F(z_k))$ for $k = 1, \ldots, c$. Then,

$$dv_k = \frac{f(z_k)}{\phi\left(\Phi^{-1}(F(z_k))\right)}\, dz_k\,, \quad \text{for } k = 1, 2, \ldots, c$$

and for $k \neq j$,

$$\begin{aligned} v_{jk}^* &= \Phi^{-1}(F(z_{jk}^*)) \\ &= \Phi^{-1}(F(\mu_j - \mu_k + z_j)) \end{aligned}$$

$$= \quad \Phi^{-1}(F(\mu_j - \mu_k + F^{-1}(\Phi(v_j)))) \tag{24}$$

since $z_k^* = \mu_j - \mu_k + z_j$. For $k = j$, note that $v_j \to -\infty$ as $z_j \to -\infty$ and $v_j \to \infty$ as $z_j \to \infty$. Therefore, we get

$$P_j = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{v_{j1}^*} \cdots \int_{-\infty}^{v_{j(j-1)}^*} \int_{-\infty}^{v_{j(j+1)}^*} \cdots \int_{-\infty}^{v_{jc}^*} \phi_c\left(v_1, v_2, \ldots, v_c; \mathbf{0}, \boldsymbol{R}\right) \prod_{k \neq j} dv_k \right\} dv_j \tag{25}$$

where v_{jk}^* is given by (24). This can be further simplified by breaking down the multivariate density function ϕ_c. The multivariate normal density can be written as a product of a conditional density and a marginal density. That is, we can write

$$\phi_c\left(v_1, v_2, \ldots, v_c; \mathbf{0}, \boldsymbol{R}\right) = \phi_{c-1}\left(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_c; \boldsymbol{\eta}^{(j)}, \boldsymbol{\Sigma}^{(j)}\right) \phi(v_j), \tag{26}$$

where $\boldsymbol{\eta}^{(j)}$ and $\boldsymbol{\Sigma}^{(j)}$ are the mean vector and covariance matrix of $(V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_c)$ given V_j. Combining (25) and (26) we get,

$$\begin{aligned} P_j &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{v_{j1}^*} \cdots \int_{-\infty}^{v_{j(j-1)}^*} \int_{-\infty}^{v_{j(j+1)}^*} \cdots \int_{-\infty}^{v_{jc}^*} \right. \\ &\quad \left. \phi_{c-1}\left(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_c; \boldsymbol{\eta}^{(j)}, \boldsymbol{\Sigma}^{(j)}\right) dv_c \ldots dv_{j+1}\, dv_{j-1} \ldots dv_1 \right\} \phi(v_j) dv_j \\ &= \int_{-\infty}^{\infty} \Phi_{c-1}\left(v_{j1}^*, \ldots, v_{j(j-1)}^*, v_{j(j+1)}^*, \ldots, v_{jc}^*; \boldsymbol{\eta}^{(j)}, \boldsymbol{\Sigma}^{(j)}\right) \phi(v_j) dv_j. \end{aligned} \tag{27}$$

2.5 PROBABILITIES FOR EQUICORRELATED CHOICES

We will derive the expressions for $\boldsymbol{\eta}^{(j)}$ and $\boldsymbol{\Sigma}^{(j)}$ in (27) in the case where the correlation matrix $\boldsymbol{R}$ is a structured matrix with parameter ρ. More specifically, we assume $\boldsymbol{R}$ is equicorrelated matrix given in (13).

Note that if $(V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_c, V_j)$ is a permutation of $(V_1, V_2, \ldots, V_c)$ and $(V_1, V_2, \ldots, V_c) \sim N(\mathbf{0}, \boldsymbol{R})$ then $(V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_c, V_j)$ is also $N(\mathbf{0}, \boldsymbol{R})$ for the equicorrelated structure $\boldsymbol{R}$. We will need a well known property regarding the conditional distribution of the multivariate normal distribution stated in the Appendix as Theorem 2. To use the theorem, let us partition the equicorrelation matrix $\boldsymbol{R}$ as

$$\boldsymbol{R}_{c\times c} = \left(\begin{array}{cccc|c} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \hline \rho & \rho & \rho & \dots & 1 \end{array}\right) = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{pmatrix},$$

where $\boldsymbol{R}_{11} = \boldsymbol{R}_{c-1\times c-1}$, $\boldsymbol{R}_{21} = \boldsymbol{R}'_{12} = (\rho, \rho, \cdots, \rho)$, and $\boldsymbol{R}_{22} = 1$. Now,

$$\begin{aligned} \boldsymbol{R_{12}R_{22}^{-1}R_{21}} &= \boldsymbol{R_{12}R_{21}} \\ &= \begin{pmatrix} \rho \\ \rho \\ \vdots \\ \rho \end{pmatrix}_{c-1\times 1} \begin{pmatrix} \rho & \rho & \dots & \rho \end{pmatrix}_{1\times c-1} \\ &= \begin{pmatrix} \rho^2 & \rho^2 & \dots & \rho^2 \\ \rho^2 & \rho^2 & \dots & \rho^2 \\ \vdots & \vdots & \dots & \vdots \\ \rho^2 & \rho^2 & \dots & \rho^2 \end{pmatrix}_{c-1\times c-1}. \end{aligned}$$

Therefore by Theorem 2,

$$\boldsymbol{\Sigma}^{(j)} = \boldsymbol{R_{11} - R_{12}R_{22}^{-1}R_{21}} = \begin{pmatrix} 1-\rho^2 & \rho-\rho^2 & \dots & \rho-\rho^2 \\ \rho-\rho^2 & 1-\rho^2 & \dots & \rho-\rho^2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho-\rho^2 & \rho-\rho^2 & \dots & 1-\rho^2 \end{pmatrix}_{c-1\times c-1}$$

and

$$\boldsymbol{\eta}^{(j)} = \boldsymbol{R_{12}R_{22}^{-1}}(v_j - 0) = v \begin{pmatrix} \rho \\ \rho \\ \vdots \\ \rho \end{pmatrix}_{c-1\times 1} = \begin{pmatrix} \rho v_j \\ \rho v_j \\ \vdots \\ \rho v_j \end{pmatrix}_{c-1\times 1}.$$

Thus

$$(V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_c \mid V_j = v_j) \sim N(\boldsymbol{\eta}^{(j)}, \boldsymbol{\Sigma}^{(j)}), \text{where}$$

$$\boldsymbol{\eta}^{(j)} = \begin{pmatrix} \rho v_j \\ \rho v_j \\ \vdots \\ \rho v_j \end{pmatrix}_{c-1\times 1} \quad \text{and} \quad \boldsymbol{\Sigma}^{(j)} = \begin{pmatrix} 1-\rho^2 & \rho-\rho^2 & \ldots & \rho-\rho^2 \\ \rho-\rho^2 & 1-\rho^2 & \ldots & \rho-\rho^2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho-\rho^2 & \rho-\rho^2 & \ldots & 1-\rho^2 \end{pmatrix}_{c-1\times c-1}$$

Using these results, we can see the choice probability given by (27), when the correlation $\boldsymbol{R}$ is equicorrelated, simplifies to

$$P_{ij} = \int_{-\infty}^{\infty} \Phi_{c-1}\left(v^*_{ij1}, \ldots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \ldots, v^*_{ijc}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \phi(v)\, dv \qquad (28)$$

where $v^*_{ijk} = \Phi^{-1}\left(F(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v)))\right)$ for $k \neq j$. We wrote $v_j = v$ in the above since it is simply a variable of integration. Here

$$\boldsymbol{\eta}^{(ij)} = \begin{pmatrix} \rho v \\ \rho v \\ \vdots \\ \rho v \end{pmatrix}_{c-1\times 1} \quad \text{and} \quad \boldsymbol{\Sigma}^{(ij)} = \begin{pmatrix} 1-\rho^2 & \rho-\rho^2 & \ldots & \rho-\rho^2 \\ \rho-\rho^2 & 1-\rho^2 & \ldots & \rho-\rho^2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho-\rho^2 & \rho-\rho^2 & \ldots & 1-\rho^2 \end{pmatrix}_{c-1\times c-1}. \qquad (29)$$

Note that the integrand in (28) is a function of the deterministic components of the utilities, namely, it depends on the vector $(\mu_{i1}, \mu_{i2}, \ldots, \mu_{ic})$ corresponding to ith consumer. To be specific, the choice probability (28) depends on the deterministic components through the differences $(\mu_{ij} - \mu_{ik})$ only.

2.5.1 INDUCED CORRELATION

The correlation matrix $\boldsymbol{\Sigma}^{(ij)}$ in (28) is a function of ρ, which determines the Gaussian copula. The parameter ρ is the correlation of the normal random variables. These normal variables are transformed into the random components of the utilities which are distributed as Gumbel. It would be interesting to find the relation between

ρ and the correlation r of the Gumbel random variables. We study the relationship using simulations. The idea is to simulate correlated normal variables, then transform them to Gumbel using inverse transformation method and estimate the correlation between the Gumbel variables. The formal steps of the simulations is given below.

Vary ρ from 0.01 to 0.99. For any fixed ρ:

Step 1 Generate N pairs of bivariate normal random variables $(X_{1i}, X_{2i}) \sim BVN(0, 0, 1, 1, \rho)$.

Step 2 Let $F(\,\cdot\,)$ be the Gumbel distribution function given in (17). Obtain

$$\begin{aligned} Z_{1i} &= F^{-1}(X_{1i}) = -\log(-\log(\Phi(X_{1i}))), \\ Z_{2i} &= F^{-1}(X_{2i}) = -\log(-\log(\Phi(X_{2i}))). \end{aligned}$$

Step 3 Calculate sample correlation r from $(Z_{1i}, Z_{2i}), i = 1, 2, \ldots, N$.

In Figure 2, we have plotted the difference $(\rho - r)$ of the correlation coefficient between the two standard normal variables and the sample correlation coefficient to the two simulated Gumbel random variables as a function of ρ. As can be seen in the plot, $\rho - r$ is very close to 0 throughout the entire positive range of ρ which shows that ρ can be treated as the correlation between the random components of the utilities.

2.5.2 PROPERTIES

Note that the choice probability P_{ij} given in (28) is of the form

$$P_{ij} = \int_{-\infty}^{\infty} \Phi_{c-1}(\boldsymbol{v}(x))\, \phi(x)\, dx$$

where $\boldsymbol{v}$ is a function of x. Since $0 \leq \Phi_{c-1}(\boldsymbol{v}(x))\, \phi(x) \leq \phi(x)$, we have

$$0 \leq P_{ij} = \int_{-\infty}^{\infty} \Phi_{c-1}(\boldsymbol{v}(x))\, \phi(x)\, dx \leq \int\limits_{-\infty}^{\infty} \phi(x)\, dx = 1,$$

and therefore $0 \leq P_{ij} \leq 1$.

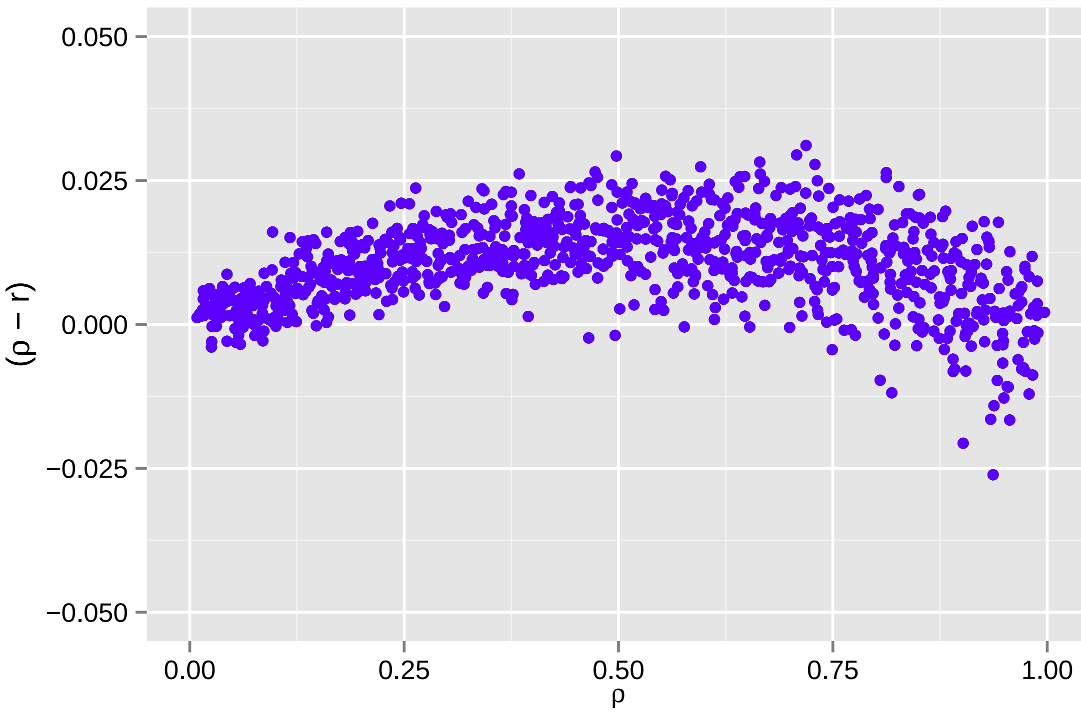

Figure 2: Difference between correlation of the copula and induced correlation plotted against the correlation of the copula

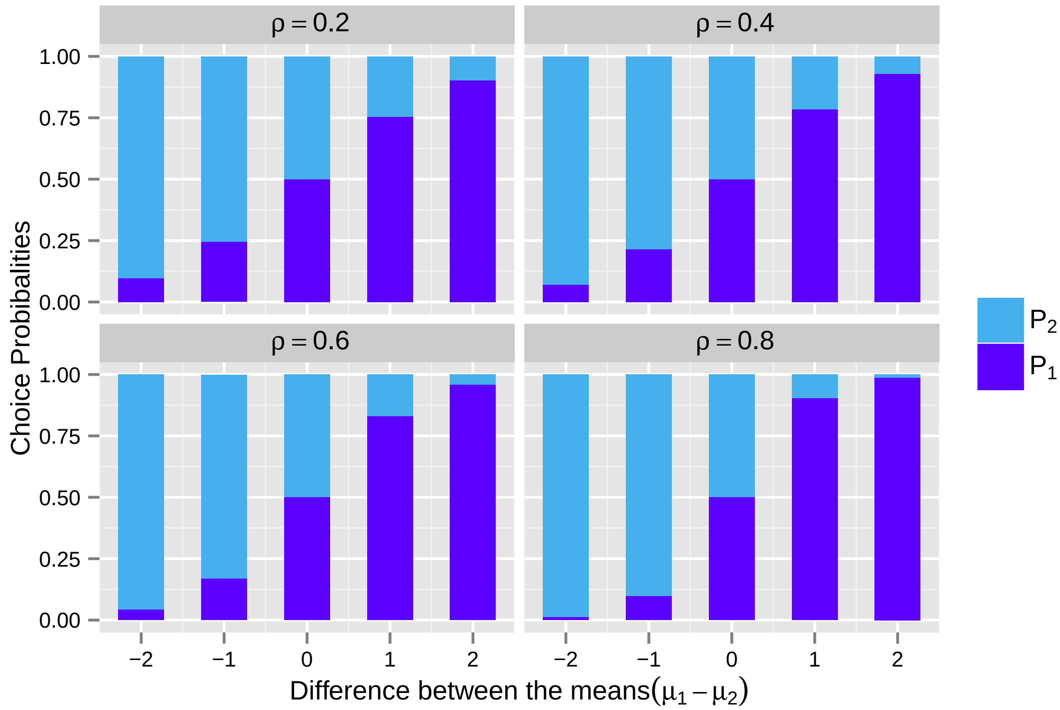

Figure 3: Choice probabilities for two choices

Next, we need to cross check that the choice probabilities for any customer will sum to 1, that is, $\Sigma_{j=1}^{c} P_{ij} = 1$ for all $i = 1, 2, \ldots, n$. It is very difficult to prove this analytically, because the expression (28) of the choice probability is an integral of a complicated multivariate function where the arguments themselves are composite and implicit functions of the variable of integration. However we will check this property using numerical calculations.

We first consider the case where there are only two choices in the choice set $(c = 2)$. Ignoring the index i for the customer, the choice probabilities become

$$\begin{aligned} P_1 &= \int_{-\infty}^{\infty} \Phi\left(\frac{v_2^* - \rho v}{\sqrt{1-\rho^2}}\right) \phi(v)\, dv, \quad v_2^* = \Phi^{-1}\left(F(\mu_1 - \mu_2 + F^{-1}(\Phi(v)))\right), \\ P_2 &= \int_{-\infty}^{\infty} \Phi\left(\frac{v_1^* - \rho v}{\sqrt{1-\rho^2}}\right) \phi(v)\, dv, \quad v_1^* = \Phi^{-1}\left(F(\mu_2 - \mu_1 + F^{-1}(\Phi(v)))\right). \end{aligned}$$

For different values of the correlation parameter ρ, we calculated P_1 and P_2 using numerical integration. The R code is in Appendix A. We plot these probabilities in a stacked bar plot for different values of $\rho = 0.2, 0.4, 0.6, 0.8$ and different pairs of means of the utility functions of the two choices (μ_1, μ_2). We chose (μ_1, μ_2) as $(3, 5), (3, 4), (3, 3), (3, 2)$ and $(3, 1)$, so that the difference $(\mu_1 - \mu_2)$ is $-2, -1, 0, 1$ and 2 respectively. The plots are given in Figure 3. The horizontal axes in the plots denote $\mu_1 - \mu_2$ and the vertical axes signify probabilities. As can be seen in the plots, P_1 and P_2 always add up to 1, for all values of ρ, μ_1 and μ_2. This proves that the choice probabilities for a single consumer add up to 1 for two choices. The calculations can easily be extended to any number of choices. Tables 3 and 4 display the probabilities in the case of three choices for different values of the means and correlation. Once again the three probabilities sum to one.

Besides verifying that the choice probabilities for a consumer sum up to 1, Figure 3 also showcases two important characteristics of the choice probabilities when the choices have the same correlation. When $\mu_1 - \mu_2 = 0$, $P_1 = P_2$ for any ρ. This property is intuitive and expected of any choice model since an equality of the utilities will result in equal probabilities of being chosen by the consumer, no matter what the correlation is among the choices. Also, when $\mu_1 - \mu_2 < 0$, P_1 decreases and P_2 increases as ρ increases. The reverse happens when $\mu_1 - \mu_2 > 0$. This tells us that an increasing correlation always increases the chance of the choice with highest utility to be chosen more than it increases the chance of a choice with lower utility. In short, an increasing correlation always favors the choice with higher utility.

In Table 3, we have shown the choice probabilities for three choices for varying ρ and means of the utility function, denoted by (μ_1, μ_2, μ_3). The choice probabilities are rounded up to two places of decimal. Each row in the table carries the sets of three choice probabilities for a given $\rho = 0, 0.1, \ldots, 0.9$ and for the mean vectors $(1, 1.5, 2)$, $(-4, -3.5, -3)$ and $(101, 101.5, 102)$ respectively. Needless to say, $\mu_2 - \mu_1 = 0.5$ and $\mu_3 - \mu_2 = 0.5$, for all the three mean vectors.

It can be observed that all the sets of choice probabilities add up to 1. Additionally, all three sets of probabilities in each row are the same. This is due to the fact the differences in the utilities in the three mean vectors are the same and goes to show that the choice probability described in (28) for the equicorrelated choices that we developed using the Gaussian copula, conforms to the idea presented in Section 1.5.3

Table 3: Choice probabilities for three equicorrelated choices and same differences between utilities

ρ	(μ_1, μ_2, μ_3)		
	(1, 1.5, 2)	(-4, -3.5, -3)	(101, 101.5, 102)
0	0.19, 0.31, 0.50	0.19, 0.31, 0.50	0.19, 0.31, 0.50
0.1	0.17, 0.30, 0.53	0.17, 0.30, 0.53	0.17, 0.30, 0.53
0.2	0.17, 0.30, 0.53	0.17, 0.30, 0.53	0.17, 0.30, 0.53
0.3	0.16, 0.29, 0.55	0.16, 0.29, 0.55	0.16, 0.29, 0.55
0.4	0.15, 0.29, 0.56	0.15, 0.29, 0.56	0.15, 0.29, 0.56
0.5	0.13, 0.28, 0.59	0.13, 0.28, 0.59	0.13, 0.28, 0.59
0.6	0.12, 0.27, 0.61	0.12, 0.27, 0.61	0.12, 0.27, 0.61
0.7	0.09, 0.25, 0.66	0.09, 0.25, 0.66	0.09, 0.25, 0.66
0.8	0.07, 0.22, 0.71	0.07, 0.22, 0.71	0.07, 0.22, 0.71
0.9	0.03, 0.16, 0.81	0.03, 0.16, 0.81	0.03, 0.16, 0.81

that only differences in utilities matter.

In Table 4, the choice probabilities according to (28) for three choices are calculated for three different sets of values of ρ and different values of the mean vector (μ_1, μ_2, μ_3) of the utilities. We have chosen (μ_1, μ_2, μ_3) to be $(2, 2.1, 2.5)$, $(0.4, 0.42, 0.5)$ and $(10, 10.5, 12.5)$. The second and third sets are scaled versions of the first and are obtained by multiplying 1/5 and 5 with the first set respectively. By doing this, we are just scaling of the mean component of the utility by a positive factor and not the actual utility.

We notice that in Table 4 too, all the triplets of choice probabilities have unity as their sums. But unlike in Table 3, the probabilities in a row are not same. As we have scaled only the means in the utilities, the differences in the mean have changed and so have the choice probabilities. This proves the idea presented in Section 1.5.4 that if we scale either the mean or the random component of the utility, the choice probability changes.

2.5.3 CASE OF INDEPENDENT UTILITIES

Suppose that the correlation parameter ρ equals zero. This corresponds to the assumption that the random components in the utilities are independent. We will

Table 4: Choice probabilities for three equicorrelated choices with scaled means of utilities

ρ	$(\mu_1,\ \mu_2,\ \mu_3)$		
	(2, 2.1, 2.5)	(0.4, 0.42, 0.5)	(10, 10.5, 12.5)
0	0.27, 0.29, 0.44	0.32, 0.33, 0.35	0.07, 0.11, 0.82
0.1	0.26, 0.29, 0.45	0.31, 0.32, 0.37	0.07, 0.11, 0.82
0.2	0.25, 0.28, 0.47	0.32, 0.32, 0.36	0.06, 0.10, 0.84
0.3	0.25, 0.28, 0.47	0.32, 0.32, 0.36	0.05, 0.08, 0.87
0.4	0.24, 0.28, 0.48	0.31, 0.32, 0.37	0.04, 0.07, 0.89
0.5	0.23, 0.27, 0.50	0.31, 0.32, 0.37	0.03, 0.06, 0.91
0.6	0.22, 0.26, 0.52	0.31, 0.32, 0.37	0.02, 0.04, 0.94
0.7	0.21, 0.25, 0.54	0.31, 0.32, 0.37	0.01, 0.03, 0.96
0.8	0.18, 0.23, 0.59	0.30, 0.32, 0.38	0.00, 0.01, 0.99
0.9	0.13, 0.18, 0.69	0.29, 0.31, 0.40	0.00, 0.00, 1.00

show in this case expression (28) reduces to a simpler form, which was originally derived by McFadden (1974). When $\rho = 0$, clearly the mean vector and the covariance matrix in (28) are $\boldsymbol{\eta}^{(ij)} = \mathbf{0}_{c-1}$, a $(c-1)$-dimensional vector with each element as 0, and $\boldsymbol{\Sigma}^{(ij)} = \boldsymbol{I}_{c-1}$, an identity matrix of order $(c-1)$. This means that the multivariate distribution function $\Phi_{c-1}(\cdot\ ;\ \mathbf{0}_{c-1}, \boldsymbol{I}_{c-1})$ is the product of $(c-1)$ univariate normal distribution functions. Thus in this case (28) reduces to

$$P_{ij} = \int_{-\infty}^{\infty} \left(\prod_{k \neq j} \Phi(v_{ik}^*) \right) \phi(v)\, dv, \tag{30}$$

where

$$\Phi(v_{ik}^*) = \Phi\left(\Phi^{-1}\left(F\left(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v))\right)\right)\right) = F\left(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v))\right) \tag{31}$$

The Gumbel distribution has the cumulative distribution function $F(z) = \exp(-\exp(-z))$. Its inverse is $F^{-1}(z) = -\log(-\log(z))$. Then

$$\begin{aligned} F\left(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v))\right) &= \exp\big(-\exp\big[\mu_{ik} - \mu_{ij} + \log(-\log(\Phi(v)))\big]\big) \\ &= (\Phi(v))^{\tau_k}, \end{aligned} \tag{32}$$

where $\tau_k = \exp\big(\mu_{ik} - \mu_{ij}\big)$. Using (31) and (32), the above integral (30) can be written as

$$
\begin{aligned}
P_{ij} &= \int_{-\infty}^{\infty} \left(\prod_{k \neq j} F\left(\mu_{ij} - \mu_{ik} + F^{-1}\left(\Phi(v)\right)\right) \right) \phi(v)\, dv \\
&= \int_{-\infty}^{\infty} \left(\prod_{k \neq j} (\Phi(v))^{\tau_k} \right) \phi(v)\, dv \\
&= \int_{-\infty}^{\infty} (\Phi(v))^{(\sum_{k \neq j} \tau_k)} \phi(v)\, dv. \qquad (33)
\end{aligned}
$$

Making a change of variable $\omega = \Phi(v)$ and noting that $d\omega = \phi(v)\, dv$ we see that (33) reduces to

$$
\begin{aligned}
P_{ij} &= \int_{0}^{1} w^{(\sum_{k \neq j} \tau_k)}\, dw \\
&= \frac{1}{(\sum_{k \neq j} \tau_k) + 1}. \qquad (34)
\end{aligned}
$$

Recall that $\tau_k = \exp\big(\mu_{ik} - \mu_{ij}\big)$, and hence $\tau_j = 1$. Therefore

$$
\sum_{k \neq j} \tau_k + 1 = \sum_{k=1}^{c} \tau_k = \left(\sum_{k=1}^{c} \exp(\mu_{ik}) \right) \Big/ \exp(\mu_{ij})
$$

and thus

$$
P_{ij} = \frac{1}{\left(\sum_{k \neq j} \tau_k\right) + 1} = \frac{\exp(\mu_{ij})}{\sum_{k=1}^{c} \exp(\mu_{ik})}, \qquad (35)
$$

which is the choice probability for the conditional logit model originally derived by McFadden (1974).

2.5.4 MAXIMUM LIKELIHOOD ESTIMATION

In this section, we present the expressions needed for maximum likelihood estimation of the parameters involved. Here is a quick review of the discrete model setup. We assume that there are n consumers and c choices in our discrete choice setup. The response is an indicator variable Y_{ij}, which takes the value one if i th consumer chooses j th alternative. We assume for any customer the joint distribution of the random components of the utilities is induced by the Gaussian copula with equicorrelated correlation matrix and the marginals are Gumbel. The choice probability corresponding to the ith consumer and the jth choice is given by P_{ij}, as defined in (28). Note that P_{ij} is a function of μ_{ik}'s, and ρ, the correlation parameter of the Gaussian copula.

We assume that there are p choice-specific covariates. For the ith consumer and the kth choice, X_{ikm} is the mth covariate, $i = 1, 2, \ldots, n$, $k = 1, 2, \ldots, c$, and $m = 1, 2, \ldots, p$. If $\boldsymbol{X}_{ik} = (X_{ik1}, X_{ik2}, \ldots, X_{ikp})$ then the mean of the (i, k)th utility is $\mu_{ik} = \boldsymbol{X}'_{ik}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ is a vector of regression coefficients for the p choice-specific covariates.

If $\boldsymbol{\theta} = (\boldsymbol{\beta}, \rho)$ is the vector of model parameters, then the log-likelihood $\ell(\boldsymbol{\theta})$ for n consumers is

$$\ell(\boldsymbol{\theta}) \;=\; \log\left(\prod_{i=1}^{n}\prod_{j=1}^{c} P_{ij}^{Y_{ij}}\right) = \sum_{i=1}^{n}\sum_{j=1}^{c} Y_{ij}\log(P_{ij}),$$

since $P_{ij} = P(Y_{ij} = 1)$ and the consumers are independent. The maximum likelihood estimate (MLE) of the parameter $\boldsymbol{\theta}$ is obtained by maximizing $\ell(\boldsymbol{\theta})$ over the parameter-space, or simply it is obtained by solving the score equation $\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$. Considering we have multiple parameters for the covariates ($\boldsymbol{\beta}$) and one correlation parameter ρ, the expressions for the first order derivatives of the log-likelihood are

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \left[\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}\ \frac{\partial\ell(\boldsymbol{\theta})}{\partial\rho}\right] \\ &= \left[\frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta_1}, \ldots, \frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta_p}, \ \frac{\partial\ell(\boldsymbol{\theta})}{\partial\rho}\right],\end{aligned}$$

with the first order partial derivatives being

$$\frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta_m} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{c} Y_{ij}\left(\frac{1}{P_{ij}}\frac{\partial P_{ij}}{\partial\beta_m}\right), m = 1, 2, \ldots, p, \text{ and}$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = \sum_{i=1}^{n} \sum_{j=1}^{c} Y_{ij} \left(\frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \rho} \right).$$

To get the partial derivatives stated above, we have to deduce the derivatives of the choice probability P_{ij} with respect to β_m, $m = 1, 2, \ldots, p$ and ρ. In Section 2.7, we provide detailed derivations and expressions for the partial derivatives of the choice probability. But before that, in the next section, we establish a theorem on the conditional distribution in the case of a multivariate normal distribution with equicorrelation correlation structure.

2.6 CONDITIONAL DISTRIBUTION OF A SUBSET OF NORMAL VARIABLES WITH AN EQUICORRELATION CORRELATION STRUCTURE

We will need the theorem below to derive the score equations for obtaining the maximum likelihood estimates.

Theorem **1.** Let $\boldsymbol{V} = (V_1, V_2, \ldots, V_c)$ be a column vector of dimension c and assume that it is distributed as normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{R}$, which is a equicorrelated correlation matrix with correlation parameter ρ. Partition $\boldsymbol{V}$ into $\boldsymbol{V}_1 = (V_1, V_2, \ldots, V_s)$, and $\boldsymbol{V}_2 = (V_{s+1}, V_{s+2}, \ldots, V_c)$, $s < c$. Then the conditional distribution of $\boldsymbol{V}_1$ given $\boldsymbol{V}_2 = \boldsymbol{v}_2$, where $\boldsymbol{v}_2 = (v_{s+1}, v_{s+2}, \ldots, v_c)$, is normal with mean

$$\boldsymbol{\mu}_{1|2} = \begin{pmatrix} \mu_1 + \frac{\rho}{1+(c-s-1)\rho} \sum\limits_{i=s+1}^{c} (v_i - \mu_i) \\ \mu_2 + \frac{\rho}{1+(c-s-1)\rho} \sum\limits_{i=s+1}^{c} (v_i - \mu_i) \\ \vdots \\ \mu_s + \frac{\rho}{1+(c-s-1)\rho} \sum\limits_{i=s+1}^{c} (v_i - \mu_i) \end{pmatrix}_{s\times 1} \tag{36}$$

and covariance matrix

$$\boldsymbol{R}_{1|2} = \frac{(1-\rho)(1+(c-s)\rho)}{1+(c-s-1)\rho}\begin{pmatrix} 1 & \frac{\rho}{1+(c-s)\rho} & \dots & \frac{\rho}{1+(c-s)\rho} \\ \frac{\rho}{1+(c-s)\rho} & 1 & \dots & \frac{\rho}{1+(c-s)\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho}{1+(c-s)\rho} & \frac{\rho}{1+(c-s)\rho} & \dots & 1 \end{pmatrix}_{s\times s}. \tag{37}$$

Proof. Let $\boldsymbol{V}_1 = (v_1, v_2, \dots, v_s)$ and $\boldsymbol{V}_2 = (V_{s+1}, V_{s+2}, \dots, V_c)$, $s < c$. Accordingly, let $\boldsymbol{\mu}$ be partitioned as $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ and $\boldsymbol{R}$ be partitioned as

$$\boldsymbol{R} = \left(\begin{array}{cccc|cccc} 1 & \rho & \dots & \rho & \rho & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho & \rho & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \dots & 1 & \rho & \rho & \dots & \rho \\ \hline \rho & \rho & \dots & \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & \dots & \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \dots & \rho & \rho & \rho & \dots & 1 \end{array}\right) = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{pmatrix}.$$

The dimensions of the sub-matrices are: $\boldsymbol{R}_{11}$ is $s\times s$, $\boldsymbol{R}_{12}$ is $s \times c - s$, $\boldsymbol{R}_{21}$ is $c - s \times s$ and $\boldsymbol{R}_{22}$ is $c - s \times c - s$. It follows from the properties of the multivariate normal distribution and Theorem 2 in the Appendix, $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ and $\boldsymbol{V}_1 \mid \boldsymbol{V}_2 = \boldsymbol{v}_2$ follow multivariate normal distributions of appropriate dimensions with the parameters $(\boldsymbol{\mu}_1, \boldsymbol{R}_{11})$, $(\boldsymbol{\mu}_2, \boldsymbol{R}_{22})$ and $(\boldsymbol{\mu}_{1|2}, \boldsymbol{R}_{1|2})$ respectively, where $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}(\boldsymbol{v}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{R}_{1|2} = \boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}$. Using the well known formula for the inverse of an equicorrelated matrix we can check that

$$\begin{aligned} \boldsymbol{R}_{22}^{-1} &= \frac{1}{1-\rho}\left(\boldsymbol{I}_{c-s} - \frac{\rho}{1+(c-s-1)\rho}\mathbf{1}\mathbf{1}'\right) \\ &= \frac{1}{1-\rho}\begin{pmatrix} 1-\frac{\rho}{1+(c-s-1)\rho} & -\frac{\rho}{1+(c-s-1)\rho} & \dots & -\frac{\rho}{1+(c-s-1)\rho} \\ -\frac{\rho}{1+(c-s-1)\rho} & 1-\frac{\rho}{1+(c-s-1)\rho} & \dots & -\frac{\rho}{1+(c-s-1)\rho} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{\rho}{1+(c-s-1)\rho} & -\frac{\rho}{1+(c-s-1)\rho} & \dots & 1-\frac{\rho}{1+(c-s-1)\rho} \end{pmatrix}_{c-s\times c-s} \end{aligned}$$

and

$$\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1} = \frac{\rho}{1+(c-s-1)\rho}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{s\times c-s}.$$

Therefore, the mean of $\boldsymbol{V}_1$ given $\boldsymbol{V}_2 = \boldsymbol{v}_2$ is

$$\begin{aligned} \boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}(\boldsymbol{v}_2 - \boldsymbol{\mu}_2) \\ &= \begin{pmatrix} \mu_1 + \frac{\rho}{1+(c-s-1)\rho}\sum_{i=s+1}^{c}(v_i - \mu_i) \\ \mu_2 + \frac{\rho}{1+(c-s-1)\rho}\sum_{i=s+1}^{c}(v_i - \mu_i) \\ \vdots \\ \mu_s + \frac{\rho}{1+(c-s-1)\rho}\sum_{i=s+1}^{c}(v_i - \mu_i) \end{pmatrix}_{s\times 1} \end{aligned} \tag{38}$$

Also,

$$\begin{aligned} \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21} &= \frac{\rho}{1+(c-s-1)\rho}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{s\times c-s}\begin{pmatrix} \rho & \rho & \dots & \rho \\ \rho & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \dots & \rho \end{pmatrix}_{c-s\times s} \\ &= \frac{(c-s)\rho^2}{1+(c-s-1)\rho}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{s\times s}. \end{aligned}$$

Thus the covariance of the conditional distribution of $\boldsymbol{V}_1$ given $\boldsymbol{V}_2 = \boldsymbol{v}_2$ is

$$\boldsymbol{R}_{1|2} = \boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}$$

$$= \frac{(1-\rho)(1+(c-s)\rho)}{1+(c-s-1)\rho} \begin{pmatrix} 1 & \frac{\rho}{1+(c-s)\rho} & \cdots & \frac{\rho}{1+(c-s)\rho} \\ \frac{\rho}{1+(c-s)\rho} & 1 & \cdots & \frac{\rho}{1+(c-s)\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho}{1+(c-s)\rho} & \frac{\rho}{1+(c-s)\rho} & \cdots & 1 \end{pmatrix}_{s\times s}. \quad (39)$$

It is interesting to note that the conditional covariance matrix (39) given above has also an equicorrelated structure. This proves that if we start with an equicorrelated multivariate normal distribution and derive the conditional distribution of any of its subset given the rest, the equicorrelation behavior will get carried over to the conditional distribution. The special cases where $s = c-1$, $s = c-2$, and $s = c-3$ are presented explicitly since we will need them later.

2.6.1 SPECIAL CASES

The conditional distribution of $s = c-1$ random variables given one variable $V_c = v_c$ is multivariate normal with mean and covariance matrix given by

$$\begin{pmatrix} \mu_1 + \rho(v_c - \mu_c) \\ \mu_2 + \rho(v_c - \mu_c) \\ \vdots \\ \mu_{c-1} + \rho(v_c - \mu_c) \end{pmatrix}_{c-1} \text{ and } (1-\rho^2) \begin{pmatrix} 1 & \frac{\rho}{1+\rho} & \cdots & \frac{\rho}{1+\rho} \\ \frac{\rho}{1+\rho} & 1 & \cdots & \frac{\rho}{1+\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho}{1+\rho} & \frac{\rho}{1+\rho} & \cdots & 1 \end{pmatrix}_{c-1}.$$

In the case $s = c-2$ the mean vector and covariance matrix of the conditional distribution are given by

$$\begin{pmatrix} \mu_1 + \frac{\rho}{1+\rho} \sum_{i=c-1}^{c} (v_i - \mu_i) \\ \mu_2 + \frac{\rho}{1+\rho} \sum_{i=c-1}^{c} (v_i - \mu_i) \\ \vdots \\ \mu_{c-2} + \frac{\rho}{1+\rho} \sum_{i=c-1}^{c} (v_i - \mu_i) \end{pmatrix}_{c-2} \text{ and } \frac{(1-\rho)(1+2\rho)}{1+\rho} \begin{pmatrix} 1 & \frac{\rho}{1+2\rho} & \cdots & \frac{\rho}{1+2\rho} \\ \frac{\rho}{1+2\rho} & 1 & \cdots & \frac{\rho}{1+2\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho}{1+2\rho} & \frac{\rho}{1+2\rho} & \cdots & 1 \end{pmatrix}_{c-2}.$$

The mean vector and the covariance matrix of $c-3$ random variables given $V_{c-2} =$

$v_{c-2}, V_{c-1} = v_{c-1}, V_c = v_c$ are

$$\begin{pmatrix} \mu_1 + \frac{\rho}{1+2\rho} \sum_{i=c-2}^{c} (v_i - \mu_i) \\ \mu_2 + \frac{\rho}{1+2\rho} \sum_{i=c-2}^{c} (v_i - \mu_i) \\ \vdots \\ \mu_{c-3} + \frac{\rho}{1+2\rho} \sum_{i=c-2}^{c} (v_i - \mu_i) \end{pmatrix}_{c-3} \text{ and } \frac{(1-\rho)(1+3\rho)}{1+2\rho} \begin{pmatrix} 1 & \frac{\rho}{1+3\rho} & \cdots & \frac{\rho}{1+3\rho} \\ \frac{\rho}{1+3\rho} & 1 & \cdots & \frac{\rho}{1+3\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho}{1+3\rho} & \frac{\rho}{1+3\rho} & \cdots & 1 \end{pmatrix}_{c-3}.$$

2.7 DERIVATIVES

In this section we will derive formulas for the derivatives of the choice probabilities with respect to the regression and correlation parameters. Let,

$$\boldsymbol{V}^{(ij)} = (v^*_{ij1}, \ldots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \ldots, v^*_{ijc})$$

where $v^*_{ijk} = \Phi^{-1}\left(F(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v))\right)$ for $k \neq j$. Therefore, the choice probability (28) can be written as

$$P_{ij} = \int_{-\infty}^{\infty} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \phi(v)\, dv$$

Recall that $\mu_{ik} = \boldsymbol{X'}_{ik}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$. For $m = 1, 2, \ldots, p$,

$$\begin{aligned} \frac{\partial P_{ij}}{\partial \beta_m} &= \frac{\partial}{\partial \beta_m} \int_{-\infty}^{\infty} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \phi(v)\, dv \\ &= \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial \beta_m} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \right\} \phi(v)\, dv \\ &= \int_{-\infty}^{\infty} \left\{ \sum_{k=1, k\neq j}^{c} \frac{\partial}{\partial v^*_{ijk}} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \frac{\partial v^*_{ijk}}{\partial \beta_m} \right\} \phi(v)\, dv \end{aligned}$$

Using Theorem 4, the derivative of $\Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right)$ with respect to v^*_{ijk} becomes

$$\frac{\partial}{\partial v^*_{ijk}} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) = \Phi_{c-2}\left(\boldsymbol{V}^{(ij)}_{-(k)}; \boldsymbol{\eta}^{(ij)}_{-(k)}, \boldsymbol{\Sigma}^{(ij)}_{-(k)}\right) \phi(v^*_{ijk}; \rho v, (1-\rho^2)),$$

where $\boldsymbol{V}^{(ij)}_{-(k)}$ is obtained by deleting the kth element in $\boldsymbol{V}^{(ij)}$. $\boldsymbol{\eta}^{(ij)}_{-(k)}$ and $\boldsymbol{\Sigma}^{(ij)}_{-(k)}$ are the conditional mean and variance of $\boldsymbol{V}^{(ij)}_{-(k)}$ given v^*_{ijk} and can be obtained from (38) and (39). To find the derivative of v^*_{ijk} with respect to β_m, we remember that $v^*_{ijk} = \Phi^{-1}\left(F(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v)))\right)$ for $k \neq j$ and thus,

$$\frac{\partial v^*_{ijk}}{\partial \beta_m} = \frac{f((\boldsymbol{X}_{ij} - \boldsymbol{X}_{ik})'\boldsymbol{\beta} + F^{-1}(\Phi(v)))}{\phi(\Phi^{-1}\left(F((\boldsymbol{X}_{ij} - \boldsymbol{X}_{ik})'\boldsymbol{\beta} + F^{-1}(\Phi(v)))\right))}(X_{ijm} - X_{ikm})$$

Hence,

$$\begin{aligned}\frac{\partial P_{ij}}{\partial \beta_m} &= \int_{-\infty}^{\infty} \left\{ \sum_{k=1,k\neq j}^{c} \Phi_{c-2}\left(\boldsymbol{V}^{(ij)}_{-(k)}; \boldsymbol{\eta}^{(ij)}_{-(k)}, \boldsymbol{\Sigma}^{(ij)}_{-(k)}\right) \phi(v^*_{ijk}; \rho v, (1-\rho^2)) \right. \\ &\qquad \left. \frac{f((\boldsymbol{X}_{ij} - \boldsymbol{X}_{ik})'\boldsymbol{\beta} + F^{-1}(\Phi(v)))}{\phi(\Phi^{-1}\left(F((\boldsymbol{X}_{ij} - \boldsymbol{X}_{ik})'\boldsymbol{\beta} + F^{-1}(\Phi(v)))\right))}(X_{ijm} - X_{ikm}) \right\} \phi(v)\, dv,\end{aligned}$$

where

$$\boldsymbol{V}^{(ij)}_{-(k)} = \begin{cases} (v^*_{ij1}, \ldots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \ldots, v^*_{ij(k-1)}, v^*_{ij(k+1)}, \ldots, v^*_{ijc}) & \text{, if } k > j \\ \\ (v^*_{ij1}, \ldots, v^*_{ij(k-1)}, v^*_{ij(k+1)}, \ldots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \ldots, v^*_{ijc}) & \text{, if } k < j \end{cases}$$

$$\boldsymbol{\eta}^{(ij)}_{-(k)} = \begin{pmatrix} \frac{\rho(v+v^*_{ijk})}{1+\rho} \\ \frac{\rho(v+v^*_{ijk})}{1+\rho} \\ \vdots \\ \frac{\rho(v+v^*_{ijk})}{1+\rho} \end{pmatrix}_{c-2\times 1} = \frac{\rho(v+v^*_{ijk})}{1+\rho}\mathbf{1}_{c-2} \text{ and}$$

$$\boldsymbol{\Sigma}^{(ij)}_{-(k)} = \begin{pmatrix} \frac{1+\rho-2\rho^2}{1+\rho} & \frac{\rho(1-\rho)}{1+\rho} & \cdots & \frac{\rho(1-\rho)}{1+\rho} \\ \frac{\rho(1-\rho)}{1+\rho} & \frac{1+\rho-2\rho^2}{1+\rho} & \cdots & \frac{\rho(1-\rho)}{1+\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\rho(1-\rho)}{1+\rho} & \frac{\rho(1-\rho)}{1+\rho} & \cdots & \frac{1+\rho-2\rho^2}{1+\rho} \end{pmatrix}_{c-2\times c-2}$$

Now, we shall find the expression for the derivative of the choice probability with respect to the correlation parameter ρ of the Gaussian copula.

$$\begin{aligned}\frac{\partial P_{ij}}{\partial \rho} &= \frac{\partial}{\partial \rho}\int_{-\infty}^{\infty} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \phi(v)\, dv \\ &= \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial \rho}\Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \right\} \phi(v)\, dv.\end{aligned}$$

We proceed now to find the derivative of $\Phi_{c-1}\left(\boldsymbol{V}^{(ij)};\boldsymbol{\eta}^{(ij)},\boldsymbol{\Sigma}^{(ij)}\right)$ with respect to ρ. As $\boldsymbol{\Sigma}^{(ij)}$ is a covariance matrix, it would be easier to find the derivative if we transform that to a correlation matrix. Note that

$$\begin{aligned}
&\frac{\partial}{\partial\rho}\Phi_{c-1}\left(\boldsymbol{V}^{(ij)};\boldsymbol{\eta}^{(ij)},\boldsymbol{\Sigma}^{(ij)}\right)\\
&=\frac{\partial}{\partial\rho}\Phi_{c-1}\left(w^*_{ij1},\ldots,w^*_{ij(j-1)},w^*_{ij(j+1)},\ldots,w^*_{ijc};\mathbf{0},\boldsymbol{R}_{c-1}(\lambda)\right)
\end{aligned}$$

where $w_{ijk}=(v_{ijk}-\rho v)/(\sqrt{1-\rho^2})$ for $k\neq j$ and

$$\boldsymbol{R}_{c-1}(\lambda)=\begin{pmatrix} 1 & \lambda & \lambda & \ldots & \lambda\\ \lambda & 1 & \lambda & \ldots & \lambda\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \lambda & \lambda & \lambda & \ldots & 1\end{pmatrix}_{c-1},\quad \lambda=\frac{\rho}{1+\rho}.$$

The above derivative can be calculated using the chain rule- by first taking the partial derivatives of the $\Phi_{c-1}(.)$ function with respect to w_{ijk}'s, and then by taking the derivative with respect to λ. So, if $\boldsymbol{W}^{(ij)}=(w_{ij1},\ldots,w_{ij(j-1)},w_{ij(j+1)},\ldots,w_{ijc})$, then

$$\begin{aligned}
&\frac{\partial}{\partial\rho}\Phi_{c-1}\left(\boldsymbol{V}^{(ij)};\boldsymbol{\eta}^{(ij)},\boldsymbol{\Sigma}^{(ij)}\right)\\
&=\frac{\partial}{\partial\rho}\Phi_{c-1}\left(\boldsymbol{W}^{(ij)};\mathbf{0},\boldsymbol{R}_{c-1}(\lambda)\right)\\
&=\sum_{k=1,k\neq j}^{c}\left\{\frac{\partial}{\partial w_{ijk}}\Phi_{c-1}\left(\boldsymbol{W}^{(ij)};\mathbf{0},\boldsymbol{R}_{c-1}(\lambda)\right)\frac{\partial w_{ijk}}{\partial\rho}\right\}\\
&\quad+\frac{\partial}{\partial\lambda}\Phi_{c-1}\left(\boldsymbol{W}^{(ij)};\mathbf{0},\boldsymbol{R}_{c-1}(\lambda)\right)\frac{\partial\lambda}{\partial\rho}.
\end{aligned}$$

Using Theorem 4 given in the Appendix we get

$$\begin{aligned}
&\frac{\partial}{\partial w_{ijk}}\Phi_{c-1}\left(\boldsymbol{W}^{(ij)};\mathbf{0},\boldsymbol{R}_{c-1}(\lambda)\right)\\
&\quad=\sum_{k=1,k\neq j}^{c}\Phi_{c-2}\left(\boldsymbol{W}^{(ij)}_{-(k)};\lambda w_{ijk}\mathbf{1}_{c-2},(1-\lambda^2)\boldsymbol{R}_{c-2}\left(\frac{\lambda}{1+\lambda}\right)\right)\phi(w_{ijk};0,1).
\end{aligned}$$

and using Theorem 5 we have

$$\begin{aligned}
&\frac{\partial}{\partial\lambda}\Phi_{c-1}\left(\boldsymbol{W}^{(ij)};\boldsymbol{0},\boldsymbol{R}_{c-1}(\lambda)\right) \\
&\quad= \sum_{k=1}^{c}\sum_{l=k+1}^{c}\Phi_{c-3}\Bigg(\boldsymbol{W}^{(ij)}_{-(kl)};\frac{\lambda}{1+\lambda}(w_{ijk}+w_{ijl})\boldsymbol{1}_{c-3}, \\
&\qquad\qquad \frac{(1-\lambda)(1+2\lambda)}{1+\lambda}\boldsymbol{R}_{c-3}\left(\frac{\lambda}{1+\lambda}\right)\Bigg)\phi_2\left(\boldsymbol{W}^{(ij)}_{(kl)};\boldsymbol{0},\boldsymbol{R}_2(\lambda)\right).
\end{aligned}$$

Furthermore

$$\frac{\partial w_{ijk}}{\partial\rho} = \frac{\rho v_{ijk}-v}{(1-\rho^2)^{3/2}}, \quad \frac{\partial\lambda}{\partial\rho} = \frac{1}{(1+\rho^2)}.$$

Substituting these we get

$$\begin{aligned}
\frac{\partial P_{ij}}{\partial\rho} &= \int_{-\infty}^{\infty}\sum_{k=1,k\neq j}^{c}\Phi_{c-2}\left(\boldsymbol{W}^{(ij)}_{-(k)};\lambda w_{ijk}\boldsymbol{1}_{c-2},(1-\lambda^2)\boldsymbol{R}_{c-2}\left(\frac{\lambda}{1+\lambda}\right)\right) \\
&\qquad \phi(w_{ijk};0,1)\frac{\rho v_{ijk}-v}{(1-\rho^2)^{3/2}}\,dv \\
&\quad + \int_{-\infty}^{\infty}\sum_{k=1}^{c}\sum_{l=k+1}^{c}\Phi_{c-3}\Bigg(\boldsymbol{W}^{(ij)}_{-(kl)};\frac{\lambda}{1+\lambda}(w_{ijk}+w_{ijl}\boldsymbol{1}_{c-3}, \\
&\qquad \frac{(1-\lambda)(1+2\lambda)}{1+\lambda}\boldsymbol{R}_{c-3}\left(\frac{\lambda}{1+\lambda}\right)\Bigg) \\
&\qquad \phi_2\left(\boldsymbol{W}^{(ij)}_{(kl)};\boldsymbol{0},\boldsymbol{R}_2(\lambda)\right)\frac{1}{(1+\rho^2)}\,dv.
\end{aligned}$$

This completes the derivation of the score functions for estimating the parameters using maximum likelihood.

2.8 MODEL FITTING

In this section, we illustrate an application of the multivariate discrete choice

Table 5: Analysis of Travel Mode Data

Variable	MDCG Equicorrelation			CNL		
	Estimate	SD	P Value	Estimate	SD	P Value
Intercept Air	4.9635	1.1552	<0.0001	5.2047	0.9052	<0.0001
Train	4.0475	0.7952	<0.0001	4.3606	0.5107	<0.0001
Bus	3.5090	0.5160	<0.0001	3.7632	0.5063	<0.0001
Car	0	—	—	0	—	—
Waiting Time	-0.0983	0.0155	<0.0001	-0.1037	0.0109	<0.0001
Travel Cost	-0.0795	0.0206	<0.0001	-0.0849	0.0194	<0.0001
Travel Time	-0.0125	0.0030	<0.0001	-0.0133	0.0025	<0.0001
Generalized Cost	0.0665	0.0224	<0.0001	0.0693	0.0174	<0.0001
ρ	0.2206	0.1093	<0.0001	—	—	—
AIC	394.65			384.01		
$\boldsymbol{R^2_M}$	0.3332			0.3507		
$\boldsymbol{R^2_{M,Adj}}$	0.3050			0.3255		

Gumbel model with equicorrelation structure (MDCG-Equicorrelation) that we developed using the Gaussian copula, on a real life data. To compare this model with an existing and popular model, we chose the conditional logit model (CNL) model for which the choice probability is expressed in (11).

Consider the discrete choice data given in Table 21.2 of Greene (2003). The data consists of information on 210 travelers' trips between Sydney and Melbourne in Australia for non-business purposes. The choice set is a collection a four modes of travel, namely Air, Train, Bus and Car. Among all the 210 travelers, 58 (27.6%) chose to travel by air, 63 (30%) chose to travel by train, 30 (14.3%) chose to travel by bus and 59 (28.1%) chose to drive a car.

We choose several choice-specific variables for covariates such as waiting time, travel cost, travel time and generalized cost for each traveler and for each mode of choice. Additionally, there being four choices, we include three intercept terms by making the intercept for the choice car to be 0. Our goal is to find maximum likelihood estimates for different parameters and their standard errors.

The choice probability for the MDCG-Equicorrelation model given in (28) is a very complex function as it involves integration of the multivariate normal cumulative distribution function (CDF) on the real line. Evaluation of the multivariate normal CDF is itself a difficult task and numerous authors have proposed methods to solve this problem. See Johnson et al. (2000) and Kotz et al. (2000) for a detailed description of the methods. We have used the R-package 'mvtnorm' in our program to compute the multivariate normal CDF for the MDCG-equicorrelation model. To obtain the standard errors of the parameter estimates we will need the Hessian which involves computing second order derivatives of the log-likelihood. However analytical expressions for the second order derivatives are extremely complex. Therefore as an alternative we use the bootstrap method to get the standard errors. The bootstrap method is a resampling procedure that aids in the calculation of standard errors. An excellent reference to the bootstrap method is Efron (1970). Below we present detailed steps of the algorithm to calculate standard errors of the maximum likelihood estimates.

Step 1: For $b = 1, 2, \ldots, B$

a. Generate a random sample $\boldsymbol{I}_b$ of 210 integers by sampling with replacement from the set $\{1, 2, \ldots, 210\}$. Let $\boldsymbol{I}_b = \{I_1, I_2, \ldots, I_{210}\}$. $1 \leq I_k \leq 210$ and clearly, I_k's may not be unique, $k = 1, 2, \ldots, 210$.

b. Generate the bth bootstrap sample S_b^* by including data of the I_kth consumers in the original sample, where $I_k \in \boldsymbol{I}_b$ and $k = 1, 2, \ldots, 210$. Clearly, all the consumers in S_b^* may not be unique.

c. Run the MDCG equicorrelation model on the bootstrap sample S_b^* and calculate the maximum likelihood estimates for the bth bootstrap iteration.

Step 2: Calculate the standard error of the estimates of the parameters of the MDCG Equicorrelation by using the following formulae: if $\hat{\theta}_b^*$ is the bth bootstrap estimate of a parameter θ, then the bootstrap estimate and standard error of θ are

$$\bar{\hat{\theta}}^* = \frac{1}{B}\sum_{i=1}^{B} \hat{\theta}_b^*, \quad \hat{se}(\hat{\theta}^*) = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2}$$

We chose $B = 50$. For the estimates of the parameters and their standard errors using the CNL model, we used "Proc MDC" in the SAS software.

Table 5 provides point estimates, standard errors and p-values for t-tests for both the MDCG-equicorrelation and the CNL models. It also presents the AIC criterion and McFadden's R^2 and adjusted R^2 for comparison of likelihoods of the two models. The R code used to generate the results in this table is provided in the Appendix of this dissertation.

As can be seen from the estimates of the parameters in the results, both models show almost similar behavior. Though train was the most preferred mode of travel measured by raw numbers (30%), looking at the estimates of the intercepts in both the models, it can be argued that that random utility maximization theory suggests the travelers assigned more utility to traveling by air. The negative coefficients for waiting time, travel cost and travel time indicate that consumers prefer cheaper mode of transportation with less waiting time and/or travel time. The estimate of the equicorrelation parameter is 0.22, and it is significant.

The AIC, R^2_M and $R^2_{M,Adj}$ statistics for the two models are very close, but they show that the MDCG with equicorrelation model does not perform better than the CNL model for this data to capture the choice behavior of the travelers . It is to be kept in mind that the fitting the models on this data set is just an exercise for demonstration. The MDCG-equicorrelation is suited when any two alternatives in the choice set have equal correlation- which is not the case for this data set. Air, train and bus are public transports while car is a private transport. Hypothetically, for traveling modes, if we had only public transport modes (or only private transport modes) in our choice set, the suitability of the MDCG equicorrelation model would be far more appropriate.

CHAPTER 3

ORDERED CHOICE MODEL WITH GAUSSIAN COPULA

3.1 ORDERED CHOICES

In some discrete choice problems, the choices are qualitative and there are situations where a natural order is inherently present within those qualitative choices. In other words the choices could be categories that are ordinal in nature. Also there could be a natural measure of distance between the choices. In these scenarios, from consumer point of view, utilities of two choices close to each other will be highly correlated compared to the ones that are further apart. For example, consider a survey where consumers are asked to rate an application (app) that they installed on their smart-phone. The ratings could range from “Very Bad”, “Bad”, “Mediocre”, “Good”, and “Excellent”, or it could be a numeral rating from 1 to 5. Clearly a rating of “Excellent” is better than “Good”, which is better than “Mediocre” and so on. Thus there is a natural order among the choices. Another example is how much a family spends on their weekly groceries. In this example the choices can be categorized as “less than \$50”, “\$50-\$100”, “\$100-\$200” and “\$200 or more.” and there is a natural ordering of the choices. Please note that ordering of the choices is different and should not be confused with the ordering of the utilities. The latter ordering is consumer dependent unlike the former which is independent of the consumers.

In the rating of the app example, the response of a consumer will depend on several covariates such as frequency of the app usage, whether or not advertisements appear while the user was using the app, and connectivity of the app to the internet etc. On the other hand, in the second example possible covariates that influence the choices are family income, size of the household, special dietary needs, and several other factors. The ordered choice is selected by the consumer after ordering the utilities which are covariate dependent. Several models have been developed to deal with ordered discrete choice data. McCullagh (1980) derived the ordered logit model to do regression analysis of ordered discrete responses by using the logistic distribution. The ordered probit model, originally proposed by Aitchison and Silvey (1957), uses the normal distribution function to fit regression models for ordered categorical data. Though these two models are widely used when the responses in a regression problem are discrete and can be ordered, they do not use the random utility maximization theory, which is the foundation for the discrete choice models. Based on the GEV family of discrete choice models by McFadden (1978), Small (1987) introduced the ordered generalized extreme value distribution (OGEV) model. This model assumes that the choices are grouped into intersecting nests. And choices that are more closer to each other in the ordering have higher correlations. In this chapter our goal is to generalize the conditional logit model by incorporating a correlation parameter in the case of ordered choices.

3.2 CHOICE PROBABILITY USING GAUSSIAN COPULA WITH AR(1) STRUCTURE

To begin with, we consider the same choice situation as in Chapter 2, where there are n consumers each facing c choices with one difference. Unlike the previous model we assume that the choices have a natural ordering. As in Chapter 2, we assume that the random components Z_{ij}'s of the utilities are distributed marginally as Gumbel and they are correlated. However, the model that we consider in this chapter differs in the correlation structure. For the ordered and dependent choices a reasonable correlation structure is the model where for any i, $corr(Z_{ij}, Z_{ik})$ depends on $|j-k|$ in such a way that if $|j-k|$ increases then $corr(Z_{ij}, Z_{ik})$ decreases. Thus an appropriate correlation model is the autoregressive of order 1 or AR(1) correlation matrix given by

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{c-1} \\ \rho & 1 & \rho & \dots & \rho^{c-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{c-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{c-1} & \rho^{c-2} & \rho^{c-3} & \dots & 1 \end{pmatrix}_{c\times c}. \tag{40}$$

The determinant of $\boldsymbol{R}$ is $(1-\rho^2)^{c-1}$. Though $\boldsymbol{R}$ is positive definite for all values of ρ

in $(-1, 1)$, we will restrict ρ to the positive range of $(0, 1)$. In summary in this chapter we consider the same discrete choice model given in Chapter 2 with one major change. Before we employed the Gaussian copula with equicorrelated structure and in this chapter we will use the Gaussian copula but replace the equicorrelated structure with AR(1) structure given in (40). Since the expression (15) for the choice probability is valid for any correlation structure the probability that the ith customer picks jth choice is given by

$$\begin{aligned} P_{ij} &= \int_{-\infty}^{\infty} \Pr\left(Z_{ik} < z^*_{ijk}, k = 1, 2, \ldots, c, k \neq j \mid Z_{ij} = z_{ij}\right) \cdot f(z_{ij})\, dz_{ij}, \\ &= \int_{-\infty}^{\infty} \Phi_{c-1}\left(v^*_{ij1}, \ldots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \ldots, v^*_{ijc}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \phi(v)\, dv, \end{aligned}$$

where $v^*_{ijk} = \Phi^{-1}\left(F(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v)))\right)$ for $k \neq j$, $\boldsymbol{\eta}^{(ij)}$ and $\boldsymbol{\Sigma}^{(ij)}$ are the mean vector and covariance of $(V_{i1}, \ldots, V_{i(j-1)}, V_{i(j+1)}, \ldots, V_{ic} \mid V_{ij} = v)$. Here

$$(V_{i1}, V_{i2}, \ldots, V_{ic}) \sim N(\mathbf{0}, \boldsymbol{R}), \tag{41}$$

where the correlation matrix $\boldsymbol{R}$ has the autoregressive structure given in (40). To compute the choice probability we need to derive simplified expressions for $\boldsymbol{\eta}^{(ij)}$ and $\boldsymbol{\Sigma}^{(ij)}$, which we will do in the succeeding sections.

3.2.1 INDUCED CORRELATIONS

We assumed that any two choices have a correlation among themselves which can be ordered by the distance of the choices in the choice set and hence used the AR(1) matrix in the Gaussian copula with ρ being the correlation parameter as expressed in (41), but ρ is not exactly the correlation between two adjacent choices. When we use the formula for the choice probability to maximize the likelihood to find estimates of the parameters involved, we shall get an estimate of ρ, and it might be misinterpreted as the correlation among two adjacent choices. Though in our method, there is no direct way of estimating the correlation between the choices, ρ can be estimated.

In order to see how close the actual correlations between the ordered choices are to the correlations in the copula in the case of four choices, for example, we first generate four Gumbel random variables from four simulated normal random variables with an AR(1) correlation structure and calculate the correlations between the pairs of Gumbel random variables. Here are the steps of the algorithm.

Vary ρ from 0.01 to 0.99 and for any fixed ρ:

Step 1 Generate N random vectors $(X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{5i}) \sim N(\mathbf{0}, \boldsymbol{R}_5)$, $i = 1, 2 \ldots, N$, where

$$\boldsymbol{R}_5 = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}.$$

Step 2 Obtain
$Z_{1i} = -\log(-\log(\Phi(X_{1i})),$
$Z_{2i} = -\log(-\log(\Phi(X_{2i})),$
$Z_{3i} = -\log(-\log(\Phi(X_{3i})),$
$Z_{4i} = -\log(-\log(\Phi(X_{4i})),$
$Z_{5i} = -\log(-\log(\Phi(X_{5i})).$
Clearly, $Z_1, \ldots, Z_5$ are univariate Gumbel random variables.

Step 3 Calculate four sample correlation coefficients- r_1 as correlation between (Z_{1i}, Z_{2i}), r_2 as correlation between (Z_{1i}, Z_{3i}), r_3 as correlation between (Z_{1i}, Z_{4i}), r_4 as correlation between $(Z_{1i}, Z_{5i}), i = 1, 2, \ldots, N$.

Step 4 Calculate $(\rho - r_1), (\rho^2 - r_2), (\rho^3 - r_3)$ and $(\rho^4 - r_4)$.

We chose $N = 100,000$ and plotted $(\rho - r_1), (\rho^2 - r_2), (\rho^3 - r_3)$ and $(\rho^4 - r_4)$ separately against different values of ρ. These plots are depicted in Figure 4. All the four plots show that the difference between the correlation of the copula and the induced correlation is very small. In fact, as we shall focus on ρ between 0 and 1, its estimate can well be interpreted as an estimate of the correlation between the parameters since the plots show that in the range $(0, 1)$ the four differences are very close to 0.

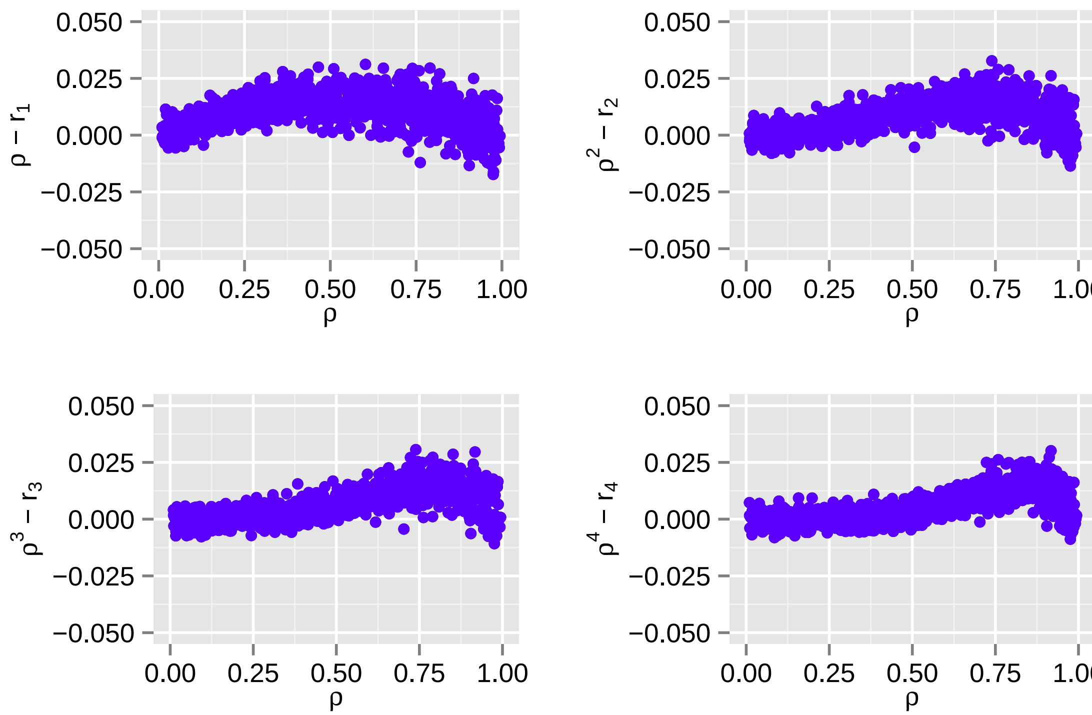

Figure 4: Difference between correlation of the copula and induced correlation plotted against the correlation of the copula

The first plot in Figure 4 is similar to that of Figure 2. This makes sense as in the case of two choices the AR(1) model is exactly same as the equicorrelation model and thus the plot of $\rho - r_1$ looks the same. Each of the four plots show that the difference between the induced correlations and the corresponding correlation of the copula stay in the proximity of 0. Hence, the estimate of the correlation parameter ρ of the copula can be treated as the estimated correlation coefficient between two consecutive choices in the ordered choice set.

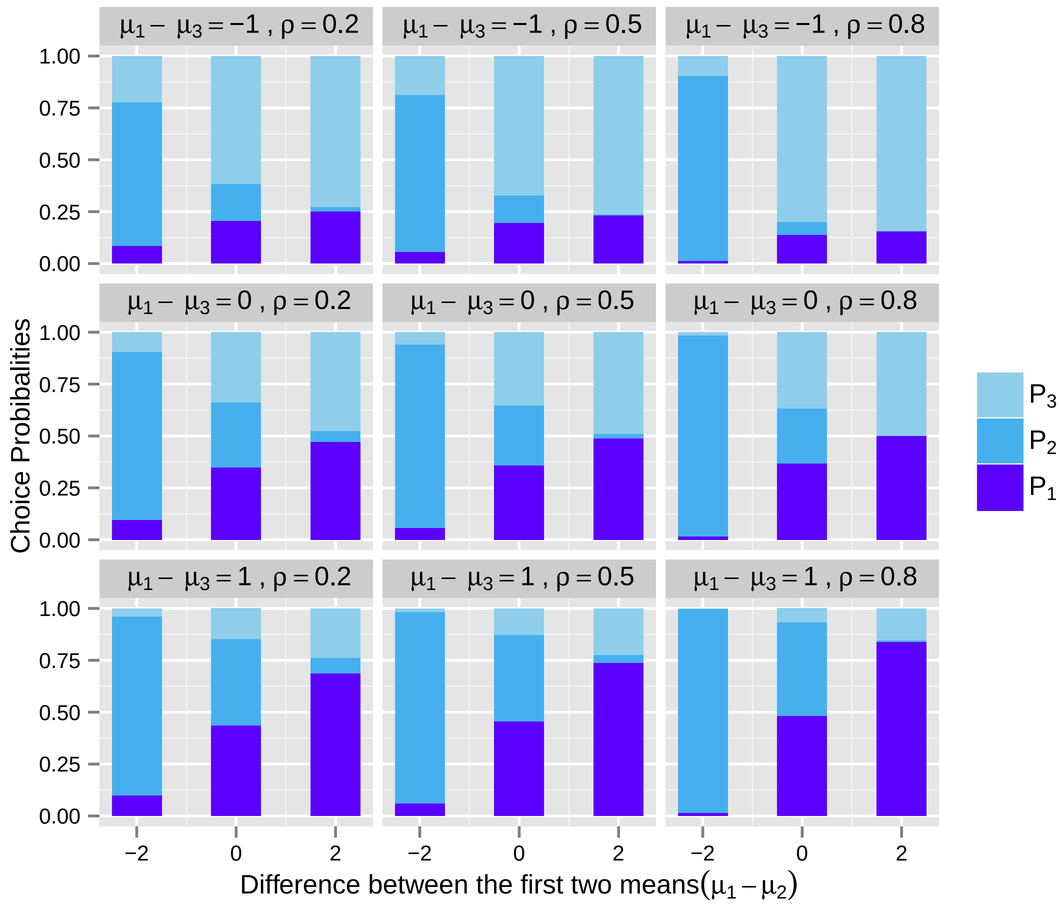

Figure 5: Simulated choice probabilities for three ordered choices

3.2.2 PROPERTIES

It can be shown, using the same proof as in Section 2.5.2, that the choice probability for ordered choices in (46) is such that $0 \leq P_{ij} \leq 1$.

To show that the choice probabilities for the ith consumer add up to 1, we use simulations. In (46) let us omit i for the time being. For $c = 3$, the choice probability then is a function of (μ_1, μ_2, μ_3) and ρ. We fixed μ_1 as 3. We sequentially chose μ_2 as $5, 3$ and 1, μ_3 as $4, 3$ and 2, and ρ as $0.2, 0.5$ and 0.8. For nine different combinations of $\mu_1 - \mu_3$ and ρ, we plot the choice probabilities in stacked bar charts against the three values of $\mu_1 - \mu_2$. These plots are illustrated in Figure 5. All of the bars in all nine plots climb up to the value 1 in the vertical axis, indicating that, in fact, the choice probabilities add up to 1.

Table 6: Choice probabilities for three ordered choices and same differences between utilities

ρ	$(\mu_1,\ \mu_2,\ \mu_3)$		
	(1, 1.5, 2)	(-4, -3.5, -3)	(101, 101.5, 102)
0	0.19, 0.31, 0.50	0.19, 0.31, 0.50	0.19, 0.31, 0.50
0.1	0.18, 0.29, 0.53	0.18, 0.29, 0.53	0.18, 0.29, 0.53
0.2	0.18, 0.28, 0.54	0.18, 0.28, 0.54	0.18, 0.28, 0.54
0.3	0.18, 0.27, 0.55	0.18, 0.27, 0.55	0.18, 0.27, 0.55
0.4	0.17, 0.26, 0.57	0.17, 0.26, 0.57	0.17, 0.26, 0.57
0.5	0.17, 0.25, 0.58	0.17, 0.25, 0.58	0.17, 0.25, 0.58
0.6	0.16, 0.24, 0.60	0.16, 0.24, 0.60	0.16, 0.24, 0.60
0.7	0.14, 0.22, 0.64	0.14, 0.22, 0.64	0.14, 0.22, 0.64
0.8	0.11, 0.19, 0.70	0.11, 0.19, 0.70	0.11, 0.19, 0.70
0.9	0.07, 0.14, 0.79	0.07, 0.14, 0.79	0.07, 0.14, 0.79

In Table 6, we calculate the three choice probabilities P_1, P_2, P_3 for ρ in $(0, 0.1, \ldots, 0.9)$ and three combinations of means of utilities $(\mu_1, \mu_2, \mu_3) = (1, 1.5, 2), (-4, -3.5, -3)$ and $(101, 101.5, 102)$. It is easy to verify that sum of P_1, P_2 and P_3 is always 1 no matter what the values of (μ_1, μ_2, μ_3) and ρ are. This too, along with Figure 5, illustrates that the sum of the choice probabilities for the ith customer is 1, for all i. Also, in each row we get the same set values for P_1, P_2, P_3 no matter what (μ_1, μ_2, μ_3) is. This property can be attributed to the equal differences between μ_1, μ_2 and μ_3 for a fixed ρ and this precisely proves that the choice probability for ordered choices derived using the Gaussian copula with AR(1) correlation structure adheres to the idea presented in section 1.5.3 that choice probabilities are translation invariant in their utilities.

In Table 7, the choice probabilities according to (46) for three choices are calculated for three different sets of values of ρ and different values of the mean vector (μ_1, μ_2, μ_3) of the utilities. We have chosen (μ_1, μ_2, μ_3) to be $(2, 2.1, 2.5)$, $(0.4, 0.42, 0.5)$ and $(10, 10.5, 12.5)$. The second and third sets are scaled versions of the first and are obtained by multiplying 1/5 and 5 with the first set respectively. By doing this, we are just scaling of the mean component of the utility by a positive factor and not the actual utility.

We notice that in Table 7 also, all the triplets of choice probabilities have unity as

Table 7: Choice probabilities for three ordered choices with scaled means of utilities

ρ	$(\mu_1,\ \mu_2,\ \mu_3)$		
	(2, 2.1, 2.5)	(0.4, 0.42, 0.5)	(10, 10.5, 12.5)
0	0.27, 0.29, 0.44	0.32, 0.33, 0.35	0.07, 0.11, 0.82
0.1	0.26, 0.28, 0.46	0.32, 0.31, 0.37	0.08, 0.11, 0.81
0.2	0.26, 0.27, 0.47	0.32, 0.30, 0.38	0.08, 0.09, 0.83
0.3	0.27, 0.26, 0.47	0.33, 0.30, 0.37	0.07, 0.08, 0.85
0.4	0.27, 0.25, 0.48	0.33, 0.29, 0.38	0.06, 0.07, 0.87
0.5	0.27, 0.23, 0.50	0.34, 0.28, 0.38	0.06, 0.05, 0.89
0.6	0.26, 0.22, 0.52	0.34, 0.27, 0.39	0.04, 0.04, 0.92
0.7	0.25, 0.20, 0.55	0.34, 0.26, 0.40	0.03, 0.03, 0.94
0.8	0.24, 0.18, 0.58	0.34, 0.25, 0.41	0.02, 0.01, 0.97
0.9	0.19, 0.13, 0.68	0.33, 0.23, 0.44	0.00, 0.00, 1.00

their sums. Unlike Table 6, the probabilities in a row are not same. As we have scaled only the means in the utilities, the differences in the mean have changed and so have the choice probabilities. This proves the idea presented in section 1.5.4 that if we scale either the mean or the random component of the utility, the choice probability changes.

The additional property that this table showcases is that for some of the choices, for a fixed ρ, there is a change in the relative magnitude of the values of P_1, P_2 and P_3. For example, starting from $\rho = 0.1$ to $\rho = 0.9$, for the second set of means, $(\mu_1, \mu_2, \mu_3) = (0.4, 0.42, 0.5)$, the second choice has the lowest choice probability in contrast to the first and third sets of means where the first choice has the lowest probability. This reversal of probabilities can be attributed to the small differences between the three mean components for the second set and its interweaving with the correlation parameter ρ in the Gaussian copula with AR(1) structure and random components which are distributed as Gumbel variables.

3.3 CONDITIONAL MEAN AND VARIANCE FOR A MULTIVARIATE NORMAL DISTRIBUTION WITH AN AR(1) COVARIANCE STRUCTURE

Let $\boldsymbol{V} = (V_1, V_2, \ldots, V_c)'$ be a c-dimensional random vector which is distributed as $N(\boldsymbol{\mu}, \boldsymbol{R})$, where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_c)'$ and $\boldsymbol{R}$ has an AR(1) structure given in (40) with parameter ρ. Our goal is to find the conditional distribution of $(V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_c)'$ given V_j. In order to do this, we introduce the notation $\boldsymbol{V}_{-j} = (V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_c)'$. Accordingly, we denote the mean of $\boldsymbol{V}_{-j}$ by $\boldsymbol{\mu}_{-j}$. Swapping the jth row and jth column with the last row and last column we can write the AR(1) correlation matrix as

$$
\begin{aligned}
\boldsymbol{R} &= \left(\begin{array}{ccccccc|c}
1 & \rho & \ldots & \rho^{j-2} & \rho^{j} & \ldots & \rho^{c-1} & \rho^{j-1} \\
\rho & 1 & \ldots & \rho^{j-3} & \rho^{j-1} & \ldots & \rho^{c-2} & \rho^{j-2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\rho^{j-2} & \rho^{j-3} & \ldots & 1 & \rho^{2} & \ldots & \rho^{c-j+1} & \rho \\
\rho^{j} & \rho^{j-1} & \ldots & \rho^{2} & 1 & \ldots & \rho^{c-j-1} & \rho \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\rho^{c-1} & \rho^{c-2} & \ldots & \rho^{c-j+1} & \rho^{c-j-1} & \ldots & 1 & \rho^{c-j} \\
\hline
\rho^{j-1} & \rho^{j-2} & \ldots & \rho & \rho & \ldots & \rho^{c-j} & 1
\end{array}\right), \\
&= \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & 1 \end{pmatrix}, \text{ where} \\
\boldsymbol{R}_{11} &= \begin{pmatrix}
1 & \rho & \ldots & \rho^{j-2} & \rho^{j} & \ldots & \rho^{c-1} \\
\rho & 1 & \ldots & \rho^{j-3} & \rho^{j-1} & \ldots & \rho^{c-2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\rho^{j-2} & \rho^{j-3} & \ldots & 1 & \rho^{2} & \ldots & \rho^{c-j+1} \\
\rho^{j} & \rho^{j-1} & \ldots & \rho^{2} & 1 & \ldots & \rho^{c-j-1} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\rho^{c-1} & \rho^{c-2} & \ldots & \rho^{c-j+1} & \rho^{c-j-1} & \ldots & 1
\end{pmatrix}, \\
\boldsymbol{R}_{21} &= \left(\rho^{j-1}, \rho^{j-2}, \ldots \rho, \rho, \ldots \rho^{c-j}\right), \ \boldsymbol{R}_{12} = \boldsymbol{R}_{21}'.
\end{aligned}
$$

Note that

$$
\boldsymbol{R}_{12}\,\boldsymbol{R}_{21} = \begin{pmatrix} \rho^{j-1} \\ \rho^{j-2} \\ \vdots \\ \rho \\ \rho \\ \vdots \\ \rho^{c-j} \end{pmatrix} \begin{pmatrix} \rho^{j-1}, & \rho^{j-2}, & \dots & \rho, & \rho, & \dots & \rho^{c-j} \end{pmatrix}
$$

$$
= \begin{pmatrix} \rho^{2j-2} & \rho^{2j-3} & \dots & \rho^{j} & \rho^{j} & \dots & \rho^{c-1} \\ \rho^{2j-3} & \rho^{2j-4} & \dots & \rho^{j-1} & \rho^{j-1} & \dots & \rho^{c-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{j} & \rho^{j-1} & \dots & \rho^{2} & \rho^{2} & \dots & \rho^{c-j+1} \\ \rho^{j} & \rho^{j-1} & \dots & \rho^{2} & \rho^{2} & \dots & \rho^{c-j+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{c-1} & \rho^{c-2} & \dots & \rho^{c-j+1} & \rho^{c-j+1} & \dots & \rho^{2c-2j} \end{pmatrix}.
$$

From Theorem 2 stated in the Appendix, we know that $\boldsymbol{V}_{-j} \mid V_j = v_j$ follows a normal distribution with mean

$$
\begin{aligned}
\boldsymbol{\mu}_{-j|j} &= \boldsymbol{\mu}_{-j} + \boldsymbol{R}_{12}(v_j - \mu_j) \\
&= \begin{pmatrix} \mu_1 + \rho^{j-1}(v_j - \mu_j) \\ \vdots \\ \mu_{j-1} + \rho(v_j - \mu_j) \\ \mu_{j+1} + \rho(v_j - \mu_j) \\ \vdots \\ \mu_c + \rho^{c-j}(v_j - \mu_j) \end{pmatrix}.
\end{aligned} \tag{42}
$$

and covariance matrix

$$
\boldsymbol{R}_{-j|j} = \boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{21}
$$

$$= \begin{pmatrix} 1-\rho^{2j-2} & \dots & \rho^{j-2}-\rho^{j} & 0 & \dots & 0 \\ \rho-\rho^{2j-3} & \dots & \rho^{j-3}-\rho^{j-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{j-2}-\rho^{j} & \dots & 1-\rho^2 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1-\rho^2 & \dots & \rho^{c-j-1}-\rho^{c-j+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \rho^{c-j-1}-\rho^{c-j+1} & \dots & 1-\rho^{2c-2j} \end{pmatrix}. \tag{43}$$

It is interesting to note that the off diagonals elements of the matrix in (43) are zero. This establishes that given $V_j = v_j$, the two vectors $(V_1, \dots, V_{j-1})$ and $(V_{j+1}, \dots, V_c)$ are independent. Another view of this result is that when the correlation structure is AR(1), conditional on the present, the future is independent of the past. In the literature this is known as the Markov property.

3.3.1 ALTERNATIVE DERIVATION

An alternative way to establish the Markov property of the multivariate normal distribution with AR(1) correlation structure is to consider the trivariate normal distribution. Consider three components V_k, V_l, V_j, of $\boldsymbol{V}$, where $k < l$ and j is arbitrary. Note that the correlation matrix of (V_k, V_l, V_j) is given by

$$\boldsymbol{R}_3 = \left(\begin{array}{cc|c} 1 & \rho^{|k-l|} & \rho^{|k-j|} \\ \rho^{|k-l|} & 1 & \rho^{|j-l|} \\ \hline \rho^{|k-j|} & \rho^{|j-l|} & 1 \end{array}\right)$$

From Theorem 8 stated in the appendix we get the conditional covariance of V_k and V_l given V_j is given by

$$\sigma_{k,l|j} = \rho^{|k-l|} - \rho^{|j-k|+|j-l|} \tag{44}$$

This shows that

$$\sigma_{k,l|j} \;=\; \begin{cases} \rho^{l-k} - \rho^{2j-k-l}, & \text{if } k < l < j \\ 0, & \text{if } k < j < l \\ \rho^{l-k} - \rho^{k+l-2j}, & \text{if } j < k < l \end{cases} . \tag{45}$$

Thus for $k < j < l$, V_k and V_l are independent given V_j or the sequence V_k, V_j and V_l satisfy the Markov property. Another method to establish the Markov property is using partial correlation concept. The partial correlation $\rho_{k,l|j}$ is defined as the correlation between V_k and V_l given V_j and it is given by the formula

$$\rho_{k,l|j} = \frac{\rho_{k,l} - \rho_{k,j}\,\rho_{j,l}}{\sqrt{1-\rho_{k,j}^2}\sqrt{1-\rho_{j,l}^2}}.$$

Putting $\rho_{k,l} = \rho^{|k-l|}$, for $k < j < l$, we get

$$\rho_{k,l|j} = \frac{\rho^{l-k} - \rho^{j-k}\,\rho^{l-j}}{\sqrt{1-\rho^{2(j-k)}}\sqrt{1-\rho^{2(l-j)}}} = \frac{\rho^{l-k} - \rho^{l-k}}{\sqrt{1-\rho^{2(j-k)}}\sqrt{1-\rho^{2(l-j)}}} = 0,$$

which establishes the Markov property.

3.4 CHOICE PROBABILITY

Using (42) and (43), the choice probability of an ordered choice model using Gaussian copula with AR(1) correlation structure is given by

$$P_{ij} = \int_{-\infty}^{\infty} \Phi_{c-1}\left(v^*_{ij1}, \ldots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \ldots, v^*_{ijc}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right)\, \phi(v)\, dv$$

where $v^*_{ijk} = \Phi^{-1}\left(F(\mu_{ij} - \mu_{ik} + F^{-1}(\Phi(v)))\right)$ for $k \neq j$,

$\boldsymbol{\eta}^{(ij)} = v\left(\rho^{j-1}, \rho^{j-2}, \ldots, \rho, \rho, \rho^2, \ldots, \rho^{c-j}\right)'$ and

$\boldsymbol{\Sigma}^{(ij)} =$

$$\begin{pmatrix} 1-\rho^{2j-2} & \dots & \rho^{j-2}-\rho^{j} & 0 & \dots & 0 \\ \rho-\rho^{2j-3} & \dots & \rho^{j-3}-\rho^{j-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{j-2}-\rho^{j} & \dots & 1-\rho^{2} & 0 & \dots & 0 \\ 0 & \dots & 0 & 1-\rho^{2} & \dots & \rho^{c-j-1}-\rho^{c-j+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \rho^{c-j-1}-\rho^{c-j+1} & \dots & 1-\rho^{2c-2j} \end{pmatrix}. \tag{46}$$

3.5 DERIVATIVES

In this section, we will find the expressions for the derivatives of the choice probability in case of Gaussian copula with AR(1) structure. As before we assume $\mu_{ij} = \boldsymbol{X'}_{ik}\,\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_1.\beta_2, \dots, \beta_p)'$ is the regression vector. Now

$$\begin{aligned} &\frac{\partial P_{ij}}{\partial \beta_m} \\ = &\int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial \beta_m} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \right\} \phi(v)\, dv \\ = &\int_{-\infty}^{\infty} \left\{ \sum_{k=1,k\neq j}^{c} \Phi_{c-2}\left(\boldsymbol{V}^{(ij)}_{-(k)}; \boldsymbol{\eta}^{(ij)}_{-(j,k)}, \boldsymbol{\Sigma}^{(ij)}_{-(j,k)}\right) \phi(v^*_{ijk}; \rho^{|j-k|}v, (1-\rho^{2|j-k|})) \right. \\ &\left. \frac{f((\boldsymbol{X}_{ij}-\boldsymbol{X}_{ik})'\boldsymbol{\beta} + F^{-1}(\Phi(v)))}{\phi(\Phi^{-1}\left(F((\boldsymbol{X}_{ij}-\boldsymbol{X}_{ik})'\boldsymbol{\beta} + F^{-1}(\Phi(v)))\right))} (X_{ijm} - X_{ikm}) \right\} \phi(v)\, dv, \end{aligned} \tag{47}$$

where,

$$\boldsymbol{V}^{(ij)}_{-(k)} = \begin{cases} (v^*_{ij1}, \dots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \dots, v^*_{ij(k-1)}, v^*_{ij(k+1)}, \dots, v^*_{ijc}) & , \text{if } k > j \\ \\ (v^*_{ij1}, \dots, v^*_{ij(k-1)}, v^*_{ij(k+1)}, \dots, v^*_{ij(j-1)}, v^*_{ij(j+1)}, \dots, v^*_{ijc}) & , \text{if } k < j \end{cases}$$

Next we will find the derivative of the choice probability with respect to the correlation parameter ρ of the Gaussian copula.

$$\begin{aligned}\frac{\partial P_{ij}}{\partial \rho} &= \frac{\partial}{\partial \rho}\int_{-\infty}^{\infty} \Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \phi(v)\, dv \\ &= \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial \rho}\Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right)\right\} \phi(v)\, dv.\end{aligned}$$

The quantity $\Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right)$ is a cumulative distribution of a $c-1$ dimensional multivariate normal distribution with mean $\boldsymbol{\eta}^{(ij)}$ and variance $\boldsymbol{\Sigma}^{(ij)}$. It would be easier to find the derivative if we transform $\boldsymbol{\Sigma}^{(ij)}$ to a correlation matrix. Consider the transformation

$$w_{ijk} = \frac{v_{ijk} - \rho^{|j-k|}v}{\sqrt{1-\rho^{2|j-k|}}} \text{ for } k \neq j.$$

Then

$$\begin{aligned}&\frac{\partial}{\partial \rho}\Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right) \\ &= \frac{\partial}{\partial \rho}\Phi_{c-1}\left(w^*_{ij1}, \ldots, w^*_{ij(j-1)}, w^*_{ij(j+1)}, \ldots, w^*_{ijc}; \mathbf{0}, \boldsymbol{R}^{(ij)}\right) \qquad (48)\end{aligned}$$

where $\boldsymbol{R}^{(ij)} = \left(r_{k,l|j}\right)_{c-1\times c-1}$ and

$$r_{k,k|j} = 1,\; k \neq j,$$

$$r_{k,l|j} = \begin{cases} \frac{\rho^{l-k} - \rho^{2j-k-l}}{\sqrt{(1-\rho^{2|j-k|})(1-\rho^{2|j-l|})}}, & \text{if } k < l < j \\ 0, & \text{if } k < j < l \\ \frac{\rho^{l-k} - \rho^{k+l-2j}}{\sqrt{(1-\rho^{2|j-k|})(1-\rho^{2|j-l|})}}, & \text{if } j < k < l \end{cases}.$$

The derivative (48) can be calculated using the chain rule- by first taking the partial derivatives of the $\Phi_{c-1}(.)$ function with respect to w_{ijk}'s, and then by taking the derivative with respect to ρ. Let $\boldsymbol{W}^{(ij)} = (w_{ij1}, \ldots, w_{ij(j-1)}, w_{ij(j+1)}, \ldots, w_{ijc})$. Then

$$\frac{\partial}{\partial \rho}\Phi_{c-1}\left(\boldsymbol{V}^{(ij)}; \boldsymbol{\eta}^{(ij)}, \boldsymbol{\Sigma}^{(ij)}\right)$$

$$= \frac{\partial}{\partial \rho} \Phi_{c-1} \left(\boldsymbol{W}^{(ij)}; \boldsymbol{0}, \boldsymbol{R}^{(ij)} \right)$$

$$= \sum_{k=1, k \neq j}^{c} \left\{ \frac{\partial}{\partial w_{ijk}} \Phi_{c-1} \left(\boldsymbol{W}^{(ij)}; \boldsymbol{0}, \boldsymbol{R}^{(ij)} \right) \frac{\partial w_{ijk}}{\partial \rho} \right\}$$

$$+ \sum_{k=1}^{c} \sum_{l=k+1}^{c} \frac{\partial}{\partial r_{k,l|j}} \Phi_{c-1} \left(\boldsymbol{W}^{(ij)}; \boldsymbol{0}, \boldsymbol{R}^{(ij)} \right) \frac{\partial r_{k,l|j}}{\partial \rho}.$$

The partial derivatives involved are

$$\frac{\partial}{\partial w_{ijk}} \Phi_{c-1} \left(\boldsymbol{W}^{(ij)}; \boldsymbol{0}, \boldsymbol{R}^{(ij)} \right)$$

$$= \sum_{k=1, k \neq j}^{c} \Phi_{c-2} \left(\boldsymbol{W}^{(ij)}_{-(k)}; \boldsymbol{\eta}^{(ij)0}_{-(k)}, \boldsymbol{R}^{(ij)0}_{-(k)} \right) \phi(w_{ijk}; 0, 1),$$

where $\boldsymbol{\eta}^{(ij)0}_{-(k)}$ and $\boldsymbol{R}^{(ij)0}_{-(k)}$ are the conditional mean and covariance of $\boldsymbol{W}^{(ij)}_{-(k)}$ given w_{ijk}. Here $\boldsymbol{W}^{(ij)}_{-(k)}$ is the vector obtained by removing w_{ijk} from $\boldsymbol{W}^{(ij)}$. It is easy to check that

$$\frac{\partial w_{ijk}}{\partial \rho} = \frac{\alpha \rho^{\alpha-1} \left(\rho^{\alpha} (v_{ijk} - \rho^{\alpha} v) - v(1 - \rho^{2\alpha}) \right)}{(1 - \rho^{2\alpha})^{3/2}}, \quad \alpha = |j - k|, \tag{49}$$

and

$$\frac{\partial}{\partial r_{k,l|j}} \Phi_{c-1} \left(\boldsymbol{W}^{(ij)}; \boldsymbol{0}, \boldsymbol{R}^{(ij)} \right)$$

$$= \sum_{k=1}^{c} \sum_{l=k+1}^{c} \Phi_{c-3} \left(\boldsymbol{W}^{(ij)}_{-(kl)}; \boldsymbol{\eta}^{(ij)0}_{-(kl)} \boldsymbol{R}^{(ij)0}_{-(kl)} \right) \phi_2 \left(\boldsymbol{W}^{(ij)}_{(kl)}; \boldsymbol{0}, \boldsymbol{R}^{(ij)}_{(kl)} \right),$$

where $\boldsymbol{\eta}^{(ij)0}_{-(kl)}$ and $\boldsymbol{R}^{(ij)0}_{-(kl)}$ are the conditional mean and covariance of $\boldsymbol{W}^{(ij)}_{-(kl)}$ given $\boldsymbol{W}^{(ij)}_{(kl)}$. Here once again $\boldsymbol{W}^{(ij)}_{-(kl)}$ is obtained removing $W^{(ij)}_{(kl)} = (w_{ijk}, w_{ijl})$ from $\boldsymbol{W}^{(ij)}$. The derivatives of the elements of the conditional correlation matrix $\boldsymbol{R}^{ij}$ with respect to ρ are

$$\frac{\partial r_{k,l|j}}{\partial \rho} = \begin{cases} \frac{\left((\alpha-\gamma)\rho^{\alpha-\gamma-1}-(\alpha+\gamma)\rho^{\alpha+\gamma-1}\right)}{\left((1-\rho^{2|\alpha|})(1-\rho^{2|\gamma|})\right)} \\ -\frac{\left(\rho^{\alpha-\gamma}-\rho^{\alpha+\gamma}\right)\left((1-\rho^{2|\alpha|})^{2|\gamma|-1}+(1-\rho^{2|\gamma|})^{2|\alpha|-1}\right)}{\left((1-\rho^{2|\alpha|})(1-\rho^{2|\gamma|})\right)^{3/2}} & , \text{ if } j<k<l \\ 0 & , \text{ if } k<j<l \\ \frac{\left((\alpha-\gamma)\rho^{\alpha-\gamma-1}+(\alpha+\gamma)\rho^{-(\alpha+\gamma+1)}\right)}{\left((1-\rho^{2|\alpha|})(1-\rho^{2|\gamma|})\right)} \\ -\frac{\left(\rho^{\alpha-\gamma}-\rho^{-\alpha-\gamma}\right)\left((1-\rho^{2|\alpha|})^{2|\gamma|-1}+(1-\rho^{2|\gamma|})^{2|\alpha|-1}\right)}{\left((1-\rho^{2|\alpha|})(1-\rho^{2|\gamma|})\right)^{3/2}} & , \text{ if } k<l<j \end{cases}, \tag{50}$$

where $\alpha = j - k$ and $\gamma = j - l$. The first derivative follows from Theorem 4 and the third derivative follows from Theorem 5 given in the Appendix. In summary the derivative of the choice probability with respect to the correlation parameter ρ can be written as

$$\begin{aligned} \frac{\partial P_{ij}}{\partial \rho} &= \int_{-\infty}^{\infty} \sum_{k=1,k\neq j}^{c} \Phi_{c-2}\left(\boldsymbol{W}_{-(k)}^{(ij)}; \boldsymbol{\eta}_{-(k)}^{(ij)0}, \boldsymbol{R}_{-(k)}^{(ij)0}\right) \phi(w_{ijk};0,1)\frac{\partial w_{ijk}}{\partial \rho}\, dv \\ &+ \int_{-\infty}^{\infty} \sum_{k=1}^{c} \sum_{l=k+1}^{c} \Phi_{c-3}\left(\boldsymbol{W}_{-(kl)}^{(ij)}; \boldsymbol{\eta}_{-(kl)}^{(ij)0} \boldsymbol{R}_{-(kl)}^{(ij)0}\right) \phi_2\left(\boldsymbol{W}_{(kl)}^{(ij)}; \boldsymbol{0}, \boldsymbol{R}_{(kl)}^{(ij)}\right) \\ &\qquad \frac{\partial r_{k,l|j}}{\partial \rho}\, dv, \end{aligned} \tag{51}$$

where the expressions for $\partial w_{ijk}/\partial \rho$ and $\partial r_{k,l|j}/\partial \rho$ are in (49) and (50) respectively.

3.6 MODEL FITTING

In this section we will use simulated data to illustrate the use of the multivariate discrete choice Gumbel model with AR(1) correlation structure (MDCG-AR(1)). Our simulated data set consists of $n = 300$ consumers and a choice set with $c = 3$ choices. First we generate the random components (errors) from a standard trivariate normal

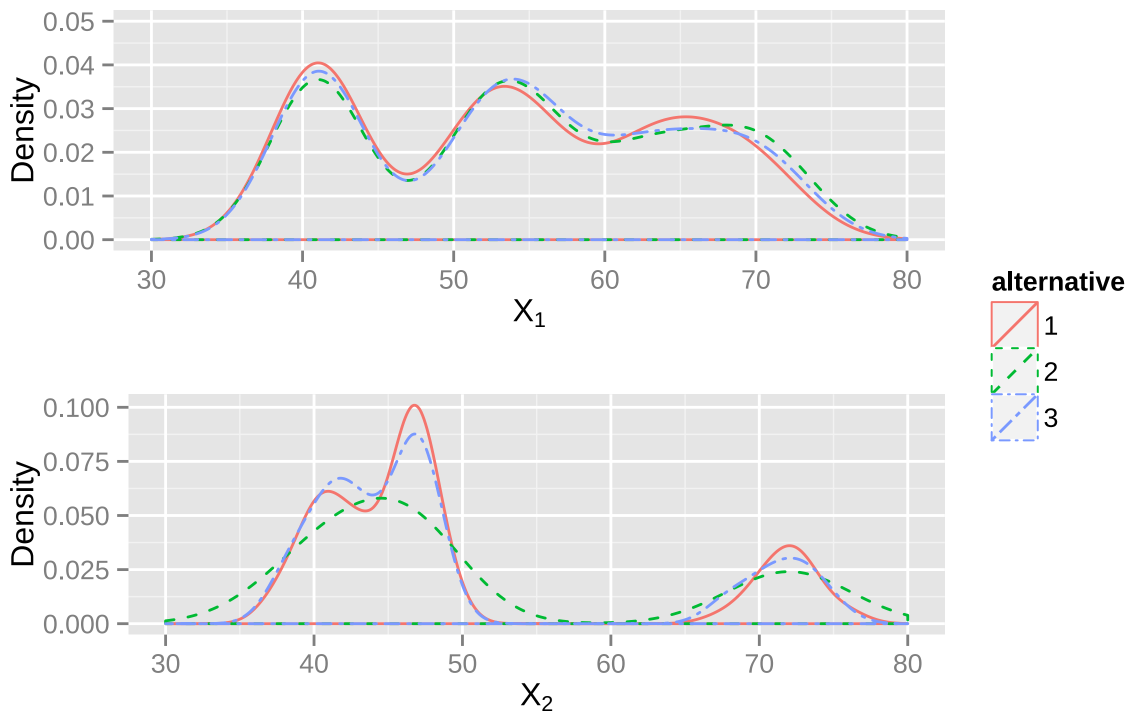

Figure 6: Density plots of the simulated covariates by chosen alternatives

distribution with mean $\boldsymbol{\mu} = (0, 0, 0)$ and AR(1) correlation matrix with $rho = 0.7$,

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0.7 & 0.49 \\ 0.7 & 1 & 0.7 \\ 0.49 & 0.7 & 1 \end{pmatrix}.$$

This means that if z_{ij} is the random component for the ith consumer and the jth choice, $i = 1, 2, \ldots, 300$ and $j = 1, 2, 3$, then

$$(z_{i1}, z_{i2}, z_{i3}) \sim N_3(\boldsymbol{\mu}, \boldsymbol{R}), \text{ for all } i.$$

We generate $p = 2$ covariates namely X_1 and X_2, both of them being simulated from mixture distributions (see McLachlan (2000)). We choose X_1 to be a mixture of three uniform distributions with parameters $(40, 42)$, $(50, 57)$ and $(60, 73)$ respectively, where the mixing probabilities are 0.3, 0.35 and 0.35. Similarly, X_2 is chosen to be a mixture of three univariate normal distributions with means and variances

$(41, 4)$, $(47, 1)$ and $(72, 4)$ and the mixing probabilities are 0.4, 0.35 and 0.25 respectively. The density plots of the two mixture distributions are shown in Figure 6. As each of them come from mixture distributions with three distant distributions in each mix, we can see that both the density plots of have three peaks. Since there are three choices, we also introduce two intercepts as covariates, namely Int_1 and Int_2 where

$$Int_1 = \begin{cases} 1, & \text{if alternative} = 1 \\ 0, & \text{if otherwise} \end{cases} \quad \text{and} \quad Int_2 = \begin{cases} 1, & \text{if alternative} = 2 \\ 0, & \text{if otherwise} \end{cases}.$$

Therefore, based on (3) and (6), the utility for the ith consumer and the jth alternative, say U_{ij}, can be calculated as

$$U_{ij} = \beta_{01} Int_1 + \beta_{02} Int_2 + \beta_1 X_1 + \beta_2 X_2 + z_{ij}.$$

For the purpose of the exercise of fitting the choice probability model for ordered choices using the Gaussian copula given in (46), we chose $\beta_{01} = 1.4, \beta_{02} = 2, \beta_1 = 1$ and $\beta_2 = 1.5$. Based on these values and the simulated data, we calculate the utility U_{ij} and hence the choice variable Y_{ij} for the ith consumer as

$$Y_{ij} = \begin{cases} 1 & \text{if } \ U_{ik} < U_{ij},\ k = 1, 2, \ldots, c,\ \forall\, k \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

An empirical summary of the simulated data shows 98 or 32.67% consumers chose alternative 1, 109 or 36.33% consumers chose alternative 2 and 93 or 31% consumers chose alternative 3. We fit the MDCG-AR(1) and CNL models for the simulated data using the maximum likelihood. Estimates and standard errors for the CNL model are obtained by using the "Proc MDC" in the SAS software. The estimates for the MDCG-Ar(1) model are obtained by maximizing the log-likelihood function in R. Standard errors for the were obtained using 50 bootstrap samples. The R programs that were developed are provided in the Appendix. The results are displayed in Table 8. It is to be noted that, the p-values are tests comparing the likelihood estimates with the values of the parameters that were used to simulate the data. For example, the estimate of the parameter β_{01}, associated with the intercept for alternative 1, is compared against the value 1.4. Similarly, estimates of $\beta_{02}, \beta_1, \beta_2$ and ρ are compared

Table 8: Analysis of simulated data of ordered choices

Variable	MDCG AR(1)			CNL		
	Estimate	SD	P Value	Estimate	SD	P Value
Int 1	1.3655	0.1945	0.23	2.0603	0.6205	0.2875
2	1.8767	0.0995	< 0.001	3.3611	0.7577	0.0728
3	0	—	—	0	—	—
X_1	0.964	0.1464	0.0985	3.8311	0.6754	< 0.001
X_2	1.4321	0.1628	0.0063	-0.6874	0.1295	< 0.001
ρ	0.738	0.1406	0.0702	—	—	—
AIC	27.32			74.90		
$\boldsymbol{R^2_M}$	0.9735			0.898		
$\boldsymbol{R^2_{M,Adj}}$	0.9583			0.8858		

against the numbers $2, 1, 1.5$ and 0.7 respectively. High p-values indicate acceptance of the null hypothesis. Please note that the hypothesis of $\rho = 0.7$ is accepted for the MDCG-AR(1) model. High values of McFadden's R^2 and adjusted R^2 are expected since the data is simulated from the model that we are fitting. The three goodness of fit statistics described in Section 1.6 show that the MDCG AR(1) model is a better fit to the data than the CNL model.

CHAPTER 4

SUMMARY

We have studied discrete choices models in this dissertation. In all walks of life, from housing, transportation, health care and grocery shopping, consumers face many choices or products and have to make decisions on selecting a choice or picking a product. Discrete choice models were introduced by econometricians and statisticians to aid in understanding the consumers' choice preferences. These models are based on the fundamental assumption that the consumers assign utilities to the choices and select the choice or product that maximizes their utility.

A popular and widely used discrete choice model is the conditional logit model introduced by Luce (1959). This model was brought into limelight by McFadden (1974) who laid the mathematical foundation, elucidated and showed practical applications of the model. The conditional logit model assumes that the unobserved utility for a choice is the sum of two components, a deterministic and a random component. The model assumes that the random components are independent and follow a Gumbel distribution. A major advantage of this model is that the probability a consumer selects a particular choice, known as the choice probability, has a closed form expression. However in practice the independence assumption of the random components is unreasonable and a better model should account for the dependence or correlation present among the choices.

In this dissertation we generalized McFadden's conditional logit model to account for the correlation between the choices. We have accomplished this objective using the Gaussian copula to construct a joint distribution for the random components. In Chapter 2, we studied a parsimonious model where we assume the correlation matrix of the Gaussian copula is equicorrelated, which is determined by a single parameter. There are examples where this assumption of equal correlation between the choices is reasonable, especially for choices that are nominal in nature. We derived an expression for the choice probabilities, and studied their behavior as a function of the correlation parameter. We obtained analytical expressions for the gradient vector of the choice probabilities and used them to develop R code for maximum likelihood estimation of the parameters. Using a real life data we showed that our model is comparable to existing models such as the conditional logit model.

In Chapter 3, our focus was on the situation where there is a natural order present among the choices, that is, the choices are ordinal in nature. An appropriate model for the ordinal choices that we studied in this chapter involves the AR(1) correlation structure. We showed that the multivariate normal distribution with AR(1) correlation structure has the property that the past and future are independent given the present. This property was used to derive simpler expressions for the choice probability and its gradient. We used the well known Plackett's formula to obtain computationally easier forms of the score equations. We developed another R program to implement this model and illustrated on a simulated data. Due to the difficulty in deriving analytical expressions for the Hessian matrix, we used bootstrap method to estimate standard errors for both equicorrelated and AR(1) models.

Future work will focus on developing faster, more efficient R code and more accurate estimation of the standard errors of the parameter estimates.

REFERENCES

Agresti, A. (2002), *Categorical Data Analysis, Second Edition*, New Jersey: John Wiley & Sons Inc..

Agresti, A., and Natarajan, R. (2001), "Modeling Clustered Ordered Categorical Data: A Survey," *International Statistical Review*, 69, 345-371.

Aitchison, J., and Silvey, S. (1957), "The Generalization of Probit Analysis to the Case of Multiple Responses," *Biometrika* 44:131-140.

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," *Second International Symposium on Information Theory*, Budapest: Akadmiai Kiado, pp. 267-281.

Ashford, J. R. and Sowden, R. R. (1970), "Multi-variate Probit Analysis," *Biometrics*, 26, 535-546.

Bhat, C.(1995), "A Heteroscedastic Extreme Value Model of Intercity Mode Choice," *Transportation Research* B 29, 471-483.

Daganzo, C. (1979), *Multinomial Probit: The Theory and its Application to Demand Forecasting*, New York: Academic Press.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics* (Vol. 1), 71-126.

Greene, W. H. (2003), *Econometric Analysis* (5th Ed.), Upper Saddle River, NJ: Printice Hall.

Hausman, J. and Wise, D. (1978), "A Conditional Probit Model For Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preference" *Econometrica*, 48, 403-429.

Johnson, N.L., Kotz, S., Balakrishnan, N. (2000), *Continuous Univariate Distributions, Volume 2, Second Edition*, New York: John Wiley and Sons Inc..

Luce, D. (1959), *Individual Choice Behavior*, New York: John Wiley and Sons.

Luce, D. and P. Suppes (1965), “Preferences, Utility and Subjective Probability”, eds. R. Luce, R. Bush, and E. Galander, *Handbook of Mathematical Psychology*, New York : John Wiley and Sons, 249-410.

Kotz, S., Balakrishnan, N., Johnson, N.L. (2000), *Continuous Multivariate Distributions, Volume 1: Models and Applications* (2nd Ed.), New York: John Wiley and Sons Inc..

Marschak, J. (1960), “Binary Choice Constraints on Random Utility Indications,” *Stanford Symposium on Mathematical Methods in the Social Sciences*, ed. K. Arrow, Stanford, CA: Stanford University Press, 312-329.

McCullagh, P. (1980), “Regression Models for Ordinal Data,” *Journal of the Royal Statistical Society*, Ser. B, 42, 2, 109-142.

McFadden, D. (1974), “Conditional Logit Analysis of Qualitative Choice Behavior,” *Frontiers in Econometrics*, ed. P. Zambreka , New York: Academic Press, 105-142.

McFadden, D. (1978), “Modeling the Choice of Residential Location,” *Spatial Interaction Theory and Planning Models*, eds. A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, North-Holland, Amsterdam, pp. 75-96.

McFadden, D. (1987), “Regression-based Specification Tests for the Multinomial Logit Model,” *Journal of Econometrics*, 34, 63-82.

McFadden, D. (2001), “Economic Choices,” *American Economic Review*, 91, 351-378.

McLachlan, G., Peel, D. (2000), *Finite Mixture Models*, Wiley Series in Probability and Mathematical Statistics, Canada: John Wiley & Sons.

Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer Series in Statistics, New York: Springer.

Plackett, R. (1954), “A Reduction Formula for Normal Multivariate Integrals,” *Biometrika*, 41, 3/4, 351-360.

Small, K. (1987), “A Discrete Choice Model for Ordered Alternatives,” *Econometrica*, 55, 409-424.

Tong, Y. L.(2012), *The Multivariate Normal Distribution*, Springer Science & Business Media, 2012.

Thurstone, L. L. (1927), “A Law of Comparative Judgment,” *Psychological Review*, 34(4), 273-286.

Train, K. (2004), *Discrete Choice Models with Simulations* (2nd Ed.), Cambridge University Press.

APPENDIX A

RESULTS ON THE MULTIVARIATE NORMAL DISTRIBUTION FUNCTION

A.1 DERIVATIVE OF MULTIVARIATE NORMAL CDF WITH RESPECT TO IT'S ARGUMENTS

We will use the well known property, stated here for completeness, of the multivariate normal distribution.

Theorem **2.** Let $\mathbf{X}$ be t dimensional vector that follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. If $\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ are partitioned as follows

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{X}_1$, $\boldsymbol{\mu}_1$ and Σ_{11} are of dimension $s < t$, then

$$\mathbf{X}_1 \mid \mathbf{X_2} = \mathbf{x_2} \sim N(\boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}). \tag{52}$$

Let $\Phi_t(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the cumulative distribution function of $\boldsymbol{X}$. Denote by $\boldsymbol{X}_{-k}$ and $\boldsymbol{\mu}_{-k}$, the vectors $\mathbf{X}$ and $\boldsymbol{\mu}$ after deleting the k component respectively. Let us permute and partition $\boldsymbol{\Sigma}$ as follows

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{(k)} & \boldsymbol{\Sigma}_{12}^{(k)} \\ \boldsymbol{\Sigma}_{21}^{(k)} & \sigma_{kk} \end{pmatrix}.$$

Note that $\boldsymbol{\Sigma}_{11}^{(k)}$ is the covariance matrix of $\boldsymbol{X}_{-k}$, σ_{kk} is the variance of X_k, the kth component of $\boldsymbol{X}$, and $\boldsymbol{\Sigma}_{21}^{(k)}$ is the covariance between $\boldsymbol{X}_{-k}$ and X_k. The next theorem gives a formula for the derivative of the multivariate normal distribution with respect to one argument.

Theorem **3.** *Suppose* $\boldsymbol{\mu}_{-k|k} = \boldsymbol{\mu}_{-k} + \boldsymbol{\Sigma}_{11}^{(k)} \sigma_{kk}^{-1}(x_k - \mu_k)$ *and* $\boldsymbol{\Sigma}_{-k|k} = \boldsymbol{\Sigma}_{11}^{(k)} - \boldsymbol{\Sigma}_{12}^{(k)} \sigma_{kk}^{-1} \boldsymbol{\Sigma}_{21}^{(k)}$ *denote the conditional mean and variance of* $\boldsymbol{X}_{-k}$ *given* $X_k = x_k$ *respectively, then*

$$\frac{\partial}{\partial x_k} \Phi_t(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi(x_k; \mu_k, \sigma_{kk})\ \Phi_{t-1}(\boldsymbol{x}_{-k}; \boldsymbol{\mu}_{-k|k}, \boldsymbol{\Sigma}_{-k|k}).$$

Proof.

$$\begin{aligned}
&\frac{\partial}{\partial x_k} \Phi_t(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \int_{-\infty}^{x_t} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \frac{\partial}{\partial x_k} \phi_t(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) dz_1\, dz_2\, \ldots dz_t \\
&= \int_{-\infty}^{x_t} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \frac{\partial}{\partial x_k} \phi(z_k; \mu_k, \sigma_{kk}) \\
&\qquad\qquad \phi_{t-1}(\boldsymbol{z}_{-k}; \boldsymbol{\mu}_{-k} + \boldsymbol{\Sigma}_{11}^{(k)} \sigma_{kk}^{-1}(z_k - \mu_k), \boldsymbol{\Sigma}_{-k|k}) dz_1\, dz_2\, \ldots dz_t \\
&= \int_{-\infty}^{x_k} \frac{\partial}{\partial x_k} \int_{-\infty}^{\boldsymbol{x}_{-k}} \phi_{t-1}(\boldsymbol{z}_{-k}; \boldsymbol{\mu}_{-k} + \boldsymbol{\Sigma}_{11}^{(k)} \sigma_{kk}^{-1}(z_k - \mu_k), \boldsymbol{\Sigma}_{-k|k})\, \boldsymbol{dz_{-k}}\, \phi(z_k; \mu_k, \sigma_{kk})\, dz_k,
\end{aligned}$$

$$\text{where } \int_{-\infty}^{\boldsymbol{x}_{-k}} = \int_{-\infty}^{x_t} \cdots \int_{-\infty}^{x_{k+1}} \int_{-\infty}^{x_{t-1}} \cdots \int_{-\infty}^{x_1} \text{ and}$$

$$\boldsymbol{dz_{-k}} = dz_t \ldots dz_{k+1} dz_{k-1} \ldots dz_1.$$

$$\begin{aligned}
&\frac{\partial}{\partial x_k} \Phi_t(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \int_{-\infty}^{x_k} \frac{\partial}{\partial x_k} \Phi_{t-1}(\boldsymbol{z}_{-k}; \boldsymbol{\mu}_{-k} + \boldsymbol{\Sigma}_{11}^{(k)} \sigma_{kk}^{-1}(z_k - \mu_k), \boldsymbol{\Sigma}_{-k|k})\, \boldsymbol{dz_{-k}}\ \ \phi(z_k; \mu_k, \sigma_{kk})\, dz_k \\
&= \Phi_{t-1}(\boldsymbol{x}_{-k}; \boldsymbol{\mu}_{-k} + \boldsymbol{\Sigma}_{11}^{(k)} \sigma_{kk}^{-1}(x_k - \mu_k), \boldsymbol{\Sigma}_{-k|k})\ \phi(x_k; \mu_k, \sigma_{kk}) \\
&= \phi(x_k; \mu_k, \sigma_{kk})\ \Phi_{t-1}(\boldsymbol{x}_{-k}; \boldsymbol{\mu}_{-k|k}, \boldsymbol{\Sigma}_{-k|k}).
\end{aligned}$$

A generalization of Theorem 3 is given next.

Theorem **4.** Let $\boldsymbol{X}_1$ of dimension s and $\boldsymbol{X}_2$ be of dimension of $t-s$, for $s<t$, be a partition of $\boldsymbol{X}$. Then,

$$\frac{\partial^s}{\partial x_1 \partial x_2 \dots \partial x_s} \Phi_t(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \phi_s(\boldsymbol{x}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\,\Phi_{t-s}(\boldsymbol{x}_2;\boldsymbol{\mu}_{2|1}(\boldsymbol{x}_1),\boldsymbol{\Sigma}_{2|1}),$$

where $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}$ are the mean and covariance of $\boldsymbol{X}_1$ respectively, and $\boldsymbol{\mu}_{2|1}(\boldsymbol{x}_1), \boldsymbol{\Sigma}_{2|1}$ are the mean and covariance of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$.

Proof. In accordance of the dimensions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, let us partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ into $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Note that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are also then normally distributed with parameters $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. The covariance between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is given by the matrix $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^{/}$.

The conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{z}_1$ is $N(\boldsymbol{\mu}_{2|1}, \boldsymbol{\Sigma}_{2|1})$ where $\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma_{21}\Sigma_{11}^{-1}(z_1 - \mu_1)}$ and $\boldsymbol{\Sigma}_{2|1} = \boldsymbol{\Sigma}_{22} + \boldsymbol{\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}$. Clearly $\boldsymbol{\mu}_{2|1}$ is a function of $\boldsymbol{z}_1$ and henceforth we will refer to it as $\boldsymbol{\mu_{2|1}(z_1)}$. We denote the t-dimensional CDF and PDF of a normal distribution as $\Phi_t()$ and $\phi_t()$ respectively.

Our goal is to derive an expression for

$$\frac{\partial^s}{\partial x_1 \partial x_2 \dots \partial x_s} \Phi_t(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}).$$

We know that,

$$\Phi_t(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_{-\infty}^{x_t} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \phi_t(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma})\, dz_1 \dots dz_{t-1}\, dz_t,$$

where $\phi_t(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma})$ can be written as

$$\phi_t(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \phi_s(\boldsymbol{z}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\,\phi_{t-s}(\boldsymbol{z}_2;\boldsymbol{\mu}_{2|1}(\boldsymbol{z}_1),\boldsymbol{\Sigma}_{2|1}).$$

Therefore,

$$
\begin{aligned}
&\Phi_t(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\\
&= \int_{-\infty}^{x_t}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\phi_s(\boldsymbol{z}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\,\phi_{t-s}(\boldsymbol{z}_2;\boldsymbol{\mu}_{2|1}(\boldsymbol{z}_1),\boldsymbol{\Sigma}_{2|1})\,dz_1\,dz_2\ldots dz_t\\
&= \int_{-\infty}^{x_s}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\phi_s(\boldsymbol{z}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\left\{\int_{-\infty}^{x_t}\cdots\int_{-\infty}^{x_{s+2}}\int_{-\infty}^{x_{s+1}}\phi_{t-s}(\boldsymbol{z}_2;\boldsymbol{\mu}_{2|1}(\boldsymbol{z}_1),\boldsymbol{\Sigma}_{2|1})\right.\\
&= \qquad\qquad\qquad\qquad \left. dz_{s+1}\,dz_{s+2}\ldots dz_t\right\}dz_1\,dz_2\ldots dz_s\\
&= \int_{-\infty}^{x_s}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\phi_s(\boldsymbol{z}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\,\Phi_{t-s}(\boldsymbol{x}_2;\boldsymbol{\mu}_{2|1}(\boldsymbol{z}_1),\boldsymbol{\Sigma}_{2|1})\,dz_1\,dz_2\ldots dz_s.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\frac{\partial^s}{\partial x_1\partial x_2\ldots\partial x_s}\Phi_t(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\\
&= \frac{\partial^s}{\partial x_1\partial x_2\ldots\partial x_s}\int_{-\infty}^{x_s}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\phi_s(\boldsymbol{z}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\\
&\qquad\qquad\qquad\qquad \Phi_{t-s}(\boldsymbol{x}_2;\boldsymbol{\mu}_{2|1}(\boldsymbol{z}_1),\boldsymbol{\Sigma}_{2|1})\,dz_1\,dz_2\ldots dz_s\\
&= \phi_s(\boldsymbol{x}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\,\Phi_{t-s}(\boldsymbol{x}_2-\boldsymbol{\mu}_{2|1}(\boldsymbol{x}_1);\boldsymbol{0},\boldsymbol{\Sigma}_{2|1})\\
&= \phi_s(\boldsymbol{x}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\,\Phi_{t-s}(\boldsymbol{x}_2;\boldsymbol{\mu}_{2|1}(\boldsymbol{x}_1),\boldsymbol{\Sigma}_{2|1}).
\end{aligned}
$$

A.2 DERIVATIVE OF MULTIVARIATE NORMAL CDF WITH RESPECT TO CORRELATION COEFFICIENTS

Let $\mathbf{X}$ be t dimensional vector that follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{R}$, where $\boldsymbol{R} = (r_{ij})_{t\times t}$, r_{ij} are functions of ρ for $i \neq j$ and $r_{ij} = 1$ for $i = j$. We define $\boldsymbol{X}_{(ij)} = (X_i, X_j)$ and $\boldsymbol{X}_{-(ij)}$ as the vector obtained by removing X_i and X_j from $\boldsymbol{X}$. The next theorem gives the derivative of the multivariate normal distribution with respect to the correlation parameter.

Theorem **5.** If $\boldsymbol{\Sigma}_{(ij)}$ is the covariance of $\boldsymbol{X}_{(ij)}$ and $\boldsymbol{\mu}_{-(ij)|(ij)}$ and $\boldsymbol{\Sigma}_{-(ij)|(ij)}$ are the mean and covariance of the conditional distribution of $\boldsymbol{X}_{-(ij)}$ given $\boldsymbol{X}_{(ij)} = \boldsymbol{x}_{(ij)}$ respectively, then

$$\frac{\partial}{\partial \rho}\Phi_t(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R}) = \sum_{i=1}^{t}\sum_{j=i+1}^{t} \Phi_{t-2}\left(\boldsymbol{x}_{-(ij)};\boldsymbol{\mu}_{-(ij)|(ij)},\boldsymbol{\Sigma}_{-(ij)|(ij)}\right)\phi_2(\boldsymbol{x}_{(ij)},\boldsymbol{0},\boldsymbol{\Sigma}_{(ij)})\frac{\partial r_{ij}}{\partial \rho}.$$

Proof. We will start with the derivative of the multivariate normal CDF $\Phi_t(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R})$ with respect to ρ. Without loss of generality, let's assume $i < j$. Then,

$$\begin{aligned}\frac{\partial}{\partial r_{ij}}\Phi_t(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R}) &= \frac{\partial}{\partial r_{ij}}\int_{-\infty}^{x_t}\dots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\phi_t(\boldsymbol{z};\boldsymbol{0},\boldsymbol{R})\,dz_1\dots dz_{t-1}\,dz_t \\ &= \int_{-\infty}^{x_t}\dots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\frac{\partial}{\partial r_{ij}}\phi_t(\boldsymbol{z};\boldsymbol{0},\boldsymbol{R})\,dz_1\dots dz_{t-1}\,dz_t.\end{aligned}$$

Plackett (1954) proved that

$$\frac{\partial}{\partial r_{ij}}\phi_t(\boldsymbol{z};\boldsymbol{0},\boldsymbol{R}) = \frac{\partial^2}{\partial z_i z_j}\phi_t(\boldsymbol{z};\boldsymbol{0},\boldsymbol{R}).$$

Using this,

$$\begin{aligned}\frac{\partial}{\partial r_{ij}}\Phi_t(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R}) &= \int_{-\infty}^{x_t}\dots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\frac{\partial^2}{\partial z_i z_j}\phi_t(\boldsymbol{z};\boldsymbol{0},\boldsymbol{R})\,dz_1\dots dz_{t-1}\,dz_t \\ &= \int_{-\infty}^{x_t}\dots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\frac{\partial^2}{\partial z_i z_j}\Big\{\Phi_{t-2}\left(\boldsymbol{z}_{-(ij)};\boldsymbol{\mu}^z_{-(ij)|(ij)},\boldsymbol{\Sigma}_{-(ij)|(ij)}\right) \\ &\quad \phi_2(\boldsymbol{z}_{(ij)},\boldsymbol{0},\boldsymbol{\Sigma}_{(ij)})\Big\}\,dz_1\dots dz_{t-1}\,dz_t \\ &= \int_{-\infty}^{x_j}\int_{-\infty}^{x_i}\Big\{\frac{\partial^2}{\partial z_i z_j}\int_{-\boldsymbol{\infty}}^{\boldsymbol{x}_{-(ij)}}\Phi_{t-2}\left(\boldsymbol{z}_{-(ij)};\boldsymbol{\mu}^z_{-(ij)|(ij)},\boldsymbol{\Sigma}_{-(ij)|(ij)}\right)\boldsymbol{dz}_{-(ij)} \\ &\quad \phi_2(\boldsymbol{z}_{(ij)},\boldsymbol{0},\boldsymbol{\Sigma}_{(ij)})\Big\}dz_i dz_j\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{x_j}\int_{-\infty}^{x_i}\left\{\frac{\partial^2}{\partial z_i z_j}\phi_2(\boldsymbol{z}_{(ij)},\boldsymbol{0},\boldsymbol{\Sigma}_{(ij)})\right.\\
&\qquad\qquad \left.\Phi_{t-2}\left(\boldsymbol{x}_{-(ij)};\boldsymbol{\mu}^z_{-(ij)|(ij)},\boldsymbol{\Sigma}_{-(ij)|(ij)}\right)\boldsymbol{dz}_{-(ij)}\right\}dz_i dz_j\\
&= \Phi_{t-2}\left(\boldsymbol{x}_{-(ij)};\boldsymbol{\mu}_{-(ij)|(ij)},\boldsymbol{\Sigma}_{-(ij)|(ij)}\right)\phi_2(\boldsymbol{x}_{(ij)},\boldsymbol{0},\boldsymbol{\Sigma}_{(ij)}),\ \text{where}
\end{aligned}
$$

$$
\int_{-\infty}^{\boldsymbol{x}_{-(ij)}} = \int_{-\infty}^{x_t}\cdots\int_{-\infty}^{x_{j+1}}\int_{-\infty}^{x_{j-1}}\cdots\int_{-\infty}^{x_{i+1}}\int_{-\infty}^{x_{i-1}}\cdots\int_{-\infty}^{x_1}\ \text{and}
$$

$$
\boldsymbol{dz}_{-(ij)} = dz_t\ldots dz_{j+1}dz_{j-1}\ldots dz_{i+1}dz_{i-1}\ldots dz_1.
$$

As all the r_{ij}s are functions of ρ, to evaluate the derivative of $\Phi_t(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R})$ we have to use the chain rule of differentiation for all r_{ij}s, $i<j$. Hence,

$$
\begin{aligned}
\frac{\partial}{\partial\rho}\Phi_t(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R}) &= \sum_{i=1}^{c}\sum_{j=i+1}^{c}\frac{\partial}{\partial r_{ij}}\Phi_t(\boldsymbol{x};\boldsymbol{0},\boldsymbol{R})\frac{\partial r_{ij}}{\partial\rho}\\
&= \sum_{i=1}^{c}\sum_{j=i+1}^{c}\Phi_{t-2}\left(\boldsymbol{x}_{-(ij)};\boldsymbol{\mu}_{-(ij)|(ij)},\boldsymbol{\Sigma}_{-(ij)|(ij)}\right)\phi_2(\boldsymbol{x}_{(ij)},\boldsymbol{0},\boldsymbol{\Sigma}_{(ij)})\frac{\partial r_{ij}}{\partial\rho}.
\end{aligned}
$$

We now consider a more general case. Let $\mathbf{X}$ be t dimensional vector that follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{D} = diag(\boldsymbol{\Sigma}) = (\sigma_{ii})$ be the diagonal matrix of variances and $\boldsymbol{R} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{D}^{-\frac{1}{2}}$ be the correlation matrix.

Theorem **6.** Let ρ_{ij} be the (i,j)th element of $\boldsymbol{R}$. The derivative of the multivariate normal distribution function $\boldsymbol{\Phi}_t$ with respect to ρ_{ij} is given by

$$
\frac{\partial}{\partial\rho_{ij}}\Phi_t(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{\sigma_{ii}\sigma_{jj}}}\phi_2(\boldsymbol{y}_{(ij)};\boldsymbol{0},\boldsymbol{R}_{ij})\Phi_{t-2}(\boldsymbol{y}_{-(ij)};\boldsymbol{\mu}_{-(ij)|-(ij)},\boldsymbol{R}_{-(ij)|-(ij)}),
$$

Proof.

$$\begin{aligned}
&\frac{\partial}{\partial \rho_{ij}} \Phi_t(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \frac{\partial}{\partial \rho_{ij}} \int_{-\infty}^{x_t} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \phi_t(\boldsymbol{w}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) dw_1\, dw_2\, \ldots dw_t \\
&= \frac{\partial}{\partial \rho_{ij}} \int_{-\infty}^{x_t} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} |\boldsymbol{D}^{-\frac{1}{2}}| \phi_t(\boldsymbol{D}^{\frac{1}{2}}(\boldsymbol{w} - \boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{R}) dw_1\, dw_2\, \ldots dw_t.
\end{aligned}$$

We change the vector of integration from $\boldsymbol{w}$ to $\boldsymbol{z}$, where $\boldsymbol{z} = \boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{w} - \boldsymbol{\mu})$. Clearly, $\boldsymbol{z}$ is also a Gaussian vector with zero mean and variance matrix $\boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{D}^{-\frac{1}{2}} = \boldsymbol{R}$, and the Jacobian of this transformation is $|\boldsymbol{D}^{\frac{1}{2}}|$. Also, the upper limit of the integral changes from $\boldsymbol{x}$ to $\boldsymbol{y} = \boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{\mu})$. So, the above integral can be written as

$$\begin{aligned}
&\frac{\partial}{\partial \rho_{ij}} \Phi_t(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \frac{\partial}{\partial \rho_{ij}} \int_{-\infty}^{y_t} \cdots \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} |\boldsymbol{D}^{-\frac{1}{2}}| \phi_t(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{R}) |\boldsymbol{D}^{\frac{1}{2}}|\, dz_1\, dz_2\, \ldots dz_t \\
&= \int_{-\infty}^{y_t} \cdots \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} \frac{\partial}{\partial \rho_{ij}} \phi_t(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{R})\, dz_1\, dz_2\, \ldots dz_t.
\end{aligned}$$

Plackett(1954) proved that

$$\frac{\partial}{\partial \rho_{ij}} \phi_t(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{R}) = \frac{\partial}{\partial z_l z_s} \phi_t(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{R}).$$

So,

$$\begin{aligned}
&\frac{\partial}{\partial \rho_{ij}} \Phi_t(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \int_{-\infty}^{y_t} \cdots \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} \frac{\partial}{\partial z_l z_s} \phi_t(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{R})\, dz_1\, dz_2\, \ldots dz_t \\
&= \phi_2(\boldsymbol{y}_{(ij)}; \boldsymbol{0}, \boldsymbol{R}_{ij}) \Phi_{t-2}(\boldsymbol{y}_{-(ij)|(ij)}; \boldsymbol{\mu}_{-(ij)|(ij)}, \boldsymbol{R}_{-(ij)|(ij)}), \\
&= \frac{1}{\sqrt{\sigma_{ii}\sigma_{jj}}} \phi_2(\boldsymbol{x}_{(ij)}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}) \Phi_{t-2}(\boldsymbol{x}_{-(ij)}; \boldsymbol{\mu}_{-(ij)|(ij)}, \boldsymbol{\Sigma}_{-(ij)|(ij)}).
\end{aligned}$$

A.3 CONDITIONAL CDF OF A NORMALIZED MULTIVARIATE NORMAL DISTRIBUTION

Let $\mathbf{X}$ be t dimensional vector that follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{D} = diag(\boldsymbol{\Sigma}) = (\sigma_{ii})$ be the diagonal matrix of variances and $\boldsymbol{Y} = \boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})$. Then $\boldsymbol{Y}$ is distributed as multivariate normal with mean $\mathbf{0}$ and correlation matrix $\boldsymbol{R} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{D}^{-\frac{1}{2}}$. Let $\boldsymbol{X}_1, \boldsymbol{X}_2$ and $\boldsymbol{Y}_1, \boldsymbol{Y}_2$ be a partition of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively of dimensions s and $(t-s)$, $s < t$.

Theorem **7.** The conditional distribution functions of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ and $\boldsymbol{Y}_1$ given $\boldsymbol{Y}_2$ are equal in the sense

$$\begin{aligned} & \Phi_s\left(\boldsymbol{x}_1; \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right) \\ = \ & \Phi_s\left(\boldsymbol{y}_1; \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{y}_2, \boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}\right) \end{aligned} \tag{53}$$

where $\boldsymbol{y}_i = \boldsymbol{D}_i^{-\frac{1}{2}}(\boldsymbol{x}_i - \boldsymbol{\mu}_i)$ for $i = 1, 2$.

Proof. Note that

$$\begin{aligned} & P(\boldsymbol{Y}_1 \leq \boldsymbol{y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2) \\ = \ & P(\boldsymbol{D}_1^{-\frac{1}{2}}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1) \leq \boldsymbol{y}_1 | \boldsymbol{D}_2^{-\frac{1}{2}}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2) = \boldsymbol{y}_2) \\ = \ & P(\boldsymbol{X}_1 \leq \boldsymbol{x}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2) \end{aligned} \tag{54}$$

since $\boldsymbol{y}_i = D_i^{-\frac{1}{2}}(\boldsymbol{x}_i - \boldsymbol{\mu}_i)$ for $i = 1, 2$. Now the conditional distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is normal with mean $\mu_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. And the conditional distribution of $\boldsymbol{Y}_1$ given $\boldsymbol{Y}_2 = \boldsymbol{y}_2$ is normal with mean $\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{y}_2$ and covariance $\boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}$. Therefore (53) is equivalent to (54).

Below are some results for the trivariate normal distribution with a special correlation structure.

Theorem **8.** Let X_k, X_l, X_j, where $k < l$, be distributed as trivariate normal with mean $\widetilde{\boldsymbol{\mu}} = (\mu_k, \mu_l, \mu_j)'$ and correlation

$$\widetilde{\boldsymbol{R}} = \begin{pmatrix} 1 & \rho^{|k-l|} & \rho^{|k-j|} \\ \rho^{|k-l|} & 1 & \rho^{|j-l|} \\ \rho^{|k-j|} & \rho^{|j-l|} & 1 \end{pmatrix}.$$

Then the conditional distribution of X_k, X_l given $X_j = x_j$ is bivariate normal with means $\mu_{k|j} = \mu_k - \rho^{|k-j|}(x_j - \mu_j)$ and $\mu_{l|j} = \mu_l - \rho^{|l-j|}(x_j - \mu_j)$ and covariance

$$\boldsymbol{\Sigma}_{kl|j} = \begin{pmatrix} 1 - \rho^{2|j-k|} & \rho^{|k-l|} - \rho^{|j-k|+|j-l|} \\ \rho^{|k-l|} - \rho^{|j-k|+|j-l|} & 1 - \rho^{2|j-l|} \end{pmatrix}.$$

Proof. Let us partition $\widetilde{\boldsymbol{R}}$ as

$$\widetilde{\boldsymbol{R}} = \left(\begin{array}{cc|c} 1 & \rho^{|k-l|} & \rho^{|k-j|} \\ \rho^{|k-l|} & 1 & \rho^{|j-l|} \\ \hline \rho^{|k-j|} & \rho^{|j-l|} & 1 \end{array}\right) = \begin{pmatrix} \widetilde{\boldsymbol{R}}_{11} & \widetilde{\boldsymbol{R}}_{12} \\ \widetilde{\boldsymbol{R}}_{21} & 1 \end{pmatrix}.$$

(X_k, X_l) given X_j follows a bi-variate normal distribution with mean (X_k, X_l) given X_j is $(\mu_k, \mu_l)' + \widetilde{\boldsymbol{R}}_{12}(x_j - \mu_j)$ and variance $\widetilde{\boldsymbol{R}}_{11} - \widetilde{\boldsymbol{R}}_{12}\widetilde{\boldsymbol{R}}_{21}$. Now,

$$\begin{pmatrix} \mu_k \\ \mu_l \end{pmatrix} + \widetilde{\boldsymbol{R}}_{12}(x_j - \mu_j) = \begin{pmatrix} \mu_k + \rho^{|k-j|}(x_j - \mu_j) \\ \mu_l + \rho^{|l-j|}(x_j - \mu_j) \end{pmatrix}, \text{ and}$$

$$\begin{aligned} \widetilde{\boldsymbol{R}}_{11} - \widetilde{\boldsymbol{R}}_{12}\widetilde{\boldsymbol{R}}_{21} &= \widetilde{\boldsymbol{R}}_{11} - \begin{pmatrix} \rho^{|k-j|} \\ \rho^{|l-j|} \end{pmatrix} \begin{pmatrix} \rho^{|k-j|}, & \rho^{|l-j|} \end{pmatrix} \\ &= \begin{pmatrix} 1 - \rho^{2|j-k|} & \rho^{|k-l|} - \rho^{|j-k|+|j-l|} \\ \rho^{|k-l|} - \rho^{|j-k|+|j-l|} & 1 - \rho^{2|j-l|} \end{pmatrix}. \end{aligned}$$

If we denote the conditional mean of X_k given X_j as $\mu_{k|j}$ and covariance of X_k

and X_l given X_j as $\sigma_{k,l|j}$, then $\mu_{k|j} = \mu_k - \rho^{|k-j|}(x_j - \mu_j),\ k, j = 1, 2, \ldots, c,\ k \neq j$

$$\mu_{k|j} = \mu_k + \rho^{|k-j|}(x_j - \mu_j), \forall k, j = 1, 2, \ldots, c,\ k \neq j. \tag{55}$$

Examples: Let us consider a multivariate normal random vector of $(V_1, V_2, V_3, V_4)'$ with parameters ($\boldsymbol{\mu}_4$ and $\boldsymbol{R}$), where

$$\boldsymbol{\mu}_4 = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} \text{ and } \boldsymbol{R} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}.$$

Then, $(V_2, V_3, V_4 \mid V_1 = x_1)$ follows multivariate normal with parameters $(\boldsymbol{\mu}_{-1|1}, \boldsymbol{R}_{-1|1})$, where

$$\boldsymbol{\mu}_{-1|1} = \begin{pmatrix} \mu_2 + \rho(x_1 - \mu_1) \\ \mu_3 + \rho^2(x_1 - \mu_1) \\ \mu_4 + \rho^3(x_1 - \mu_1) \end{pmatrix} \text{ and } \boldsymbol{R}_{-1|1} = \begin{pmatrix} 1-\rho^2 & \rho-\rho^3 & \rho^2-\rho^4 \\ \rho-\rho^3 & 1-\rho^4 & \rho-\rho^5 \\ \rho^2-\rho^4 & \rho-\rho^5 & 1-\rho^6 \end{pmatrix}.$$

Also, if we consider $(V_1, V_2, V_4 \mid V_3 = x_3)$, then it is distributed as a three dimensional multivariate normal distribution with parameters $(\boldsymbol{\mu}_{-3|3}, \boldsymbol{R}_{-3|3})$, where

$$\boldsymbol{\mu}_{-3|3} = \begin{pmatrix} \mu_1 + \rho^2(x_3 - \mu_3) \\ \mu_2 + \rho(x_3 - \mu_3) \\ \mu_4 + \rho(x_3 - \mu_3) \end{pmatrix} \text{ and } \boldsymbol{R}_{-3|3} = \begin{pmatrix} 1-\rho^4 & \rho-\rho^3 & 0 \\ \rho-\rho^3 & 1-\rho^2 & 0 \\ 0 & 0 & 1-\rho^2 \end{pmatrix}.$$

APPENDIX B

SELECTED R CODE

In this section, we provide a selection of R codes that we developed. Brief descriptions of all the important functions are stated below.

1. **choice.prob.Eq**: For a given set of means ($\boldsymbol{\mu}$) and correlation parameter (ρ), calculates the choice probabilities according to (28). Output is a vector of dimension equal to the dimension of $\boldsymbol{\mu}$.

2. **like.fn.Eq**: Calculates the likelihood function for a discrete choice data set and stated values of a set of parameters using the choice probability in (28). This function is evaluated in parallel and uses a modified version of the choice.prob.Eq function mentioned above.

3. **choice.prob.AR1**: For a given set of means ($\boldsymbol{\mu}$) and correlation parameter (ρ), calculates the choice probabilities according to (46). Output is a vector of dimension equal to the dimension of $\boldsymbol{\mu}$.

4. **like.fn.AR1**: Evaluates the likelihood function for a discrete choice data set and stated values of a set of parameters using the choice probability in (46). This function is evaluated in parallel and uses a modified version of the choice.prob.AR1 function mentioned above.

```
##########################################################################
########## Choice Probability for MDCG Equicorrelation #############
##########################################################################

# PDF of Gumbel Distribution
pdfgmbl <- function(y){
  f <- exp(-y)*exp(-(exp(-y)))
  return(f)
}
# CDF of Gumbel Distribution
cdfgmbl <- function(y){
```

```
  f <- exp(-(exp(-y)))
  return(f)
}

require("mvtnorm")
choice.prob.Eq <- function(mu, rho){
  c <- length(mu)
  R <- rho*matrix(1,c,c)+(1-rho)*diag(c)
  R1 <- (1-rho^2)*(((rho/(1+rho))*matrix(1,c-1,c-1))+
                     (1-(rho/(1+rho)))*diag(c-1))
  prob <- c()
  inverse <- function(x){
    y <- qnorm(cdfgmbl(x))
    return(y)
  }
  inverse1 <- function(x){
    return(0-log(0-log(pnorm(x))))
  }
  for (i in 1: c){
    condcdf <- function(v){
      condmean <- numeric(c-1)
      upperlimit <- numeric(c-1)
      for(j in 1: c-1){
        condmean[j] <- rho*v
        if(j < i){
          upperlimit[j] <- inverse(mu[i] - mu[j]+inverse1(v))
        }
        else{
          upperlimit[j] <- inverse(mu[i]-mu[(j+1)]+inverse1(v))
        }
      }
      return(pmvnorm(lower=rep(-Inf, c-1), mean=condmean,
                     upper = upperlimit, sigma = R1))
    }
    integrand <- function(z){
      p <- condcdf(z)
      p <- p*dnorm(x = z)
      return(p)
    }
    prob[i] <- integrate(f = integrand, lower = -Inf,
                         upper = Inf)$value
  }
  return(prob)
}
```

```
# Table to show only differences in utility matter

mu1 <- c(1, 1.5, 2)
mu2 <- mu1 - 5
mu3 <- mu1 + 100
rho <- seq(0, 0.9, by = 0.1)

diff.equicorr <- matrix(0, nrow = length(rho), ncol = 4)
for (i in 1:length(rho)){
  diff.equicorr[i, 1] <- rho[i]

  diff.equicorr[i, 2] <- paste0(format(round(choice.prob.Eq(mu1,
                        rho[i]), 2), nsmall = 2),
                        collapse =", ", sep = "")
  diff.equicorr[i, 3] <- paste0(format(round(choice.prob.Eq(mu2,
                        rho[i]), 2),nsmall = 2),
                        collapse =", ", sep = "")
  diff.equicorr[i, 4] <- paste0(format(round(choice.prob.Eq(mu3,
                        rho[i]), 2), nsmall = 2),
                        collapse =", ", sep = "")
}
diff.equicorr <- as.data.frame(diff.equicorr)
colnames(diff.equicorr) <- c("rho", paste0(mu1, collapse="",
                                      sep=","),
                             paste0(mu2, collapse="", sep=","),
                             paste0(mu3, collapse="", sep=","))
View(diff.equicorr)

# Table to show scale of utility should be normalized

mu1 <- c(2, 2.1, 2.5)
mu2 <- mu1/5
mu3 <- mu1*5
rho <- seq(0, 0.9, by = 0.1)

scale.equicorr <- matrix(0, nrow = length(rho), ncol = 4)
for (i in 1:length(rho)){
  scale.equicorr[i, 1] <- rho[i]

  scale.equicorr[i, 2] <- paste0(format(round(choice.prob.Eq(mu1,
                         rho[i]), 2),nsmall = 2),
                         collapse =", ", sep = "")
  scale.equicorr[i, 3] <- paste0(format(round(choice.prob.Eq(mu2,
                         rho[i]), 2),nsmall = 2),
                         collapse =", ", sep = "")
```

```
  scale.equicorr[i, 4] <- paste0(format(round(choice.prob.Eq(mu3,
                            rho[i]), 2),nsmall = 2),
                            collapse =", ", sep = "")
}
scale.equicorr <- as.data.frame(scale.equicorr)
colnames(scale.equicorr) <- c("rho", paste0(mu1, collapse="",
                                         sep=","),
                              paste0(mu2, collapse="", sep=","),
                              paste0(mu3, collapse="", sep=","))
View(scale.equicorr)

######################################################################
########## Likelihood Function for MDCG Equicorrelation ###########
######################################################################

require(mvtnorm)
require(doParallel)
require(foreach)

detectCores()
cl <- makeCluster(6)
registerDoParallel(cl)
getDoParWorkers()

like.fn.Eq <- function(data,theta){
  y  <- data[, 1]
  x  <- data[, 2 : ncol(data)]
  beta  <- theta[1 : length(theta) - 1]
  rho  <- theta[length(theta)]
  nc  <- length(y)
  n  <- nc/c
  x  <- data.matrix(x)
  mu  <- numeric(nc)
  mu <- apply(x, 1, function(z) z%*%beta)
  Mu  <- matrix(mu, nrow  =  n, byrow = T)
  Y <- matrix(y, nrow  =  n, byrow = T)
  Prob  <- matrix(0, nrow  <- n, ncol  <- c)
  inverse <- function(x){
    y <- qnorm(cdfgmbl(x))
    return(y)
  }
  inverse1 <- function(x){
    return(0-log(0-log(pnorm(x))))
  }
  choice.prob.1 <- function(Y, mu, rho) {
```

```
  if (length(Y) != length(mu)){
    stop("Y and mu are not of same length")}
  c <- length(mu)
  probability <- c()
  for (i in 1: c){
    if ( Y[i] == 1){
      condcdf <- function(v){
        condmean=numeric(c)
        for(j in 1:c){
          condmean[j] <- rho*v
        }
        condmean <- condmean[-i]
        upperlimit=numeric(c-1)
        for(j in 1:c-1){
          if(j < i){
            upperlimit[j] = inverse(mu[i] - mu[j]+inverse1(v))
          }
          else{
            upperlimit[j] = inverse(mu[i]- mu[(j+1)]+inverse1(v))
          }
        }
        R1 <- (1-rho^2)*(((rho/(1+rho))*matrix(1,c-1,c-1))+
                         (1-(rho/(1+rho)))*diag(c-1))
        return(pmvnorm( lower=rep(-Inf, c-1), mean=condmean,
                        upper = upperlimit, sigma = R1))
      }
      integrand <- function(z){
        p <- condcdf(z)
        p <- p*dnorm(x = z)
        return(p)
      }
      probability[i] <- integrate(f = integrand,
                                  lower = -Inf, upper = Inf)$value
    }
    else{
      probability[i] <- 0
    }
  }
  return(probability)
}

# Parallelized evaluation of choice probabilities

Prob <-
  foreach(i = 1:n, .combine = rbind, .packages = c("mvtnorm"),
```

```
              .export = c("cdfgmbl")) %dopar%{
                return(choice.prob.1(Y[i, ], Mu[i, ], rho))
              }

  prob <- matrix(t(Prob))

  loglikelihood  <- 0
  for (i in 1: nc){
    if(y[i] == 1 & is.na(prob[i]) == "FALSE" & prob[i] != 0){
      loglikelihood= loglikelihood + log(prob[i])
    }
  }
  return(-loglikelihood)
}

##########################################################################
########## Model Fitting ###############################################
##########################################################################

# TRANSPORT DATA
c  <-  4
Transport <- read.table("H:/Research/Data/Transport/Transport.txt",
                          header = T)
table(Transport$MODE)

Int  <-  matrix(0,nrow(Transport),c-1)

for (i in 1:nrow(Transport)){
  remainder <- i %% c
  for(j in 1:c-1){
    if (j==remainder){
      Int[i, j]  <-  1
    }
  }
}
Data  <-   cbind(Transport,Int)
Data  <-  cbind(sort(rep(1:210,4)),Data)
colnames(Data)  <-  c("Person",colnames(Transport),
                        "Int_Air","Int_Train","Int_Bus")

# Choice specific covariates

Reduced.Data=Data[,-7]
Reduced.Data=Reduced.Data[,-7]
colnames(Reduced.Data)
```

```
# Initializ with estimates from SAS for CNL model
initial.fromSAS <- c(-0.1036, -0.0849, -0.0133, 0.0693,
                     5.2047, 4.3606, 3.7632)
initial.SD.fromSAS <- c(0.0109, 0.0194, 0.002517, 0.0174,
                        0.9052, 0.5107, 0.5063)
initial <- c(initial.fromSAS, 0.05)

system.time(
  sol.reduced <- optim(initial,
                       like.fn.Eq,
                       data = Reduced.Data[, -1],
                       method = "L-BFGS-B",
                       lower = c(initial.fromSAS - 1.96*
                                 initial.SD.fromSAS, 0.1),
                       upper = c(initial.fromSAS + 1.96*
                                 initial.SD.fromSAS, 0.8),
                       control = list(trace = 6, maxit = 500,
                                      factr = 1e-11),
                       hessian = F)
)

# Bootstrap Estimation of SE

solution = list()
for (b in 1:50){
  id=sample(1:210,210,replace=TRUE)
  newdata=matrix(0,1,9)
  for (i in 1:210){
    persondata=as.matrix(Reduced.Data[(id[i]*4-3):(id[i]*4-0),])
    newdata=rbind(newdata,persondata)
  }
  newdata <- newdata[-1,]
  print(b)
  solution[[b]] <- try(optim(initial,
                             like.fn.Eq,
                             data = newdata[, -1],
                             method = "L-BFGS-B",
                             lower = c(initial.fromSAS - 1.96*
                                       initial.SD.fromSAS, 0.1),
                             upper = c(initial.fromSAS + 1.96*
                                       initial.SD.fromSAS, 0.8),
                             control = list(trace = 0, maxit = 500,
                                            factr = 1e-11),
                             hessian = F))
}
```

```
sol.pars <- matrix(0, nrow = 0, ncol = 8)
sol.LL <- c()
for(i in 1: ncol(sol.pars)){
  if(length(solution[[i]]) != 1){
    try(sol.pars <- rbind(sol.pars, solution[[i]]$par))
  }
}
summary(sol.pars)
sol.pars.mean <- apply(sol.pars, 2, mean)
sol.pars.sd <- sqrt(apply(sol.pars, 2, var) *
                    (nrow(sol.pars)/(nrow(sol.pars)-1)))
sol.pars.t <- c()
sol.pars.pvalue <- c()
for (i in 1: ncol(sol.pars)){
  sol.pars.t[i] <-
    sqrt(nrow(sol.pars))*(sol.pars.mean[i] /sol.pars.sd[i])
  sol.pars.pvalue[i] <- 2*pt(-abs(sol.pars.t[i]),
                             df = nrow(sol.pars))
}
sol.pars.table <- rbind(sol.pars.mean,sol.pars.sd,
                        sol.pars.t, sol.pars.pvalue)
round(sol.pars.table, 4)

########################################################################
########## Choice Probability for MDCG AR(1) #######################
########################################################################

choice.prob.AR1 <- function(mu, rho) {
  c <- length(mu)
  prob <- c()
  inverse <- function(x){
    y <- qnorm(cdfgmbl(x))
    return(y)
  }
  inverse1 <- function(x){
    return(0-log(0-log(pnorm(x))))
  }

  for (i in 1:c){
    condcdf <- function(v){
      condmean<-numeric(c)
      for(j in 1:c){
        condmean[j]<-rho^(abs(i-j))*v
      }
      condmean <- condmean[-i]
```

```
    upperlimit<-numeric(c-1)
    for(j in 1:c-1){
      if(j < i){
        upperlimit[j] <- inverse(mu[i] - mu[j]+inverse1(v))
      }
      else{
        upperlimit[j] <- inverse(mu[i]- mu[(j+1)]+inverse1(v))
      }
    }
    R1 <- matrix(0, nrow = c, ncol = c)
    for(k in 1:c){
      for(l in min(c, (k+1)):c){
        if(k != i & l != i){
          if (l < i){
            R1[k, l] <- rho^(l-k) - rho^(2*i-l-k)
          }
          if (k < i & i < l){
            R1[k, l] <- 0
          }
          if (i < k){
            R1[k, l] <- rho^(l-k) - rho^(l+k-2*i)
          }
        }
      }
    }
    for(k in 1:c){
      for(l in 1:k-1){
        R1[k, l] <- R1[l, k]
      }
      R1[k, k] <- 1 - rho^(2*abs(i-k))
    }
    R1 <- R1[-i, ]
    R1 <- R1[, -i]

    return(pmvnorm( lower=rep(-Inf,c-1), mean=condmean,
                    upper = upperlimit, sigma = R1))
  }
  integrand <- function(z){
    p <- condcdf(z)
    p <- p*dnorm(x = z)
    return(p)
  }
  prob[i] <- round(integrate(f = integrand, lower = -Inf,
                             upper = Inf)$value, 2)
}
```

```
  return(prob)
}

# Table to show only differences in utility matter

mu1 <- c(1, 1.5, 2)
mu2 <- mu1 - 5
mu3 <- mu1 + 100
rho <- seq(0, 0.9, by = 0.1)

diff.AR1 <- matrix(0, nrow = length(rho), ncol = 4)
for (i in 1:length(rho)){
  diff.AR1[i, 1] <- rho[i]

  diff.AR1[i, 2] <- paste0(format(round(choice.prob.AR1(mu1, rho[i]),
                          2), nsmall = 2), collapse =", ", sep = "")
  diff.AR1[i, 3] <- paste0(format(round(choice.prob.AR1(mu2, rho[i]),
                          2), nsmall = 2), collapse =", ", sep = "")
  diff.AR1[i, 4] <- paste0(format(round(choice.prob.AR1(mu3, rho[i]),
                          2), nsmall = 2), collapse =", ", sep = "")

}
diff.AR1 <- as.data.frame(diff.AR1)
colnames(diff.AR1) <- c("rho", paste0(mu1, collapse="", sep=","),
                       paste0(mu2, collapse="", sep=","),
                       paste0(mu3, collapse="", sep=","))
View(diff.AR1)

# Table to show scale of utility should be normalized

mu1 <- c(2, 2.1, 2.5)
mu2 <- mu1/5
mu3 <- mu1*5
rho <- seq(0, 0.9, by = 0.1)

scale.AR1 <- matrix(0, nrow = length(rho), ncol = 4)
for (i in 1:length(rho)){
  scale.AR1[i, 1] <- rho[i]

  scale.AR1[i, 2] <- paste0(format(round(choice.prob.AR1(mu1,
                     rho[i]), 2), nsmall = 2),
                     collapse =", ", sep = "")
  scale.AR1[i, 3] <- paste0(format(round(choice.prob.AR1(mu2,
                     rho[i]), 2), nsmall = 2),
                     collapse =", ", sep = "")
```

```
  scale.AR1[i, 4] <- paste0(format(round(choice.prob.AR1(mu3,
                     rho[i]), 2), nsmall = 2),
                     collapse =", ", sep = "")

}
scale.AR1 <- as.data.frame(scale.AR1)
colnames(scale.AR1) <- c("rho", paste0(mu1, collapse="", sep=","),
                         paste0(mu2, collapse="", sep=","),
                         paste0(mu3, collapse="", sep=","))
View(scale.AR1)

##############################################################################
########## Likelihood Function for MDCG AR(1) ######################
##############################################################################

require(mvtnorm)
require(doParallel)
require(foreach)

detectCores()
cl <- makeCluster(6)
registerDoParallel(cl)
getDoParWorkers()

like.fn.AR1 <- function(data, theta){
  y  <- data[, 1]
  x  <- data[, 2 : ncol(data)]
  beta  <- theta[1 : length(theta) - 1]
  rho  <- theta[length(theta)]
  nc  <- length(y)
  n  <- nc/c
  x  <- data.matrix(x)
  mu  <- numeric(nc)
  mu <- apply(x, 1, function(z) z%*%beta)
  Mu  <- matrix(mu, nrow  =  n, byrow = T)
  Y <- matrix(y, nrow  =  n, byrow = T)
  Prob  <- matrix(0, nrow  <- n, ncol  <- c)
  inverse <- function(x){
    y <- qnorm(cdfgmbl(x))
    return(y)
  }
  inverse1 <- function(x){
    return(0-log(0-log(pnorm(x))))
  }
  choice.prob.1 <- function(Y, mu, rho) {
```

```
if (length(Y) != length(mu)){
  stop("Y and mu are not of same length")}
c <- length(mu)
probability <- c()
for (i in 1:c){
  if ( Y[i] == 1){
    condcdf <- function(v){
      condmean=numeric(c)
      for(j in 1:c){
        condmean[j]=rho^(abs(i-j))*v
      }
      condmean <- condmean[-i]
      upperlimit=numeric(c-1)
      for(j in 1:c-1){
        if(j < i){
          upperlimit[j] = inverse(mu[i] - mu[j]+inverse1(v))
        }
        else{
          upperlimit[j] = inverse(mu[i]- mu[(j+1)]+inverse1(v))
        }
      }
      R1 <- matrix(0, nrow = c, ncol = c)
      for(k in 1:c){
        for(l in min(c, (k+1)):c){
          if(k != i & l != i){
            if (l < i){
              R1[k, l] <- rho^(l-k) - rho^(2*i-l-k)
            }
            if (k < i & i < l){
              R1[k, l] <- 0
            }
            if (i < k){
              R1[k, l] <- rho^(l-k) - rho^(l+k-2*i)
            }
          }
        }
      }
      for(k in 1:c){
        for(l in 1:k-1){
          R1[k, l] <- R1[l, k]
        }
        R1[k, k] <- 1 - rho^(2*abs(i-k))
      }
      R1 <- R1[-i, ]
      R1 <- R1[, -i]
```

```
          return(pmvnorm( lower=rep(-Inf ,c-1), mean=condmean,
                          upper = upperlimit , sigma = R1))
        }
        integrand <- function(z){
          p <- condcdf(z)
          p <- p*dnorm(x = z)
          return(p)
        }
        probability[i] <- integrate(f = integrand,
                               lower = -Inf, upper = Inf)$value
      }
      else{
        probability[i] <- 0
      }
    }
    return(probability)
  }

  # Parallelized evaluation of choice probabilities

  Prob <-
    foreach(i = 1:n, .combine = rbind, .packages = c("mvtnorm"),
            .export = c("cdfgmbl")) %dopar%{
              return(choice.prob.1(Y[i, ], Mu[i, ], rho))
            }

  prob <- matrix(t(Prob))

  loglikelihood  <- 0
  for (i in 1: nc){
    if(y[i]== 1 & is.na(prob[i]) == F & prob[i] != 0){
      loglikelihood= loglikelihood + log(prob[i])
    }
  }
  return(-loglikelihood)
}

##########################################################################
######### Simulated Data on Ordered Choices #####################
##########################################################################

require(dplyr)
require(reshape2)
require(mvtnorm)
```

```
N <- 100000
n <- 300
rho <- 0.7
c <- 3

# Generate the AR(1) correlation matrix

R <- matrix(0, nrow= c, ncol = c)
for(i in 1:c){
  for(j in 1:c){
    R[i, j] <- rho^(abs(i-j))
  }
}
R

# Generate the errors

e1 <- rmvnorm(n, mean = rep(0, c), sigma = R)
head(e1, n = 10)
round(cor(e1), 2)
e1 <- data.frame(e1)
e1 <- sample_n(tbl_df(e1), size = n)
e1 <- e1 %>%
  mutate(id = row_number())
R
# Check the sample correlation matrix
round(cor(e1), 2)

# Check the sample covariance matrix
round(cov(e1), 2)
str(e1)

e <- melt(data.frame(e1), id = c("id"))
e <-
  e%>%
  arrange(id) %>%
  group_by(id) %>%
  mutate(alternative = row_number(id))

# Generate the 1st covariate from a mixture of uniform distributions

x1.mix.prob <- sample(1:3, prob=c(0.3,0.35,0.35),
                      size=n*c, replace=TRUE)
x1 <- round(runif(n*c, min = c(40, 50, 60)[x1.mix.prob],
                  max =c(42, 57, 73 )[x1.mix.prob]), 3)
```

```
# Generate the 2nd covariate from a mixture of normal distributions

x2.mix.prob <- sample(1:3, prob=c(0.4,0.35,0.25),
                      size=n*c, replace=TRUE)
x2 <- round(rnorm(n*c, c(41, 47, 72)[x2.mix.prob],
                       c(2, 1, 2)[x2.mix.prob]), 2)

x <- data.frame(x1, x2)
Int  <- matrix(0, nrow(x), c-1)
for (i in 1: nrow(x)){
  remainder <- i %% c
  for(j in 1: c-1){
    if (j == remainder){
      Int[i, j]  <- 1
    }
  }
}
Int <- data.frame(Int)
colnames(Int) <- paste("Int", 1: (c-1), sep = "_")
x <- cbind(x, Int)
head(x)
summary(x)
e <- select(e, -variable)
x <- cbind(x, e)
x <- select(x, id, alternative, x1, x2, Int_1, Int_2, value)

# Select the parameters

b1 <- 1
b2 <- 1.5
b_int1 <- 1.4
b_int2 <- 2

# Generate the utilities

x <- mutate(x, u = x1*b1 + x2*b2 +
            Int_1*b_int1 + Int_2*b_int2 + value)

util <- dcast(x, id ~ alternative, value.var = "u")
util <-
  x %>%
  group_by(id) %>%
  summarize(u_max = max(u))
```

```
# Generate the response variable Y

x <- left_join(x, util, by = c("id"))
x <- mutate(x, y = ifelse(u == u_max, 1, 0))
sum(x$y)
unique(table(x$id, x$y)[, 2])
table(x$y, x$alternative)
prop.table(table(x$y, x$alternative), 2)
data.simulated.oredred <-
  x %>%
  select(id, y, alternative, x1, x2, Int_1, Int_2)

data.simulated.oredred <- data.simulated.oredred[, -3]

data.simulated.oredred$alternative <- as.factor(rep(1: 3, n))
save(data.simulated.oredred, "data.simulated.oredred.rda")

######################################################################
########## Example of Bootstrap Estimation ##########################
######################################################################

# Here we estimate the parameters of the simulated ordered
# choice data
require(doParallel)
require(foreach)

detectCores()
cl <- makeCluster(6)
registerDoParallel(cl)
getDoParWorkers()

c <- 3
rho <- 0.7
b1 <- 1
b2 <- 1.5
b_int1 <- 1.4
b_int2 <- 2
initial <- c(1.9, 1.6,  1.3, 1.7, 0.5)
initial.CNL <- initial[-5]

like.fn.AR1(data.simulated.oredred[, -1], initial)

sim.sol <- optim(initial,
                 like.fn.AR1, data=data.simulated.oredred[, -1],
                 method = "L-BFGS-B",
```

```
                lower = c(0.8, 1.3, 1.2, 0.7, 0.4),
                upper = c(1.3, 1.7, 1.7, 1.2, 0.9),
                control = list(trace=6, maxit = 5000),
                hessian=FALSE)

# Bootstrap estimation of the standard errors
solution.sim.AR1 <- list()
ptm <- proc.time()

for(b in 1: 50){
  set.seed(17624 + b)
  id <- sample(1: 300, 300, replace = T)
  id <- data.frame(id)
  bootstarp.data <- left_join(id, data.simulated.oredred, by = "id")
  print(b)
  cl <- makeCluster(6)
  registerDoParallel(cl)
  getDoParWorkers()

  solution.bootstrap.AR1 <-
    try(optim(initial,
              like.fn.AR1, data=bootstarp.data[, -1],
              method = "L-BFGS-B",
              lower = c(0.8, 1.3, 1.2, 0.7, 0.4),
              upper = c(1.3, 1.7, 1.7, 1.2, 0.9),
              control = list(trace=0, maxit = 5000),
              hessian = FALSE)
    )
  solution.sim.AR1[[b]] <- solution.bootstrap.AR1
}
proc.time() - ptm

sim.AR1.pars <- matrix(0, nrow = 0, ncol = 5)
for(i in 1: 50){
  if(length(solution.sim.AR1[[i]]) != 1){
    try(sim.AR1.pars <- rbind(sim.AR1.pars, solution.sim.AR1[[i]]$par))
}
sim.AR1.pars.sd <- sqrt(apply(sim.AR1.pars, 2, var) *
                        (nrow(sim.AR1.pars)/(nrow(sim.AR1.pars)-1)))
params.AR1 <- c(b1, b2, b_int1, b_int2, 0.7)

sim.AR1.pars.t <- numeric(5)
sim.AR1.pars.pvalue <- numeric(5)
for (i in 1: 5){
  sim.AR1.pars.t[i] <-
```

```
    sqrt(nrow(sim.AR1.pars))*(sim.sol$par[i] -
                                params.AR1[i])/sim.AR1.pars.sd[i]
  sim.AR1.pars.pvalue[i] <- 2*pt(-abs(sim.AR1.pars.t[i]),
                                 df = nrow(sim.AR1.pars))
}
sim.AR1.pars.table <- rbind(sim.sol$par, sim.AR1.pars.sd,
                            sim.AR1.pars.t, sim.AR1.pars.pvalue)
round(sim.AR1.pars.table, 4)

# P-values for CNL #
sim.CNL.pars <- c(3.8311, -0.6874, 2.0603, 3.3611) # From SAS
sim.CNL.pars.sd <- c(0.6754, 0.1295, 0.6205, 0.7577) # From SAS

params.CNL <- c(b1, b2, b_int1, b_int2)

sim.CNL.pars.t <- numeric(4)
sim.CNL.pars.pvalue <- numeric(4)
for (i in 1:4){
  sim.CNL.pars.t[i] <- (sim.CNL.pars[i] - params.CNL[i])/
    sim.CNL.pars.sd[i]
  sim.CNL.pars.pvalue[i] <- 2*pt(-abs(sim.CNL.pars.t[i]),
                                 df = 900)
}
sim.CNL.pars.table <- rbind(sim.CNL.pars, sim.CNL.pars.sd,
                            sim.CNL.pars.t, sim.CNL.pars.pvalue)
round(sim.CNL.pars.table, 4)
```

VITA

Arjun Poddar
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA, 23529, USA.

Education

Ph.D	Old Dominion University, Norfolk, VA (May 2016). Major: Computational and Applied Mathematics (Statistics).
MS	Old Dominion University, Norfolk, VA (December 2014). Major: Computational and Applied Mathematics (Statistics).
M.Sc	Indian Institute of Technology, Kharagpur, WB, India (May 2009). Major: Statistics and Informatics.
B.Sc	University of Calcutta, WB, India. (July 2007) Major: Statistics.

Experience

Statistical Consultant, LTCJ Cancer Research Center, Eastern Virginia Medical School, Norfolk, VA, (08/2013 - 06/2016).

Statistical Analyst, Chesapeake Bay Program, Old Dominion University, Norfolk, VA, (08/2011 - 07/2013).

Database Research Assistant, Bridgetree Research Services Pvt. Ltd., Kolkata, WB, India, (08/2010 - 07/2011).

Typeset using LaTeX