Modelling the Stability and Determinants of Household Food Insecurity: A Multivariate Longitudinal Ordinal Logistic Regression Approach

by

Jemal Ayalew Yimam

submitted in accordance with the requirements for the degree of

DOCTOR OF PHILOSOPHY

in the subject

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: Professor **John O. Olaomi**

2019

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

DECLARATION

Name: Jemal Ayalew Yimam

Student Number: 50879235

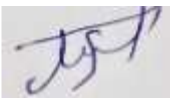
Degree: PhD

Exact wording of the title of the thesis as appearing on the electronic copy submitted for examination: 16

I declare the above thesis is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I submitted the thesis to originality checking software. The result summary is attached.

I further declare that I have not previously submitted this work, or part of it, for examination at UNISA for another qualification or at any other higher education institution.'



Signature

September. 2019

Date

Acknowledgements

First of all, I praise Allah, the almighty, for enabling me to complete this long journey. I have been very fortunate to have been supported by many people. First and foremost, my deepest appreciation goes to my Supervisor, Professor John O Olaomi. He is the most wonderful mentor who has always been a true inspiration and a role model. As my supervisor, he has given me constant support, guidance, encouragement, and good humour. I have learned a great deal from his expertise and experiences in research on modelling food insecurity determinants. I am greatly indebted to him for his patience, generosity and understanding as well as his insightful suggestions.

I would like to thank both Wollo University and University of South Africa for their valuable and unreserved arrangement of studying this education and granting both the ethical clearance and the financial support to accomplish this long journey work.

My special thanks also go to South Wollo Zone Woreda and Kebele leaders for their positive and genuine cooperation provided to me during data collection. Many thanks also for the Agricultural Agent workers participated as supervisor and data collector for their contribution to successfully accomplish the data collection.

I wish to thank my parents for their unconditional love and support throughout my life. Without it, I could have never made it this far. I am also grateful to my mother for her prayers and understanding.

Finally, I owe tremendous amount of gratitude to my wife, Seada Bezabih, and my children, Sumaya, Ameterrehman, Huzeyfa and Abdurrehman. Seada has been the wind beneath my wings, enabling me to fly so high to accomplish my goals. It was a long journey, but thanks to her love and support; it was incredible.

List of Abbreviations

2SLS	: 2-Stages Least Square
AMH	: Ali-Mikhail-Haq
C	: Copula(s)
CFSI	: Composite Food Security Index
EM	: Expectation Maximization
ERHS	: Ethiopian Rural Household Survey
FAO	: Food and Agricultural Organization
FCS	: Consumption Score
GEE	: Generalized Estimating Equations
GLS	: Generalized Least Squares
HDDS	: Household Dietary Diversity Score
HFIAS	: Household Food Insecurity Access Scale
HHS	: Household Hunger Scale
IEE	: Independent Estimation Approach
IFM	: Inference Function for Margins
IV	: Instrumental Variable
MAHFP	: Months of Adequate Household Food Provisioning
mAIC	: modified Akaike Information Criterion
ML	: Maximum Likelihood
MLE	: Maximum Likelihood Estimation
MPL	: Maximum Pairwise Likelihood
OLS	: Ordinary Least Square
PCAI	: Principal Component Analysis Index
PCC	: Pair-Copula Construction
PMF	: Probability Mass Function
SEM	: Structural Equation Modelling
SSP	: Stepwise Semi-parametric Estimator
USDA	: United States Department of Agriculture
WASH	: Water Supply, Sanitation, and Hygiene
WFS	: World Food Submit

Summary

Multivariate longitudinal ordinal data are collected for studying the dependence between multivariate ordinal outcomes, the changes over time and associated determinant factors. This emanates from the interdependence of the three dimensions of household food security statuses, the stability of these dimensions over time and the additional contribution of covariates on the dependence structure.

It is generally known that the random effect models have a lack of population-averaged interpretation for non-normally distributed outcomes in analysing ordinal data. In this thesis, we propose an alternative model for analysing multivariate longitudinal ordinal data with application to the household food insecurity by developing a pair copula construction (PCC) and cumulative logit marginal distributions-based model using the full maximum likelihood estimation (MLE) method. The simplified log-likelihood function of the D-vine pair copula multivariate discrete random variables was obtained with its parameters estimated.

Data were collected from 646 households living in selected rural Woredas of South Wollo Zone of the Amhara Regional State, Ethiopia from June 2014 to June 2015 three times at six months interval. Multistage cluster sampling was employed to select representative Woredas and households. The household food security status was determined using both the quartile score and composite index. Three distinct pair copula models with cumulative logit version were employed for multivariate, longitudinal and multivariate longitudinal ordinal data applicable for household food security.

The first model was employed to assess the dependence between food security dimensions and their corresponding determinant factors simultaneously. The copula parameter of this model indicated that household food security dimensions have significant and positive pairwise dependence. The marginal parameters showed that smaller land size, shortage of rainfall, cultivating once a year, and the presence of disease were positively associated with chronic to mild food insecurity in all dimensions. Moreover, cold agro-ecology and market price increase were associated with household food insecurity at availability and accessibility dimensions.

The second model was used to assess the stability of household food security over time and the determinant factors. The copula parameter revealed that individual household food security status is not stable over time. Moreover, the marginal parameter indicated that presence of crop disease, market price increase and medium agro-ecology were the significant recurrent factors for households to have chronic to mild food insecurity throughout the study period. One-time cultivation per year was the temporal significant factor for household food insecurity.

The third model was developed for measuring the dependence between the three dimensions, namely, their stability over time, the effects of the covariates both on the dependence structure, and stability over time simultaneously. The copula parameter of the population-average cumulative logit model revealed that food security dimensions were positively dependent to each other and the individual household food security status is not stable over time.

The marginal parameter of this model provided that lower agro-ecology, shortage of rainfall, presence of cultivation disease, increased market price, use of pesticides, cultivating smaller types of cereal crops, and cultivating once per year were positively affects the household food in-security in availability dimension. On the other hand, lower agro-ecology, increased market price, herbing small amount of livestock, hot agro-ecology and small farmland size positively affect the household food insecurity in the accessibility dimension. Furthermore, households headed by wife, divorced/widowed marital status of the household head, shortage of rainfall, and small farmland size positively affect the household food in-security in utilisation dimension.

This model provided a population-average interpretation with acceptable computational challenges in multivariate longitudinal ordinal data analysis. The study suggests that food security situation analysis is a multidimensional so that over-sighting the three dimensions over time simultaneously provides detail household food security situation than the single dimension. The pair copula population-average cumulative logit model addressed all the food security dimensions simultaneously, and the model found computationally effective. Therefore, we suggest this model to apply for other application areas for not extremely large number of outcomes and covariates.

Keywords:

Food insecurity; chronically food in-secured; composite food index; multivariate ordinal outcomes; longitudinal ordinal outcomes; multivariate longitudinal ordinal outcomes; marginal model; cumulative logit, pair copula; full maximum likelihood

Tables of Contents

List of Tables

List of Figures

Chapter One

1. Introduction

This study attempted to fill the methodological gap of jointly modelling the stability and determinants of the household food security for each dimension using multivariate longitudinal ordinal outcome. Since, multivariate joint modelling has a pretty advantage over separate modelling because it has relative efficiency to estimate the model and provide a powerful test of significance than the univariate (Gueorguieva, 2001; Molenberghs and Verbeke, 2006).

Among previous studies for multivariate longitudinal ordinal outcomes, the majority of them were concentrated around random effect models. Among these, random effect models in the context of item response theory (Liu, 2008, Liu and Hedeker, 2006), random effect models in the context of latent variable model (Cagnone et al., 2009), random effect models using subject-specific model (random intercept models) (Choi, 2012, Verbeke et al., 2014) and random effects models by introducing a continuous distributed random variable underneath the ordinal outcomes, some form of latent variable models (Laffont et al., 2014) were employed so far. However, the random effect models have lack of a population-averaged interpretation for non-normally distributed outcomes and some computational challenges (Abegaz et al., 2015, Nooraeea, 2015).

Despite numerous studies on multivariate longitudinal outcomes, relatively little research has been conducted on the marginal models for ordinal outcomes on the context of population average. Among these, Generalized Estimating Equation (GEE) was one of the most popular models. GEE was proposed for binary outcome and continuous outcome (Rochon, 1996). Furthermore, Gray and Brookmeyer (2000) proposed multivariate longitudinal models for continuous and discrete/time-to-event response variables using GEE. Another marginal model tailoring GEE was proposed for measuring multicity measured ordinal outcomes (Huang et al., 2002). Even though GEE is popular and provided consistent estimators for the regression parameters in the population-average interpretation; when the focus of the analysis includes certain aspects of the association structure, the construction of the joint model using GEE becomes more complex as it implies making assumptions about the within-outcome, the between-outcome, and the cross outcome association, and inferences of interest can be very sensitive with respect to the assumptions made (Verbeke et al., 2014). Moreover, GEE cannot incorporate the contribution of determinants of the outcome in determining the dependence between outcomes since it measures it using a working correlation independent of the determinants.

Another alternative model that helps for parameter estimation for non-normally distributed continuous or categorical data was the quasi least squares (Chaganty and Naik, 2002). A marginalised bivariate model using Kronecker product (KP) covariance structure was also proposed to handle two longitudinal ordinal outcomes (Lee et al., 2013). Marginal models using maximum likelihood estimation (MLE) on the context of multivariate t-copula were developed for multivariate longitudinal regression model for ordinal responses (Abegaz et al., 2015). The computation of the probability mass function for a discrete multivariate copula including multivariate t-copula requires 2m for evaluation while the pair-copula construction (PCC) method reduces the computation challenge of discrete multivariate copulas which requires only the evaluation of 2m (m − 1) bivariate copula functions (Lennon, 2016, Nicklas, 2013; Panagiotelis et al., 2012, Stöber et al., 2015).

In the context of PCC, a general framework for modelling multivariate repeated measurements using PCC (Shi and Yang, 2016, Shi and Zhao, 2018) and copula-based GLMM models by combining random-effects models and the D-vine copulas (Zhang et al., 2019) have been proposed for investigating multivariate longitudinal data with mixed-types of responses. All these copula-based models have yet not implemented in multivariate longitudinal ordinal outcomes.

Hence, the current study further attempted to resolve both the population-average interpretations in the random effect models and computational challenges of the multivariate copulas using pair copula construction for multivariate longitudinal ordinal data. Furthermore, pair copula construction requires the determination of the appropriate marginal distributions based on the nature of the outcomes (Aas et al., 2009, Czado, 2010, Nelsen, 2007, Panagiotelis et al., 2012). Since the responses of the current study variables are ordinal, the natural choices for ordinal data are cumulative logit or probit models. Hence, we selected cumulative logit model since the scale of the logistic is greater than the normal and this made the interpretation easier for logistic version and popular in many fields (Choi, 2012). In this thesis, the pair copula construction and cumulative logit marginal distributions are incorporated for the development of multivariate longitudinal ordinal model using the full maximum likelihood estimation (MLE) method in particular for the current application data.

The research problem informing this thesis lies in that, reviewed literature demonstrates that multivariate longitudinal ordinal outcomes both in the random effect and population-average have some limitations in providing the appropriate interpretations or computational challenges. The random effects model lacks the population-average interpretation and the population-average models are limited in number and have computational challenges. As a result, analysing multivariate longitudinal ordinal outcomes need improvement with the hope of incorporating the population-average interpretation that can resolve computational challenges. Furthermore, literature demonstrates that food security experts do not amply utilize the rigour of statistics. Despite the availability of robust statistical tools that have the rigour to satisfy the quest for assessing the stability and determinants of food insecurity, existing analysis of survey data heavily depend on the rudimentary, exploratory or descriptive statistics that lack depth. As a result, modelling stability and determinants of food insecurity analysis lack the efficiency of scientifically established evidence.

This thesis is, therefore, aimed at achieving the following main objectives:

1. To develop population-average multivariate longitudinal ordinal models using Pair Copula Construction, with application to household food insecurity; and
2. To jointly model the stability and determinants of household food insecurity using multivariate longitudinal ordinal model approach.

Furthermore, this thesis is, therefore, aimed at achieving the following specific objectives in line with the main objectives:

1. To jointly model the household food security determinants for the three food security dimensions and the dependence between them using Pair Copula Construction based Multivariate Ordinal data analysis;
2. To jointly determine household food security dependence between successive time periods and respective determinants using Pair Copula Construction based Longitudinal Ordinal data analysis;

3. To develop population-average based Pair Copula Construction models for Multivariate Longitudinal Ordinal Data;
4. To apply the population-average based Pair Copula Construction for jointly modelling the stability and determinants of household food insecurity for each dimension; and
5. To assess food security situation in selected Woredas of South Wollo Zone, Amhara Region, Ethiopia.

In this light, the study sought an answer to three specific questions:

1. Does the Pair Copula Construction approach alleviate the lack of the population-average interpretation of the random effect models in modelling multivariate longitudinal ordinal outcomes?
2. Does the Pair Copula Construction approach alleviate the existing computational challenges of population-average marginal multivariate copula models in modelling multivariate longitudinal ordinal outcomes?
3. Does the population-average based pair copula construction of the multivariate longitudinal ordinal model fits for modelling the stability and determinants of household food insecurity?

To provide answers to these questions, this thesis aims at exploring the rigour of models that help in resolving the population-average interpretation of the random effect models and the computational challenges of the multivariate copula models in the population-average version in modelling multivariate longitudinal ordinal outcomes. By doing so, the study hopes to strengthen the theory of statistics in alleviating the aforementioned limitations and enlarge statistical methods in capturing the complete information of the multivariate longitudinal ordinal data analysis. Moreover, this thesis aims at exploring the rigour of models that may help in predicting the determinants of household food insecurity for each dimension simultaneously, generating indices for the dependence of the dimensions, and the stability of the three dimensions over time. By so doing, the study hopes to strengthen food security monitoring, evaluation and reporting systems toward more robust, statistics-based predictive analysis. The usual analysis approaches often shy away from such approaches in a misconception that statistical methods are complicated and user-unfriendly. The study specifically taps into recent work in constructing jointly estimating the stability and determinants for each dimension.

Household food security becomes the multifaceted problems with multidimensional serious impacts since 1970 (Abafita and Kim, 2014). The widest acceptable definition of food security states that the existence of food security occurs when “all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life”(Pinstrup-Andersen, 2009). This definition consists of four important interlinked dimensions, namely; physical availability of food, economic and physical access to food, food utilisation and stability of the other three dimensions over time (Abafita and Kim, 2014, FAO., 2014, Napoli et al., 2011). Moreover, the usual category of household food security levels “food secured” or “food in-secured” were further disaggregated into “severe food in-secured”, “moderately food in-secured”, “mildly food in-secured” and “food secured” (Capaldo et al., 2010, Hunnes, 2013). Hence, the first three food security dimensions take these categories to classify food security status of the households.

Modelling the determinants of household food security should be assessed using the first three dimensions simultaneously over time to oversee the entire household food security situation and the stability over time (Capaldo et al., 2010, FAO, 2008). Knowing that the levels of food insecurity are ordinal in their nature (Magaña-Lemus et al., 2016) and hence modelling of food insecurity is the question of ordinal data analysis and each dimension can be considered as a response factor which have ordinal outcomes. Therefore, the methodological issue of food insecurity is the generalization of developing multivariate longitudinal ordinal data analysis. Furthermore, each of the dimensions has non-normal correlation to each other. Hence, identifying the determinant of one or two of the dimensions will not reflect the entire food security situation (FAO, 2008, FAO., 2014).

Numerous studies were conducted on household food security determinants separately for each dimension or in a composite index (Abafita and Kim, 2014, Aspelund, 2002, Birhane et al., 2014, Endale et al., 2014, Etana and Tolossa, 2017, Méthot and Bennett, 2018, Moroda et al., 2018, Motbainor et al., 2016, Ngema et al., 2018). However, the findings have been quite mixed and conflicting. This is owing to some of the food access proxy indicators which have served as food availability proxy indicators and vice versa. Similarly, some of food access proxy indicators have served for utilisation proxy indicators. This implies that understanding several concepts associated with the definition of food security are necessary before examining the determinants of food security. Moreover, despite the fact that FAO recommended that the food security dimensions has been addressed simultaneously, there is no work conducted on the subject that can accommodate the determinants for each dimension jointly as well as the stability and determinants for each dimension over time jointly in a single model.

Longitudinal food security household survey was conducted in selected rural Woredas of the South Wollo Zone of Amhara regional state in Ethiopia to illustrate the PCC model in multivariate longitudinal ordinal data. Ethiopia is one of the poorest countries in the world, and about 90% of the populations live in the rural areas. The problem of food insecurity has continued to persist in many rural households of the country. The seriousness of the problem varies from one area to another depending on the state of the natural resources and the extent of development of these resources (Asmamaw et al., 2015, Endalew et al., 2015). Rural food insecurity is one of the defining features of rural poverty, particularly in the moisture-deficit northeast highland plateaus and some pastoral areas of Ethiopian(Agidew and Singh, 2018). The study area, the South Wollo Zone, is among these areas, which is mostly affected by food insecurity (Agidew and Singh, 2018, Asmamaw et al., 2015).

Numerous studies have been conducted in Ethiopia regarding the subject of food security with different results with particular recommendations, and various measures have been taken (Abafita and Kim, 2014, Abdu et al., 2018, Abegaz, 2017, Agidew and Singh, 2018, Ahmed et al., 2017, Asmamaw et al., 2015, Assefa, 2015, Bashir and Schilizzi, 2012, Birhane et al., 2014, Castro, 2000, Endale et al., 2014, Endalew et al., 2015, Etana and Tolossa, 2017, Moroda et al., 2018, Motbainor et al., 2016, Negatu, 2004, Nigussie and Alemayehu, 2013, Shone et al., 2017). However, only two studies have been conducted for the last 15 years in South Wollo Zone in particular in Tewuledery (Agidew and Singh, 2018) and Sayint (Asmamaw et al., 2015) Woredas. These studies did not reflect the entire food security situation of the zone and they are concerned only on the food access dimension.

On the other hand, studies conducted by (Castro, 2000, Negatu, 2004) covered the overall situation of the zone. However, the studies were conducted about 15 years ago, in which the level of food security and the economy as a whole are very different from this time. Many socio-economic factors have been changed in the South Wollo Zone and the country too. Therefore, it is important to update the stability and determinants of household food security or insecurity situation incorporating all potential factors in the recent models. In this regard, our model is more robust than the previous which will lead to a more accurate result.

This thesis consists of seven chapters including this chapter as it is organized in the following chapters. Chapter 2 provides the review of food security concepts, definition, measuring methods, and determinant factors. The statistical models that can serve for multivariate, longitudinal and multivariate longitudinal ordinal outcomes are also discussed. In Chapter 3, we describe the data used for this thesis including source and type of data, sample size and sampling procedures, data collection procedures and methods and food security measuring methods.

The methods employed for modelling the determinant factors of food security are discussed in Chapter 4. In this chapter, we presented a pair copula construction based marginal cumulative logit model for multivariate, longitudinal and multivariate longitudinal outcomes to estimate the dependence between the ordinal outcomes and their respective determinants. We developed algorithm to select appropriate bivariate copula families to represent the dependence measures in the model.

In Chapter 5, we presented the analysis household data using the methods presented in Chapter 4. It consists of the findings of the pair copula construction-based multivariate ordinal cumulative logit model for assessing the dependence between food availability, accessibility and utilisation, and their respective determinants. It also presents the findings of the pair copula construction-based longitudinal ordinal cumulative logit model for assessing the dependence between food security status of households in the three rounds and the respective determinants. Lastly, the findings of the new population-average pair copula construction approach for multivariate longitudinal ordinal data to measure the dependence between food security dimensions, the stability over time and the respective determinants simultaneously are presented.

In Chapter 6, we present the discussion for the findings presented in Chapter 5. In Chapter 7, we present the contribution of this thesis for statistical methods and future works for further developments. We ended the thesis with the questionnaire, simplification of the full maximum likelihood and log-likelihood functions of the developed models as appendices.

Chapter Two

2. Literature Review

2.1 Statistical Models in Food Security

The issue of food security can be addressed using either of the three models including multivariate ordinal models, longitudinal ordinal models and multivariate longitudinal models. The statistical models employed so far for multivariate ordinal outcomes, longitudinal ordinal outcomes and multivariate longitudinal ordinal outcomes were reviewed as follows.

2.1.1 Multivariate Ordinal Models

Among different previous studies for multivariate ordinal data, (Gange, 1994) developed Generalized Estimating Equations (GEE) methods for correlated ordinal responses which extends the model developed by (Liang and Zeger, 1986) for correlated binary data. GEE method is still the applied model with different versions.

On the other hand for performing multivariate ordinal data analysis, Structural Equation Modelling (SEM) with two stage methodology Maximum Pairwise Likelihood (MPL) - Generalized Least Squares (GLS) method was developed (Liu, 2007). Large sample simulation studies of this method showed that the parameter and standard error estimates, and the test statistics are acceptable. However, standard error formulae underestimate empirical variability for small sample size less than 200.

Moreover, a multivariate non-linear model for ordinal responses was also developed (Aspelund, 2002). In this model, a linear- by-linear log linear model with independent estimation approach (IEA) along robust standard errors was used. IEA performed well if only the marginal parameter estimates were of interest.

Multivariate ordered probit model with pairwise likelihood inference was employed for multivariate ordinal responses in the continuous latent variable model (Kenne Pagui and Canale, 2016). The model was applied in PLordprod R package and found that the model reduced the computational problems related to the calculation of a q dimensional integral for each single likelihood using full likelihood. However, the model is unable to compute both the mean of the latent variables and the first threshold.

The composite likelihood methods with a latent variable specification were also applied using both the probit and the logit link functions for multivariate ordinal regression model (Hirk et al., 2018). The model was executed in mnormt R package; both link functions resulted into recovery for high correlation parameters whereas for low correlation both of the link functions were not recovered.

In line with the existing methods, copulas have been popular tools for modelling multivariate outcomes since copulas have several attractive properties. The first attractive properties of copulas are allowing us to construct separately the dependence structure from the joint dependence structure and the marginal probabilities. The second properties are invariant, under continuous and increasing transformations. The third one is, unlike correlation, they do not require elliptically distributed for the marginals. Lastly, they can be used to measure tail dependences of the joint distribution (Syring, 2013) . "A copula is a function which joins a multivariate distribution function to its one-dimensional marginal distribution functions (Nelsen, 2007)." With a closed form, MLE is straightforward for copula functions. For m-dimensional data, the probability mass function can be computed using 2^m finite differences of the copula function. As a result, the approach is computationally intensive, and becomes infeasible for high dimensional problems (Panagiotelis et al., 2012).

Another extension of copulas that have attractive properties for discrete data have been developed. This extension is called elliptical copulas, in particular Gaussian copulas that can capture both positive and negative dependence under closed marginalization. As many elliptical copulas, including the Gaussian copula, cannot be written down in closed form, MLE through taking finite differences is not a feasible option (Panagiotelis et al., 2012). In a similar vein, to estimate models based on Gaussian copulas, Bayesian methods have been used (Pitt et al., 2006). In general, both Frequentist and Bayesian techniques discussed above are computationally intensive, and may not be applied easily to higher dimensions (Panagiotelis et al., 2012).

Among copula methods applied for multivariate ordinal data, a multivariate ordered logit regression with the notion of multivariate copula was modelled (Dardanoni and Forcina, 2008).

This model describes how the joint distribution of a set of ordinal response variables depends on exogenous regressors. The nature of the main properties of the marginal parameterization and the global interaction copula was found to be nonparametric. The model was found to be an efficient model for estimation purpose for small response variables.

In response to these challenges, a pair-copula construction (PCC) method has been developed for multivariate copulas using only bivariate copulas. PCC was originally developed for continuous random variables and then extended for discrete random variables (Panagiotelis et al., 2012, Syring, 2013). The advantages of this approach are one PCC provide a highly flexible framework for constructing copulas exhibiting a wide range of dependence characteristics. Second, the computation of the probability mass function for a discrete PCC only requires the evaluation of 2m (m − 1) bivariate copula functions, whereas the multivariate copulas requires 2^m for evaluation. As a result, MLE is feasible even for higher (Panagiotelis et al., 2012).

2.1.2 Longitudinal Ordinal Models

GLM extended to the longitudinal setting in two types of generalizations includes subject-specific and marginal models (Fitzmaurice et al., 2009; Koper and Manseau, 2009). The subject-specific models are the class of generalised linear mixed models (GLMM) that consider the association between ordinal outcomes within a subject by treating some of the model parameters as random variables. Maximum likelihood (ML) estimation method is commonly used to compute the fixed and random effect parameters. The consequence of random effect in longitudinal ordinal outcome is that the association is always positive and interpretations of the fixed parameter estimates for the population of subjects are not straightforward (Fitzmaurice et al., 2009**).**

Among different previous studies for analysing clustered data with ordinal responses in the GLLM classes, mixed-effect model was introduced for the first time by (Harville and Mee, 1984). The estimates of the random effects were approximated by Taylor series expression. This was also advanced for parameter estimation purpose through numerical quadrature method with one random effect by (Jansen, 1990).

The marginal model (population-average) interpretation is not obvious owing to the complication of integrating out the random effects since generally assumed as normally distributed. To overcome these issues, different random effect models were proposed. The random effect model that captured the limitation in marginal model was developed which is the mean response depends only on the fixed-effect and not the individual effect with complex correlation structure (Tutz and Hennevogl, 1996). Furthermore, a maximum likelihood utilizing quasi-Newton algorithms with Monte Carlo integration of the random effects was developed for the random effects of the longitudinal ordinal data (Lee, 2008). The model can be executable in any software which have independence proportional odds model (IPOM). However, the models were particularly useful in longitudinal analyses with a moderate to large number of repeated measurements per subject.

In contrast with GLMMs, marginal models consider the association between ordinal outcomes at population level. In the class of marginal models, Fitzmaurice and Laird (1993) developed a marginal model or population-averaged model with maximum likelihood (ML) approach for repeated ordinal responses that captured only the mean response effects on the particular specified predictors of interest but not on the individual effects (Fitzmaurice and Laird, 1993). Since ML approach for fitting marginal model is awkward, GEE was developed as alternative method for the first time in place to fit marginal model through cumulative logit (Lipsitz et al., 1994). GEE provides consistent estimators for the regression parameters when the model has been correctly specified even if it has the limitation of treating the association structure as nuisance parameters. Moreover, GEE2 was also developed for modelling the association structure using global odds ratio, while in marginal model, it is considered as a nuisance (Heagerty and Zeger, 1996). However, the interpretation of the association structure is difficult for ordinal outcomes because GEE2 does not lead to a multivariate distribution for the ordinal outcomes and thus complicates the interpretation of association structures. On the contrary, to improve of the existing models during the century, Perin (2009) developed a model called "alternate formulation of alternating logistic regressions model" using orthogonalised residuals to consider the association structure in marginal models for longitudinal ordinal data. Similarly, an alternative logistic regression (ALR) was proposed to provide insight and some advantages in the marginal model estimated via GEE and subject-specific models estimated via GLMMs (Bhatnagar et al., 2015). The model was executed in SAS/STAT version 9.3 and the model behaved similarly to marginal models estimated via GEE for mean effects. However, it was difficult to ascribe clustering to the correct level, particularly for ALRs.

To improve the association structure different forms of GEE have been proposed. Of these among the recent one, local odds ratios parameterization structure executed in multgee R package (Touloumis et al., 2013) and weighted score methods executed in weighted Scores R package (Nikoloulopoulos, 2017) are found. The latter allowed latent correlation structure for the selection of the correlation structure not restricted to an exchangeable or unstructured one and reduces the computational challenge for large dimensions.

Alternatively, Nooraee et al., (2016) developed an approximate marginal logistic distribution model for the analysis of longitudinal ordinal data that can accommodate majority of the limitations in the existing models. The model was executed in existing packages in R and provided comparable interpretation with GEE. Moreover, the model can be applied without having to use additional analysis such as multiple imputation over the other methods (GEE) if incomplete outcome data fulfils the ignobility assumptions and sample size are not too small (n>=100). However, the model expected to be sensitive to strong deviations from the multivariate t-distribution for latent variables for estimation of the correlation coefficient.

In line with the existing methods, among recently developed models for modelling repeated or longitudinal outcomes, the most popular tools are copula. A copula model with bivariate copula function was one of the presented model-to-model repeated ordered categorical data (Vandenhende and Lambert, 2000). In this model, the standard cumulative regression models were used to model the marginal distributions and the copula function was used to model the dependence between repeated responses. Even though, the copula models addressed both the marginal parameters and association structures in repeated ordinal data as well, it might not be suitable to quantify dependence over the bounds.

Another extension of copulas that have attractive properties for longitudinal ordinal data has been developed. This extension is the multivariate ordered probit model on the basis of multivariate copula representation for obtaining the maximum likelihood estimates of the parameters of longitudinal ordinal model (Kurada, 2011). The model was executed in Mprobit package in R software and the result showed that the model was computationally challenging to implement it.

In response to these challenges, a pair-copula construction (PCC) method has been developed for multivariate copulas using only bivariate copulas. PCC was originally developed for continuous random variables and then extended for discrete random variables (Panagiotelis et al., 2012, Syring, 2013). The advantages of this approach are one PCC provides a highly flexible framework for constructing copulas exhibiting a wide range of dependence characteristics. Moreover, the computation of the probability mass function for a discrete PCC only requires the evaluation of 2m (m − 1) bivariate copula functions whereas the multivariate copulas requires 2^m for evaluation. As a result, MLE is feasible even for higher and it was executed in R software (Panagiotelis et al., 2012).

2.1.3 Multivariate Longitudinal Ordinal Models

The statistical analysis of multivariate longitudinal ordinal data for assessing the changes across time can be addressed either by reducing the dimension of the multivariate longitudinal data to univariate longitudinal data using some kind of summary measures, or jointly addressing the associations/dependencies across multivariate covariates and the changes across time points. A statistical review for the first part was produced in different literatures; Verbeke et al., 2014). Several approaches for jointly modelling multivariate longitudinal data have been proposed in the statistical literature that includes three main classes. These are the subject- specific (random effect) models, the marginal (population-average) models and full specification of the multivariate distribution for the outcomes (Copula models).

Among previous studies for multivariate longitudinal ordinal outcomes, majorities of them were concentrated around random effect models. One approach of this model in the context of item response theory that can handles three-level multivariate ordinal outcomes in longitudinal settings and can accommodates multiple random subject effects was developed using iterative Fisher scoring solution for estimating all required parameters and their corresponding standard errors (Liu, 2008; Liu and Hedeker, 2006). The model was implemented in the GAUSS language (GAUSS 3.6). Another approach for this random effect models in the context of latent variable model was also developed to account for the correlation between the time points using item-specific random effects with a full information MLE method (Cagnone et al., 2009). A FORTRAN program was written to implement the model. The models can be extended in many different directions but more difficult to be implemented computationally.

Similarly, in the setting of random effect models, another model was also developed using subject-specific model (random intercept models) for the longitudinal part and conditioning these random effect models to account for the repeated independent cross-sectional outcomes. One extension of this model was also developed that can generate multivariate correlated random effects across the repeated cross-sectional outcomes from the subject-specific random effects model which varies across each cross-sectional outcome (Choi, 2012, Verbeke et al., 2014). Furthermore, another extension of random effects models was developed to relax the independence assumption on the conditional distribution given random effects by introducing some type of latent variable models (Laffont et al., 2014). The model was implemented in R package with probit mixed effects model with a latent variable interpretation. The authors pointed out that the model worked well for their application data even if the probit model offers less flexibility than other (logistic) models, requiring only a limited number of parameters to be estimated. Therefore, extensions of the model can be considered to accommodate more complex situations.

Even though random effect models have provided many advantages especially to compute correlations among outcomes through random effects, they have lack of a population-averaged interpretation for normally distributed outcomes and some computational challenges (Abegaz et al., 2015, Nooraeea, 2015).

In line with random effect models, very little marginal models were developed for multivariate longitudinal categorical or ordinal outcomes with the hope to resolve the limitation of population-average interpretation in random effects models. One of the alternative model that helps for parameter estimation for non-normally distributed continuous or categorical data was the quasi least squares (Chaganty and Naik, 2002). Other marginal model approaches that tailoring GEE was proposed for measuring multicity measured ordinal outcomes (Huang et al., 2002) since GEE was implemented for binary and time-to event outcomes through combining two GEE models for the two outcomes, using an autoregressive-type working correlation matrix for the intra- and inter-outcome dependence over time (Rochon, 1996). Furthermore, Gray and Brookmeyer (2000) proposed multivariate longitudinal models for continuous and discrete/time-to-event response variables using GEE approach the popular population-average interpretation model. GEE approach still has limitation in modelling multivariate longitudinal outcomes since it treats association as nuisance and measures it using working correlation. Another marginalised bivariate model using Kronecker product (KP) covariance structure to capture the correlation between processes at a given time and the correlation within a process over time (serial correlation) for bivariate longitudinal ordinal data was employed (Lee et al., 2013). The model was implemented in R package but limited for only two longitudinal ordinal outcomes.

On the other hand, alternative model in the marginal models using the MLE method on the context of multivariate t-copula was developed for multivariate longitudinal regression model for ordinal responses, through a computationally efficient Monte Carlo EM approach (Abegaz et al., 2015). The computation of the probability mass function for a discrete multivariate copulas including multivariate t-copula require 2^m for evaluation (Panagiotelis et al., 2012).

In the third classes of modelling multivariate longitudinal outcomes, the full specification of the multivariate distribution for the outcomes was implemented specially using multivariate copula models. A general framework for modelling multivariate repeated measurements was also proposed for mixed type of outcomes (Shi and Yang, 2016; Shi and Zhao, 2018). The longitudinal observation of each response was separately modelled using pair copula constructions with a D-vine structure. Then the multiple D-vines were then joined by a multivariate copula. The model was executed in R package using zero inflated Poisson regression and sequential approach was used for inference purpose.

In line with the existing PCC models, copula-based GLMM models have been proposed for investigating multivariate longitudinal data with mixed-types of responses by combining random-effects models and the D-vine copulas (Zhang et al., 2019). The D-vine copula measured the correlation between multiple responses measured at a given time point. Furthermore, the non-parametric maximum likelihood method was used instead of specifying the random effects distribution. The model was executed in R package using c continuous and binary outcomes and the result showed that the non-parametric models were more efficient and flexible than the usual Gaussian models. However, the model converged slowly when the number of mass points K is large. All the copula-based model reviewed so far have not yet been implemented in multivariate longitudinal ordinal outcomes.

Hence, modelling multivariate longitudinal ordinal outcomes using pair copula construction will reduce the computational challenges of the probability mass function evaluation. Therefore, this model can resolve both the population-average interpretations in the random effect models and computational challenges of the multivariate copulas.

All the literatures reviewed herewith for multivariate, longitudinal and multivariate longitudinal ordinal models did not mean they are part of the analysis. However, we tried to show the evolution of the multivariate longitudinal ordinal model developmental process until the recent year. The thesis established itself on the implementation of the PCC model for multivariate longitudinal ordinal outcomes for jointly assessing the stability over time, the dependence between ordinal outcomes and the determinants for each ordinal outcome at the same time.

2.2 Definition and Concept of Food Security

Since 1970 food insecurity has brought multifaceted problems with multidimensional serious impacts and became the first debating issue on the development that concerns the whole of mankind (Abafita and Kim, 2014). Since then, the issue of food security has discussed and diversified immensely worldwide. The widest acceptable definition of food security was also acquired during the World Food Submit (WFS) held in 1996, which states that the existence of food security occurs when "all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life"(Pinstrup-Andersen, 2009).

This definition consists of four important interlinked dimensions, namely, physical availability of food, economic and physical access to food, food utilisation and stability of the other three dimensions over time (Abafita and Kim, 2014, FAO, 2014, Napoli et al., 2011). Therefore, the function of food unavailability, food inaccessibility, inadequate utilized food and instability of food availability, accessibility and utilisation over time at household level, resulted in household food insecurity (Etana and Tolossa, 2017).

Physical availability of food: The availability refers to the physical existence of food. It addresses the "supply side" of food security and be it from own production or on the markets so that the supply is adequate, of appropriate quality, varied and contributes to a healthy diet. On national level, it is a combination of domestic food production, commercial food imports and exports, food aid, and domestic food stocks. Moreover, on household level, it could be from own production or bought from the local markets (Godfray et al., 2010). Different scholars employed different types of methods to measure availability dimension of household food security status. Various studies conducted by FAO used ‘‘dietary energy intake” as a measure food security in-terms of food availability (Coates, 2013). Moreover, the “Months of Adequate Household Food Provisioning (MAHFP)” has served as a measure of household food security status in-terms of availability dimension (Carletto et al., 2013, Moroda et al., 2018). On the other hand, both the total annual household production of corn and bean; and the total annual corn and bean consumption per capita have served as a measure of household food security status of the availability dimension (Coates, 2013, Méthot and Bennett, 2018).

Economic and physical access to food: The presence of food availability in the community does not mean that the household accessed the food so that the accessibility is the demand side of food security. The accessibility refers to the purchasing power of a household/individual, infrastructure and existing food price at national or regional level (FAO., 2014, Pinstrup-Andersen, 2009). The affordability of the food available at the region or community was evaluated by economic access to food and the deliverability of the available food to all people who need it was also captured by physical accessibility to food (Assefa, 2015). In general, access is ensured when all households and all individuals within those households have sufficient resources to obtain appropriate foods for a nutritious diet because an adequate supply of food (food production and availability) at the national or international level does not in itself guarantee household level food security (Assefa, 2015, Carletto et al., 2013, FAO., 2014, Hunnes, 2013). Various indicators of food access were employed to measure the status of food security including, the annual net household income (Méthot and Bennett, 2018); the Household Food Insecurity Access Scale (HFIAS) (Carletto et al., 2013, Coates, 2013, Etana and Tolossa, 2017, Méthot and Bennett, 2018, Moroda et al., 2018); the Household Hunger Scale (HHS) (Ballard et al., 2011, Méthot and Bennett, 2018); the Months of Adequate Home Food Provisioning (MAHFP) (Bilinsky and Swindale, 2007, Méthot and Bennett, 2018) and the Household Dietary Diversity Score (HDDS) (Méthot and Bennett, 2018, Swindale and Bilinsky, 2007).

Food Utilisation: Proper utilisation of food is very important because the availability and accessibility of food do not reflect the appropriate food security situation (Assefa, 2015). Utilisation is commonly understood as the way the body makes the most of various nutrients in the food (FAO., 2014). It is directly linked to a safe and adequate diet, water availability and quality, sanitation systems, and is influenced by water-borne, food-borne, vector-borne, and other infectious diseases (Hunnes, 2013, Pinstrup-Andersen, 2009). In addition, sufficient energy and nutrient intake by individuals is the result of good care and feeding practices, food preparation, and diversity of the diet and intra-household distribution of food (FAO, 2008, Hunnes, 2013, Pinstrup-Andersen, 2009). In general, it is socio-economic and biological aspects of food. The composite score of the "Household Dietary Diversity Score (HDDS)" and the facilities in-terms of access and use of water supply, sanitation, and hygiene (WASH) were suggested as a measure of food security indicators in the dimension of utilisation (Carletto et al., 2013, Moroda et al., 2018). On the other hand, Consumption Score (FCS) was employed as a measure of food utilisation dimension (Méthot and Bennett, 2018).

Stability: Stability depends on local and regional food production (food availability) and on the reliability and price of food imports (food access) (Cohen and Garrett, 2010, Hunnes, 2013). On the other hand even if one's food intake is adequate today, one is still considered to be food insecure if she has inadequate access to food on a periodic basis, risking a deterioration of your nutritional status (Hunnes, 2013). Therefore, stability depends on the availability, access and utilisation dimensions of food security (FAO., 2014).

2.3 Food Security Measurements and Determinant Factors

Different types of food security measurements were employed for different purposes. We reviewed recently employed food security measurements and the corresponding statistical models employed to determine associated factors of food security. The most used food security measuring tool and statistical models to determine its associated factors are "Household Food Insecurity Access scale (HFIAS)" and multivariable logistic regression model, respectively. HFIAS is used to assess the household food security status in-terms of food access (accessibility dimension) and effects to action since 2007 (Coates, 2013).

Several studies were conducted in Ethiopia that revealed different factors affecting household food (in)-security status using HFIAS and multivariable logistic regression model. First we reviewed the recent studies implemented using both of the two methods together and continues for other methods. All the researches reviewed herewith in this thesis that were measured by HFIAS method assessed the household food security access dimension. The study conducted in Addis Ababa by (Birhane et al., 2014) showed that lower monthly income of the household, household headed by uneducated household heads, daily labourers, and government employees were more likely to have higher food insecurity whereas households living in government rental houses were less likely to be food in-secured.

A community based cross-sectional study conducted in Farta District, Northwest Ethiopia indicated that households headed by females, lack of education, large family size, few or absence of livestock, absence of income from off-farm activities, lack of irrigation and lack of perennial income were identified as associated factors for food insecurity (Endale et al., 2014). Similarly, (Motbainor et al., 2016) conducted a community based comparative cross-sectional study in east and west Gojjam zones of Amhara Region and the results revealed that five or above family size, non-merchant women, household monthly income less than 560 ETB, illiterate mothers, rural residential area, highland agro-ecology and lack of livestock were positively affects household food insecurity. Moreover, (Shone et al., 2017) conducted a community-based cross-sectional study in West Abaya District, Southern Ethiopia and the results indicated that households headed by female, households headed by persons aged >65 years, households with larger family size and owning smaller farm land were increased the risk of being food in-secured.

On the other hand, the study conducted in Addis Ababa and Arisi Zone of Oromia Region by (Etana and Tolossa, 2017) showed that lower education status, poor economic status, unemployment status and study sites provided a statistical significant effect for households to be food in-secured. As studied by (Tantu et al., 2017) in Wolaita Sodo Town, the result indicted that single household head, greater than two dependent members, households headed with daily labourers, lower monthly income and low monthly food expenditure have positive and significant relationship with food insecurity. Moreover, (Abegaz, 2017) analysed the pooled data of the sixth and seventh round of the Ethiopian Rural Household Survey (ERHS) using binary multivariable logistic regression model revealed that rain shock, lack of off-farm income, and region of the households were negatively associated with food security. Moreover, as studied by (Abdu et al., 2018) in Assayita district in Afar region through multivariate regression models revealed that age, parity, and having >2 children below five years of age were statistically associated variables with household food insecurity.

The “Household Dietary Diversity Score (HDDS)” was also used to determine food security status as proxy indicator. (Moroda et al., 2018) conducted a study in Ethiopia using logistic regression models. They found that low educational status, small farmland size, small total annual income, far distance from health facilities, access to irrigable land, far distance to road transport, far distance to input/output markets, frequent drought and the in-availability of supporting organizations were positively associated with household food in-security situation in the utilisation dimension. This paper addressed both the determinants of household food accessibility and utilisation dimensions. Moreover, a study conducted in South Africa by (Ngema et al., 2018) using binary logistic regression revealed that education, and receiving infrastructural support (irrigation), positively influenced the food security status of households. However, household income and access to credit showed a negative correlation. The work addressed the general outlook of food security. It does not indicate which food security dimension was addressed.

The “coping strategies index” has served as a means of proxy variable to measure food security status. (Napier et al., 2018) conducted a study in Durban, South Africa using logistic regression model. They found that larger household size, households spending between R700 and R900 on food monthly and households purchasing food from street vendors or informal community shops were indicators of food in-security.

The “calorie intake” tool was also employed in Pakistan as a proxy food security indicator as conducted by (Ahmed et al., 2017) using binary logistic regression model. The results showed that family size, monthly income, food prices, health expenses, the market accessibility factors (road distance and transportation cost) and debt were identified as the main factors influencing the food security status of rural households. This work provided an input in the household food availability dimension.

The study conducted in the Sekyere-Afram Plains District of Ghana using both “USDA Household Food Security Scale” and a binary logit model revealed that households headed by unmarried people, large household size, small farm size, absence of off-farm income generating activities and farmers without access to credit were identified that leads households to be food in-secured (Mensah et al., 2013; Zeray, 2017). The work of Mensah et al. (2013) addressed the deteminats of food security intermis of household food access. Similarly, (Kelly and Pemberton, 2016) conducted a study in eastern rural area of Grand Bahama Island with the same procedure revealed that higher educational level of household head, high monthly income, and access to community gardens were statistically significant predictors for food security.

Habyarimana (2015) conducted a study in Rural Households in Rwanda using “Food Consumption Groups Score” and probit model. The study revealed that rural households headed by females, large household size, limited household's farm animal, and small household asset index, were significant variables for household food security. Moreover, the study also found that limited household food acquisition level, large household food acquisition problem, small amount of household spending level, small amount of monthly food expenditure, small percent of land suitability per cell, large amount of soil erosion index per village, reduced coping strategy index and being membership to agricultural cooperative were significant variable for household food in-security (Habyarimana, 2015). This research addressed the determinants for household food accessibility dimension.

The study was conducted to determine predictors of household food security in-terms of food access in Mexico using the “Mexican Food Security Scale” and Ordered probit model (Magaña-Lemus et al., 2016). They indicated that households include those with younger, less-educated household heads, headed by single, widowed or divorced women, with disabled household members, with native language speakers, with children, as well as rural and lower-income households were more likely to be food insecure.

Bashir and Schilizzi (2012) conducted a meta-analysis that showed education level, household head’s age, input availability, technology adoption, farm size, land quality, price of inputs, and credit were associated with household food security of the availability dimension. On the contrary, income, distribution of income within the household, household size, total earning members, and family structure were associated with the access dimension of household food security. Moreover, gender and expenditure on food and health are considered as determinants of utilisation aspect.

The studies reviewed so far addressed a single dimension among the four dimensions of household food security. The next review intended on the composite multidimensional index of food security. The most known composite index of food security analysis was principal component analysis index (PCAI). Abafita and Kim (2014) employed PCAI to compute the composite food security index of food availability, accessibility and utilisation in Ethiopia. An instrumental variable (IV) regression models using 2-Stages Least Square (2SLS) was applied to select the significant predictors and the findings indicated that participation in off-farm activities, education of the household head, household size, livestock possession, rainfall index, fertilizer use and per capita consumption expenditure were statistically significant determinants with positive impact on household food security. On the contrary, remittance and credit access had a negative and statistically significant impact on household food security (Abafita and Kim, 2014).

Similarly, (Mbolanyi et al., 2017) followed the procedure of (Abafita and Kim, 2014) for the study conducted in rangeland area of Uganda using Ordinary least square (OLS) and the result indicated that age of the household head, male household head, On-Farm Income and household head level of education (second degree or above) positively affected the food security of households. On the other hand, as studied by (Wineman, 2016) in rural Zambia on three food security components of the households (food quantity, food quality and food stability) using multinomial logistic regression. The author found that both rainfall and temperature have a significant impact on a household's food security score.

The situation of food security is very difficult as it is the result of complex interaction between numerous variables. For instance, some of the food access proxy indicators have served as food availability proxy indicators and vice versa. Similarly, some of food access proxy indicators have served for utilisation proxy indicators. This implies that understanding several concepts associated with the definition of food security are necessary before examining the determinants of food security.

The composite multidimensional index of food security conducted so far did not consider the contribution of each dimension for the determination of household food security status.

Similarly, previous studies did not address the associated determinate factors for each dimension at a time in a single model. The dependence between the dimensions and the respective predictors for each dimension did not undertake in a single model simultaneously.

These gaps can be seen in three ways taking all the dimensions together. First, the dependence between food availability, accessibility and utilisation and the predictors for each dimension can be addressed in the statistical models of multivariate ordinal data analysis. Second, the stability of the composite multidimensional index of food security of the three dimensions can be addressed using the statistical models of longitudinal ordinal data analysis. Lastly, the stability of the three food security dimensions over time and predictors for each dimension can be addressed using multivariate longitudinal ordinal data analysis.

Chapter Three

3. Data

3.1 Study Area and Population

This study was conducted in South Wollo Zone, one of the 11 zones in the Amhara Region State of Ethiopia. South Wollo is located in the North East of Ethiopia with latitude and longitude of 11°07'59.99" N 39°37'59.99" E. Dessie is the capital of the zone which is 401 kilometres away from the capital city of Ethiopia, Addis Ababa. South Wollo has a population of 2,518,862, of whom 50.4% and 49.6% are women and men, respectively. The largest ethnic group of the zone was Amhara which accounted for 99.33% of the total population. Moreover, 70.89% of the population subscribe to Muslim religion, and 28.8% were practising Ethiopian Orthodox Christianity (CSA, 2007).

South Wollo has 18 rural Woredas and two urban Woredas. Each Woreda has Kebeles which is a smallest unit in the administration of the zone. The target group for the thesis is households who are farmers living in the rural Woredas. This choice will minimise the error which will come from the heterogeneity lifestyle of the households because the living style and sources of food security dimensions are the same.

3.2 Data Collection Instrument

The quantitative data collection instrument was developed through extensive review of similar literatures and recent studies (Ballard et al., 2011; Bashir and Schilizzi, 2012; Carletto et al., 2013; Castro, 2000; Coates, 2013; Cohen and Garrett, 2010; Godfray et al., 2010; Hunnes, 2013; Napoli et al., 2011; Negatu, 2004; Ryu and Bartfeld, 2012; Biesalski et al., 2017; de Bruin and Gresse, 2018). Data were collected by using semi-structured questionnaire, which allowed study participants to express more additional information and their opinions.

The questionnaire has five parts; the first part covers area identification, the second covers demographic and socio-economic characteristics of the households, the third part covers farming activities in relation to agriculture activities, environmental and climate change conditions. The fourth part intended on the information related to food security status of the households in each dimension includes Availability, accessibility, utilisation, and the stability of the three dimensions over time. The availability dimension questions were customised from several researchers (Carletto et al., 2013, Coates, 2013, Godfray et al., 2010). The questions for accessibility used in this study was the Household Food Insecurity Access Scale (HFIAS) (Swindale and Bilinsky, 2007). The utilisation tool was obtained from (Carletto et al., 2013, Faber et al., 2009). The last part covers the copying strategy that the households applied to overcome the hardship and crises of food security. The questionnaire is first prepared in English and then translated into Amharic (local language of the respondent's) attached in appendix I.

3.3 Sampling Design and Procedure

Administratively, Ethiopia is divided into 11 regions. Subsequently, regions are divided into zones. Similarly, zones are further divided into Woredas, the smaller administrative unit. Each Woredas is further subdivided into the lowest administrative unit called Kebele. For the current study, South Wollo zone from Amhara region was selected. In this area, the food security situation has yet not been updated for the last 15 years after (Castro, 2000, Negatu, 2004). South Wollo zone has 18 and 2 rural and urban Woredas, respectively. The rural Woredas are assumed to have uniform agro-ecological and homogenous in cultivation strategies. Hence, three-stage sampling procedure is the ideal sampling methods. This implied that sample of primary units (Woredas) were selected from the total rural Woredas of south Wollo Zone, then sample of secondary units (Kebeles) were chosen from each of the selected primary units (Woredas) and finally, sample of tertiary units (households) were chosen from each selected secondary unit (Kebeles). Hence, three rural Woredas were determined as optimal sample size using the ordinary cluster sampling formula as

$$m = \frac{\left(Z_{\alpha/2}+Z_{\beta}\right)^2 MV^2}{\left(Z_{\alpha/2}+Z_{\beta}\right)^2 V^2+(M-1)d^2} = \frac{(1.96+0.84)^2(18)(0.001)}{(1.96+.84)^2(0.001)+(18-1)(0.05)^2} = 2.867 \approx 3.$$

Where $\left(Z_{0.05/2}+Z_{0.2}\right)^2 = (1.96+0.84)^2$ at 5% level of significant and 80% power, $M = 18$ is the number of rural Woredas in South Wollo zone, d is the degree of precision and taken to be 0.05 and $V^2 = 0.001$ is the ratio of the variance of the error term and the variance of the food security proportion $P = 0.60$ of the study conducted in Guraghe zone, Southern Ethiopia (Nigussie and Alemayehu, 2013).

These three Woredas were selected using simple random sampling (in particular lottery method) from 18 rural Woredas. Using the same fashion, a total of six Kebeles, 2 for each were selected from the selected Woredas. The sampling procedure conducted in this research was a three-stage sampling design. Woredas, Kebeles and households were the first, second and the third stage of sampling, respectively.

Hence, there were three sample units in the first stage, and six in the second stage. For the third stage, a complete list of household heads was obtained in each of the six selected Kebeles from agriculture agent office of each Kebele. The determined sample was proportionally allocated for each Woreda and then Kebele. Based on the allocated sample size, households included in the sample were selected using systematic random sampling technique from those representative Kebeles. Then a list of names of the sampled households was prepared for each Kebele.

3.4 Sample Size

Sample size determination is a very crucial task because a huge sample costs money and a small sample reduces the power of estimation. Hence, during the determination of required sample size issues/points one has to consider are objective of the research, design of the research, cost constraint, degree of precision required for generalization, etc. Based the above information, several sample size calculation formulas were developed that conform to different research situations. Accordingly, the sample size determination formula $n = \frac{Z^2 p(1-p)}{d^2}$ (Cochran, 2007) is adopted for this study since the target population is reasonably large. Where Z is the upper $\alpha/2$ points of standard normal distribution with α =0.05 significance level, which is Z =1.96. The degree of precision d is taken to be 0.05. The parameter p represents proportion of food security of household. P=0.60 is used in this study obtained from previous study in rural areas of Guraghe Zone, Southern Ethiopia (Nigussie and Alemayehu, 2013). Accordingly, the sample size using the given formula becomes n=369. Five percent of the sample size, which is 19, is added to the determined sample size 369 to compensate for non-response rate and the sample size becomes 388. Since the sampling design is multistage, 1.75 times of the sample size should be taken to compensate the design effect. Therefore, the required sample size for the study becomes ***n = 646***. Next, based on these 646 farmer households, the following sample size allocations were employed based on proportional allocation for the selected districts and then to Kebeles as presented in Table 3.1.

Table 3. 1 Sample size allocations for the selected Woredas and then to selected Kebeles with-in the respected Woredas

Woredas	Total farmer households	Allocated Sample size	Kebeles	Total farmer households	Allocated Sample size
Kutaber	27, 443	210	Alansha (03)	1245	139
			Beshilo (06)	639	71
Kalu	31,693	235	Kedida (07)	706	89
			Degan (019)	1150	146
Tehuledere	27, 241	201	Bededo (01)	1973	108
			Jari (017)	1692	93
Total	87, 377	646		7405	646

3.5 Data Collection

The data collection process has three phases. Each phase has similar procedures to be undertaken. Data collection was carried out using trained data collectors and data collection supervisors under the direct supervision of the researcher who worked closely with them. All the data collectors and data collection supervisors were trained on sample design, survey technique, survey instruments, and confidentiality protocol both for the pre-test and main data collection. This was necessary to ensure a common understanding of the whole survey in order to reduce interviewer biases as much as possible. Both data collectors and data collection supervisors were agricultural extension workers with a minimum of diploma for data collectors and Bachelor of Science for supervisors who speak English and local language Amharic.

Before administering the questionnaires, the questionnaires were pre and pilot tested for the purpose of insuring the questions were clear and understood by the study participants. The pilot test fieldwork was conducted over half a day in one Kebele which is out of the selected Kebeles from 50 household heads. The pilot test field staff and the investigator made thorough discussions. After a while, based on lessons drawn from the pilot test exercise, the questionnaires were modified.

During the data collection, the data collectors approached the sampled household heads and requested their willingness to be take part of the study before starting the interview. Only consented household heads were interviewed face-to-face that took place usually outside of the house in the compound. But if the head was not present or available, then the spouse or an adult household member aged 18 or more and live more than six months with the family to be considered as a member of that household was interviewed with the same fashion as the household head. Once data collection ended, the data collectors told the study participants before leaving about the second phase data collection as they will come after six-month interval. The third phase data collection proceeded like phases one and two. Each interview lasted on average 30 to 45 minutes. The surveys were carried out in a local language of the household head.

The supervisors were in charge verifying everyday what each data collectors had done (how the questionnaire was filled in, omission and coherence of answers, and sometimes assisted in interviews). This was very important because enumerators could quickly rectify any mistake that had occurred by going back to the households to verify the information from their subjects when it was necessary. The team (data collectors and supervisors) met the principal investigator every morning for field feedback and every two days for logistical support. The data collection lasted for 30 working days (exclusion of Sunday) and each interviewer had to administer seven to eight questionnaires per day.

The identification code was prepared for each household head participated in the study. An appointment abstraction form was also prepared to trace the name of the household heads and the study participation code that serves for the six-month follow-up data collection process (for the second and third round data collection processes). Once the data collection process is accomplished, the follow-up appointment abstraction form detached from the data collection questionnaire and placed in the separate place to secure confidentiality issues.

3.6 Data Collection Periods

The current study was employed the longitudinal data collection approach. Three rounds of data collection were employed at six months interval. The main harvest season in most of the study locations is during the months of June and July. The first round of data collection that assessed the food security experience for the last 12 months was carried out on June, 2014. The second-round data collection that assessed the last six months food security experiences was carried out on December, 2014. Lastly, the third round of data collection that assessed the last 6 months food security experiences was employed on June, 2015.

3.7 Measuring Food Security

In this thesis, different household food security measuring methods were employed with the hope to have single index for availability, accessibility and utilisation. Moreover, a composite index was also employed to assess the stability of household food security status over time.

Measuring Household Food Security Status of the Availability Dimension

Different scholars or organisations employed different methods and food security classification for this dimension. The “Dietary Energy Intake” method is used to determine food security status as food secured or in-secured (Coates, 2013, FAO., 2014). The “Months of Adequate Household Food Provisioning (MAHFP)” is used also to determine least food insecure, moderately food and most food insecure (Carletto et al., 2013, Moroda et al., 2018). The median score is used to determine food security status, those below median score as food in-secured and above median score as food secured (Kisi et al., 2018). Based on the recommendation obtained from (Capaldo et al., 2010), we expanded the work of Kisi (Kisi et al., 2018) using the quarter score approach to determine food security as “food secured”, “mildly food in-secured”, “moderately food in-secured” and “severe food in-secured as follows.

Food availability at household level depends on own production or bought from the local markets (Godfray et al., 2010). Coates (2013) used the total annual household production and consumption of corn and bean per capita as a proxy measure of household food security status for the availability dimension.

In the study area, foods like prepared from cereal crops, fruit and vegetables, milk and milk products, and meat and meat products are more or less consumed from their own production or from the local market. Moreover, food availability depends on foods provided by food aid organisations. Twelve (12) questions were developed to assess the availability of the above food groups from their own production and/or from local markets; and from food aid organisations to measure the household food security status in terms of availability. We created a summative scale using these questions and each answer was recoded as 1, 2, 3 and 4 where 1 stands for the response "enough of the kinds of food we want to eat", 2 for "enough but not always the kinds of food we want", 3 for "sometimes not enough to eat" and 4 for "often not enough to eat".

Screening questions were used before asking the availability of food groups from their own production. If a household is not produce a particular food group or some of them, he or she does not ask about the availability of food group from own production. Similarly, a household obtains "enough of the kinds of food he or want to eat" from own production is not asked about the availability of that food group from local market. This implies that the number of question for each household may not be equal and may be less than 12 for some of them.

Based on the above criteria, the item responses were summed to compute the score of household food security status ranging between 12 and 48 points for those asked all of the 12 questions. This range divided into four equal parts based on quartile score. The scores fall in the range 12-20 grouped as "food secured", 21-29 as "mildly food in-secured", 30-38 "moderately food in-secured" and 39-48 as "chronically food in-secured". Similarly, for those asked 11 questions, the range is 11 and 44 points, for those asked 10 questions, the range is 10 and 40 points and so on for the other households asked less than 10 questions. The quartile square is applied for each range to determine the household food security status.

Measuring Household Food Security Status of the Accessibility Dimension

Majority of food security studies relayed the Household Food Insecurity Access Scale (HFIAS) for measuring household food security access status. Hence, we followed this scale for this thesis to assess the household food insecurity status in terms of accessibility. The module consists of nine items that measure the severity of a wide range of food hardships over the past 12 months. The status of the households was classified into four as availability based on the criteria given in the module. The cut-off points that serve to place households in a unique category of food security status is given in Table 3.2.

Table 3. 2: Cut-off points for household access scale

Questions	Frequencies		
	Rare	Sometimes	Often
1	Food secure	Mildly food insecure	Mildly food insecure
2	Mildly food insecure	Mildly food insecure	Mildly food insecure
3	Mildly food insecure	Moderately food insecure	Moderately food insecure
4	Mildly food insecure	Moderately food insecure	Moderately food insecure
5	Moderately food insecure	Moderately food insecure	Chronically food insecure
6	Moderately food insecure	Moderately food insecure	Chronically food insecure
7	Chronically food insecure	Chronically food insecure	Chronically food insecure
8	Chronically food insecure	Chronically food insecure	Chronically food insecure
9	Chronically food insecure	Chronically food insecure	Chronically food insecure

Food secure
Moderately food insecure
Mildly food insecure
Chronically food insecure

Measuring Household Food Security Status of the Utilisation Dimension

The composite score of the "Household Dietary Diversity Score (HDDS)" and the facilities in-terms of access and use of water supply, sanitation, and hygiene (WASH) were suggested as a measure using the three food security categories in this dimension (Carletto et al., 2013, Moroda et al., 2018). The DDS score was used to determine food security based on the median score those below the median score classified as food in-secured and above median as food secured (Faber et al., 2009). Based on the recommendation obtained from Capaldo et al. (Capaldo et al., 2010), we expanded the work of Faber et al. (Faber et al., 2009) using the quarter score approach to determine food security as "food secured", "mildly food in-secured", "moderately food in-secured", and "severe food in-secured" as follows.

Utilisation is directly linked with safe and adequate diet; and access and use of water supply, sanitation, and hygiene (WASH). "Household Dietary Diversity Score (HDDS)", which is an assessment of 12 food groups, can measure the safe and adequate diet. As a result, we developed

19 questions that can address these issues and each answer was recoded as 0 (no) and 1(yes). The first 12 questions were the HDDS component and the last seven were the WASH component.

As in food availability, the item responses in utilisation were summed to compute the score of household food security status ranging between 0 and 19 points. The scores fall in the range 15-19 classified as “food secured”, 10-14 as “mildly food in-secured”, 5-9 as “moderately food in-secured”, and 0-4 as “chronically food in-secured”.

Composite Food Security Index (CFSI)

The main objective of computing composite index was to determine the stability of household food security over the successive time periods. Three rounds of data collections were made. In each round, the household food security was measured for each dimension. To determine the stability of household food security status over time the food security measure in each round should be combined into one. We call this a composite food security index. Therefore, the following approach was made to compute the composite food security index.

In each of the data collection phase, there are three food security measures that have four levels namely “food secured”, “mildly food in-secured”, “moderately food in-secured”, and “chronically food in-secured”.

We created a summative scale using the three measures of phase one and each answer was recoded as 1, 2, 3 and 4 where 1 stands for the response “chronically food in-secured”, 2 for “moderately food in-secured”, 3 for “mildly food in-secured” and 4 for “food secured”. The item responses were summed to compute the score of household food security status ranging between 3 and 12 points. This range divided into four equal parts based on quartile score. The scores fall in the range 3-5 grouped as “chronically food in-secured”, 6-7 as “moderately food in-secured”, 8-9 as “mildly food in-secured”, and 10-12 as “food secured”. The same procedure was made for phase two and three to compute the combined food security status of the household. The framework of this computation displayed in the following diagram.

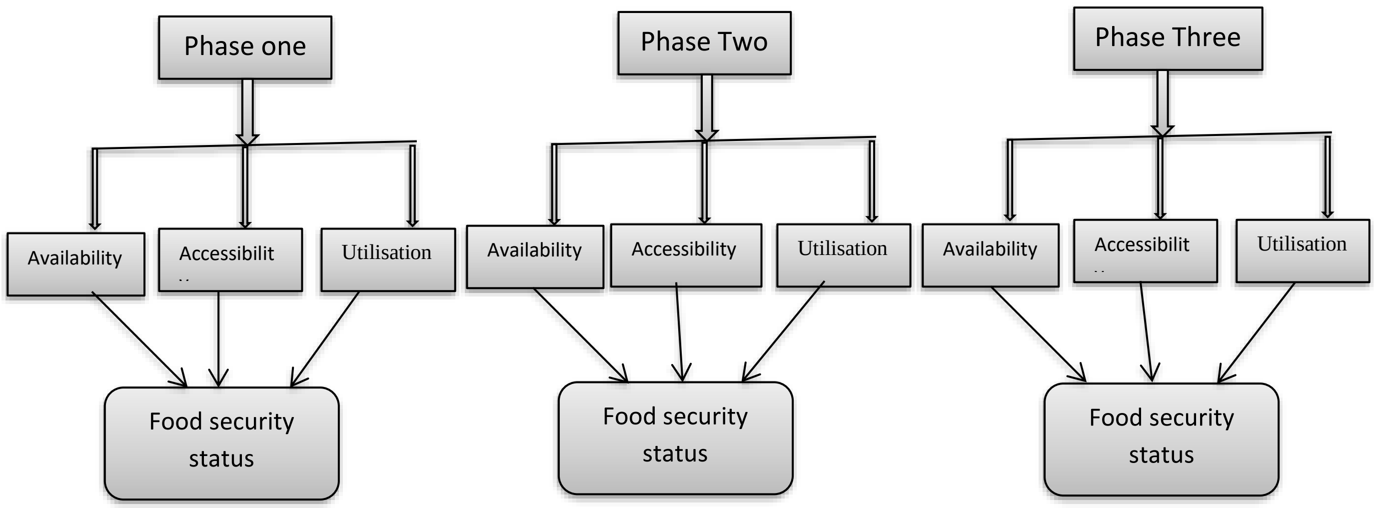


Figure 3. 1: Framework of food security of the household at three follow-up time points at six-month interval

3.8 Characteristics of Some Study Variables

Some of the determinant variables for food security are clarified as follows: Shortage of rain fall is described by the length that the rain rains; Shortage of rainfall happens if the rain stops too late and/or too early. Since the amount of rainfall depends on the nature of cultivable land, some type of land may need a high amount of rainfall or some of the lands will need few amount of rainfall. Hence, the farmers can declare the amount of rainfall happened in their village as per their type of cultivable land.

The crops/vegetables disease can be described as any type of disease reported considered as crops/vegetables disease occurred in that area. The type of weather conditions in the study area (Hot, Medium and Cold) was replaced by type of agro-ecology of the study site.

Cultivation season is the amount of cultivation season per year in their majority cultivable land (one time, two or more time per year). Moreover, the cultivable land of the study sites categorised as less than or equal to half hectare and above half hectare. Since majority of the farmers in the study site have less than half hectare cultivable land. Majority of other previous studies categorized as we have conducted.

Fertility of cultivable land is categorised as fertile, medium fertile, and less fertile. The concept of this question was to assess the fertility of majority of cultivable land. The fertility depends on the nature of the land that cultivated in the rainy season or not, resist during the dry season and the amount cultivated from that specific cultivable land. Moreover, the data collectors and supervisors were diploma and BSc in agricultural science, respectively; working around the farmers elaborated the categories in detail during data collection.

Chapter Four

4. Methodology

4.1 Introduction

For the long period of time, statisticians have been searching models that serve for measuring relationships or associations both in the continuous and discrete analogs of multivariate distribution. Mainly measuring relationships revolved on the continuous analogs whereas association on the discrete using bivariate and tri-variate distribution functions with given univariate margins. Sklar (1959), who develops new class of functions called copulas, gave the concept of univariate margins for this case. The word copula originates from the Latin word *copulare*, which means a link or connect or join. It was used by Sklar (1959) in the theorem for the first time in a mathematical or a statistical context to describe multivariate distribution functions that are constructed by joining together one-dimensional distribution functions (Sklar, 1959).

For the first time, copula was introduced to Encyclopaedia of Statistical Sciences in 1997 by (Fisher, 1997). Fisher introduced the interest of copula to statisticians on the concept of probability and statistics in this Encyclopaedia for two main reasons. The first reason is to measure scale free dependence and the second reason is to construct the starting point of bivariate distributions families (as cited by Nelson, 2007). “Copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval (0, 1)”.

Various advancements of the copula functions have been introduced since its introduction for many applications, especially, multivariate distributions. One of the recently developed copulas for cascading the multivariate distribution into bivariate distribution with the great accuracy and efficiency is the pair copula construction (PCC). Pair copula construction was first introduced for continuous margins (Aas et al., 2009), then extended for discrete margins (Panagiotelis et al., 2012). The current focus is the applicability of pair copula construction on multivariate longitudinal ordinal outcomes with the hope of measuring the dependency between longitudinal ordinal outcomes. This thesis further concerned itself on the implementation of PCC for multivariate longitudinal ordinal outcomes using the frequentist paradigm because the Bayesian paradigm requires intensive work in selecting the appropriate prior distribution for the marginal model especially in ordinal outcomes setting.

In sum, this chapter addresses the basics of copula theory, the application of pair copula construction for multivariate ordinal outcomes, for longitudinal ordinal outcomes, and for multivariate longitudinal ordinal outcomes. The chapter also addresses parameter estimation of all the three models stated above.

4.2 Basics of Copula Theory

4.2.1 Definition and Properties of Copula Theory

Definition 4.1: A multivariate distribution function with uniformly distributed margins that satisfies the following properties is called an m- dimensional copula C: $[0,1]^m \rightarrow [0,1]$.

i. For every **u** in $[0,1]$

   a. $C(u_1, u_2, ..., u_m) = 0$, if any $u_i = 0$

   b. $C(u_1, u_2, ..., u_m) = 1$, if any $u_i = 1$

ii. For any $(a_1, a_2, ..., a_m)$ and $(b_1, b_2, \ldots, b_m) \in [0,1]^m$ if $a_j \le b_j$, then $P(U_1 \in [a_1, b_1], ..., U_m \in [a_m, b_m]) \ge 0$, we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_m=1}^{2} (-1)^{i_1+..+i_m} C(u_{1,i_1}, ..., u_{m,i_m}) \ge 0$$

$where\ \ u_{i,1} = a_j,\ u_{i,2} = b_j$ and U_j are the random numbers which have uniform margins. The first property expresses the requirement of uniform marginal distributions whereas the second property expresses the rectangle inequality. A copula characterizes through these two properties; meaning that if a function C is fulfilled then it is a copula.

Sklar's theorem summarized the importance of copula in the study of multivariate distribution functions. The theorem shows how the copula coupled the univariate marginal distributions to construct multivariate distributions.

Theorem 4.1-Sklar (1959): For m-dimensional random variables $(y_1, y_2, \dots, y_m)$ with corresponding margin $F_1(y_1),\ F_2(y_2),\dots,F_m(y_m)$, their joint distribution function $F(y_1, y_2,\dots, y_m)$ can be expressed in-terms of an m-copula C functions as follows;

$$F(y_1, y_2 \dots, y_m) = C(F_1(y_1), F_2(y_2),\dots, F_m(y_m)). \tag{4.1}$$

If all $F_j(y_j)$ are continuous then unique function C is defined, if not, C is uniquely defined within the product of the regions $RanF_1(y_1)\times\ldots\times RanF_m(y_m)$, where $RanF_i(y_i)$ is the range of the ith distribution function.

Equation (4.1) gives an expression for joint distribution functions in terms of a copula and univariate distribution functions. But (4.1) can be inverted to express copulas in terms of a joint distribution function and the “inverses” of the margins. However, if a margin is not strictly increasing, then it does not possess an inverse in the usual sense. Therefore, we can use “quasi-inverses” of distribution functions.

Corollary 4.1: Let F be an m-dimensional joint distribution function with margins $F_1, F_2,\dots,F_m$, C be an m-copula and let $F^{(-1)}$ be the ith quasi-inverses of F. Then for any **u** in domain of C,

$$C(u_1,\dots,u_m) = C(F_1^{(-1)}(u_1),\dots, F_m^{(-1)}(u_m)). \tag{4.2}$$

This corollary is the unique copula satisfying equation (4.1). Given marginal and joint cumulative distribution functions, the above result allows the direct construction of a copula.

In copula theory, there are special dependence structure functions. These are the “Frechet-Hoeffding upper bound *M* named the **comonotonicity copula”**, the “Frechet-Hoeffding lower bound *W* named the **countermonotonicity** copula”, and the “independence copula *Π*” (Nicklas, 2013). There expressions are given below respectively.

$$M(u_i,\dots,u_m) = C(u_i,\dots,u_m) = \min(u_i,\dots,u_m)$$
$$W(u_i,\dots,u_m) = C(u_i,\dots,u_m) = \max(u_i + \dots + u_m - m + 1, 0)$$
$$\prod(u_i,\dots,u_m) = C(u_i,\dots,u_m) = \prod_{i=1}^{m} u_i \ .$$

Note that in arbitrary dimensions, M and Π are copulas whereas W is a copula only in the bivariate dimensions. Any copula functions are bounded point-wise by the Frechet-Hoeffding bounds.

Proposition 4.1 For every u ∈ $[0, 1]^m$ in any copula function C, the following expression holds.

$$W(u_i,\ldots,u_m) \le C(u_i,\ldots,u_m) = M(u_i,\ldots,u_m) \tag{4.3}$$

In copula theories whenever *M*, *Π*, and *W* are copulas, they have a special interpretation as stated in detail in (Nelsen, 2007).

4.2.2 Copula Density

The multivariate density $f(y_1, y_2, \ldots, y_m)$ for the continuous case can be obtained through both sides' differentiation of equation 4.1 using the chain rule, we have

$$f(y_1, y_2, \ldots, y_m) = c(F_1(y_1), F_2(y_2), \ldots, F_m(y_m)) \times f_1(y_1) \ldots f_m(y_m) \tag{4.4}$$

where $f_1(y_1) \ldots f_m(y_m)$ are the marginal density of the j^{th} margin and c(.), known as the copula density, is the copula function differentiated with respect to each of its arguments.

Even if the copula function is not unique for discrete margins, parametric copulas may still be used to model the dependence between discrete data which provides some evidence that discrete data inherit dependence properties from a parametric copula like the continuous case. In contrast with the continuous case, the probability mass function (pmf) for discrete data can be evaluated by taking differences of the copula function. Without loss of generality assuming Y ∈ N^m (where N is the set of natural numbers), the probability mass function of Y is given (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015);

$$\begin{aligned} P(Y = y) &= \sum_{i_1=0,1} \ldots \sum_{i_m=0,1} (-1)^{i_1+\ldots+i_m} P(Y_1 \le y_1 - i_1, \ldots, Y_m \le y_m - i_m) \\ &= \sum_{i_1=0,1} \ldots \sum_{i_m=0,1} (-1)^{i_1+\ldots+i_m} C(F_1(y_1 - i_1), \ldots, F_m(y_m - i_m)). \end{aligned} \tag{4.5}$$

Equation (4.5) is the special class of copula distribution called multivariate copula function for discrete data, which requires 2^m evaluations of the pmf.

Alternately, the pmf for some copula functions that do not have a closed form can be evaluated by integration over a rectangle. Hence, the probability mass function of the Gaussian copula with discrete margins can be expressed as;

$$P(Y_1 = y_1, ..., Y_m = y_m) = \int_{\psi_1^-}^{\psi_1^+} ... \int_{\psi_m^-}^{\psi_m^+} \phi_m(\psi_1, ..., \psi_m; \Gamma) d\psi_1, ..., d\psi_m, \tag{4.6}$$

where $\psi_j^- := \Phi^{-1}(p(Y_j = y_j - 1)), \quad \Phi_m(.,\Gamma) \quad and \quad \phi_m(.,\Gamma)$ respectively denote the cdf and probability density function of an m-dimensional normal distribution with mean 0 and variance matrix given by the correlation matrix Γ (Panagiotelis et al., 2012).

Both the multivariate and Gaussian copulas remain a highly challenging computational problem, especially for higher dimensions. Furthermore, this computational challenge was resolved through the introduction of vines pair copula construction (PCC) which requires 2m(m-1) evaluation of the pmf less demanding than 2^m of the former one (Panagiotelis et al., 2012, Huynh et al., 2014, Sirisrisakulchai and Sriboonchitta, 2014). This is a copula-based framework that effectively simplifies the computational cost of evaluating the pmf and also a large range of dependence characteristics can be modelled. PCC will be discussed in detail later on section 4.2.3.3.

4.2.3 Families of Copulas

Several scholars in their literature have carried out the construction of copula families and their properties. Here, we present few of the most popular in the literature and use for our purpose as follows:

4.2.3.1 Elliptical Copulas

Copulas developed from elliptical distributions are elliptical copulas. The elliptical copulas can be used to create new multivariate distribution functions by combining arbitrary margins. The

Gaussian and the t-copulas are the most commonly used elliptical copulas. The properties of these copulas were presented below one by one.

Gaussian Copulas

Let the distribution function of the multivariate normal distribution with zero mean and correlation matrix P is denoted by $\Phi_{1\ldots m}$ and the univariate standard normal distribution is by Φ. Then the m-dimensional Gaussian copula is defined by

$$C^{Ga}(u_1,\ldots,u_m) = \Phi_{1,\ldots,m}(\Phi_1^{(-1)}(u_1),\ldots,\Phi_m^{(-1)}(u_m)) \qquad (4.7)$$

Even if we can express Gaussian copula as an integral, it does not have simple closed form (Nicklas, 2013). In two dimensions, given that the covariance matrix is non-singular, we get

$$C^{Ga}(u_1,\, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{(1-p^2)}} \exp\left\{\frac{-(x_1^2 - 2px_1x_2 + x_2^2)}{2(1-p^2)}\right\} dx_1 dx_2 \qquad (4.8)$$

The dependence structure can be extracted from the multivariate normal distribution through this Gaussian copula. We can obtain the independence copula from Gaussian copula if $P = I_m$. Similarly, the comonotonicity copula can be also obtained if P is an m×m matrix consisting entirely of ones. In two dimensions, the Gaussian copula with $\rho = -1$ is equal to the countermonotonicity copula. Hence, at least in two dimensions, the dependence structure that interpolates between perfect positive and negative dependence can be thought as dependence in the Gaussian copula.

The t copula

Let the distribution function of the m-dimensional t distribution with ν degrees of freedom, zero mean vector and correlation matrix P denoted by $t_{v,1,2,\dots,m}$ and the univariate t distribution by t_v. Then the m-dimensional t copula is defined by

$$C_v^t(u_1,\dots,u_m) = t_{v,1,\dots,m}(t_v^{-1}(u_1),\dots,t_v^{-1}(u_m)), \tag{4.9}$$

In t copula, there is an additional parameter v (degrees of freedom) (Nicklas, 2013). Owing to this parameter t, copula becomes more suitable for financial application. This parameter controls the dependence of the extreme events meaning that both extreme positive and extreme negative events can be modelled equivalently.

Like the Gaussian copula, we can express t copula as an integral, and does not have simple closed form. In two dimensions, the t copula with ν degrees of freedom has the following form:

$$T_{p,v}(u_1,u_2) = \int_{-\infty}^{t_v^{-1}(u_1)} \int_{-\infty}^{t_v^{-1}(u_2)} \frac{1}{2\pi\sqrt{(1-p^2)}} \left(1 + \frac{x_1^2 - 2px_1x_2 + x_2^2}{v(1-p^2)}\right)^{-\frac{v+2}{2}} dx_1 dx_2 . \tag{4.10}$$

As in the case of the Gaussian copula, the comonotonicity copula can be obtained if P is an m×m matrix of ones. However, in contrast to the Gaussian copula, we do not obtain the independence copula from t copula if $P = I_m$. This is because uncorrelated multivariate t-distributed random variables are not necessarily independent.

4.2.3.2 Archimedean Copulas

Among the copula families in parametric dependence modelling, the most popular on is the class of Archimedean copula. Elliptical copulas have the advantage that simulating from them is easy. However, they often do not have closed-form representations and they are all radially symmetric. All Archimedean copulas are flexible in the types of dependence structures they can model, and have closed-form expressions. Unlike the previously described copulas, Archimedean copulas are not derived using marginal distributions and Sklar's theorem, though they are still easy to construct. The uniqueness of Archimedean copulas are defined by a generating function denoted by ϕ (Nicklas, 2013). Under a continuous, strictly decreasing, and convex function from I to [0, ∞] of ϕ with $\phi(1) = 0$, then the Archimedean copula is given by

$$C(u,v) = \phi^{[-1]}(\phi(u), \phi(v)) \tag{4.11}$$

where $\phi^{[-1]}$ is the pseudo-inverse of ϕ:

$$\phi^{[-1]}(t) = \begin{cases} \phi^{[-1]} & if \;\; 0 \le t \le \phi(0) \\ 0 & if \;\; \phi(0) \le t \le \infty \end{cases} \tag{4.12}$$

Since an Archimedean copula can be generated by any continuous, strictly decreasing, convex function, a huge number of Archimedean copulas have the ability to model a wide range of dependence structures. For a complete summary of families of one parameter, Archimedean copulas, refer to (Nelsen, 2007). Moreover, hereunder we give the generator and the selected bivariate one parametric Archimedean copula functions that we used only for our purpose in this study.

Clayton copula (Clayton, 1978): given that the copula parameter $\theta \in [-1, \infty] \setminus \{0\}$, the Clayton copula is given by

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$$, with generator $\phi(t) = \frac{1}{\theta}(t^{-\theta} - 1).$

For $\theta = 0$, we set $C = \Pi$ (Nicklas, 2013).

Gumbel copula (Gumbel, 1960): For the copula parameter θ ∈ [1, ∞), the Gumbel copula is defined as

$$C(u,v) = \exp\left(-\left((-\log(u))^{\theta} + (-\log(v))^{\theta} \right)^{\frac{1}{\theta}} \right),$$

With generator $\phi(t) = (-\log(t))^{\theta}$ (Nicklas, 2013).

Frank copula (Frank, 1979): For the copula parameter θ ∈ (−∞, ∞)\ {θ}(Aas et al.) the Frank copula is defined as

$$C(u,v) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right),$$

with generator $\phi(t) = -\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right).$

Again, we set C = Π for θ = 0(Nicklas, 2013).

Ali-Mikhail-Haq Copula (Ali et al., 1978): For copula parameter θ ∈ [−1, 1], The Ali-Mikhail-Haq is defined as

$$C(u,v) = \frac{uv}{1-\theta(1-u)(1-v)}$$

with generator $\phi(t) = \ln[1-\theta(1-t)]/t$

Note: Among the 22 Archimedean copulas, AMH copula is the only copula whose parameter lies on [-1, 1] and measures both, positive and negative, dependence.

4.2.3.3 Pair Copula Construction

Another family of copula called a pair-copula construction (PCC) was developed using only bivariate copulas to construct a general construction method for multivariate copulas. The classes of multivariate copulas that we have discussed so far are limited in modelling various dependence structures. Among the flexible multivariate dependence structures which are needed especially in financial applications were the centre of the distribution and the tails (the upper and lower tail) dependence parameters. The application of copula for modelling purpose has been applied to many areas including actuarial sciences, finance, neuroscience, and weather research as cited in Kim et al., (2013). Among the copula-based models which are parametric copula families, elliptical and Archimedean copulas were commonly applicable for the application areas stated above. These families are limited in some aspect like tail dependence and dependence flexibility. For instance, from elliptical copulas, the Gaussian copula allows for an arbitrary correlation matrix with zero tail dependence while the tail dependence parameter from the multivariate t-copula were driven from only a single degree of freedom parameter (Fang et al., 2002, Frahm et al., 2003).

Besides to fill the gaps of elliptical copulas, Archimedean copula classes, namely, fully and partially nested Archimedean copulas, Hierarchical Archimedean copulas and Multivariate Archimedean copulas were considered by several researchers (Joe, 1997, Nelsen, 2007, Savu and Trede, 2006, Schirmacher and Schirmacher, 2008). However, these extensions require additional parameter restrictions. These additional parameters reduced the flexibility of the extended copula functions for modelling dependence structures. To overcome the limitations of existing copula-based models, a vine copula-based model has been developed. This vine is called pair-copulas. This copula-based model can express a multivariate copula by using a cascade of bivariate copulas.

The first pair-copula construction of a multivariate copula for the continuous data were introduced by Joe (1996) in terms of distribution functions while Bedford and Cooke (2001; 2002) expressed these constructions in terms of densities and graphical way involving a sequence of nested trees, which they called regular vines. The two popular subclasses of PCC models, which are called Drawable vines, or D-vines and Canonical vines or C-vines were also identified by Bedford and Cooke. Aas et al. (2009) and Czado (2010) also conducted different extensions in the continuous data. Even though Genest and Neslehova (2007) provided some evidence that discrete data inherit dependence properties from a parametric copula in a similar way to the continuous case, Panagiotelis et al. (2012) provided different PCC models for discrete data.

To overcome these problems, pair copula have been developed first by (Joe, 1996) and extended by different scholars like (Aas et al., 2009, Bedford and Cooke, 2001, Bedford and Cooke, 2002, Czado, 2010) for continuous data and for discrete data (Panagiotelis et al., 2012). One of the contributions of Pair Copula Constructions (PCCs) in the construction of multivariate copula was to provide a highly flexible framework for constructing copulas exhibiting a wide range of dependence characteristics. This flexibility arises since any combination of bivariate copulas can be used to construct PCC models (Czado, 2010). Since PCC is the concern of the current study, details of PCC is presented in the coming consecutive topics in sections 4.3, 4.4 and 4.5.

4.2.4 Dependency Measures

For understanding complicated dependence structure, measures of dependence are the most commonly used instruments. Among commonly used measures of dependence, the most popular is Pearson's correlation coefficient. Under strictly increasing linear transformations, Pearson's correlation coefficient is invariant but not under non-linear transformations. It is also defined only for pairs of random variables with finite variances, but this can bring problems when working with heavy-tailed distributions. Therefore, measuring dependence by standard correlation is adequate in the context of multivariate elliptical distributions. There are increasing proportions of nonlinear risks like the non-normal behaviour of most financial time series. As a result, other tools are needed since estimates of risk dependence via linear correlation neglects nonlinearities and leads in most cases to underestimation of the global risk. Since copula is often the key issue for numerous models in relation to the above limitations, it is very important to find the copula that describes the complete dependence structure. For detail on these, we refer to several researchers (Fan, 2009, Genest and Nešlehová, 2007, Nelsen, 2007, Pirktl, 2007, Nicklas, 2013).

Since the copula functions are invariant under strictly increasing transformations, it makes sense to consider dependence measures which are also invariant under such transformations. Kendall's tau and Spearman's rho are the most widely known scale-invariant measures of association. Both measure the form of dependence known as concordance.

4.2.4.1 Measure of Concordance

Pair of random variables is said to be concordant, if large values of variable is associated with large values of the other variable and small values of one with small values of the other. On the other hand it is discordant, if large values of one variable are associated with small values of the other. A more formal definition is the following:

Definition: Consider, (x_i, x_j) and (y_i, y_j) two observations from a pair random vectors (X, Y). We say that (x_i, x_j) and $(y_i, y_j)$are concordant if $(x_i - x_j)(y_i - y_j) > 0$and discordant if $(x_i - x_j)(y_i - y_j) < 0$.

Definition: Kendall's tau for the pair random vectors (X, Y) and whose copula C is given by

$$\begin{aligned}\tau &= P((x_i - \mathrm{x}_j)(\mathrm{y}_i - \mathrm{y}_j) > 0) - P((x_i - \mathrm{x}_j)(\mathrm{y}_i - \mathrm{y}_j) < 0)\\ &= 4\int_{[0,1]^2} C(u,v)dC(u,v) - 1\end{aligned}$$

Hence Kendall's tau is the probability of concordance minus the probability of discordance (Nelsen, 2007).

Definition: Spearman's rho for the random vector (X, Y) and the copula C is given by

$$\begin{aligned}\rho(X,Y) &= 3P((x_i - \mathrm{x}_j)(\mathrm{y}_i - \mathrm{y}_j) > 0) - P((x_i - \mathrm{x}_j)(\mathrm{y}_i - \mathrm{y}_j) < 0)\\ &= 12\int_0^1\int_0^1 \mathrm{C(u, v)}dudv - 3.\end{aligned}$$

Kendall's tau and Spearman's rho have many common properties. They can measure the degree of monotonic dependence between random variables. Both are taking values in [-1, 1] which measures symmetric dependence. They have the value 1 when X and Y are "comonotonic" and have -1 when they are "countermonotonic", the value [-1, 1] does not necessarily imply that all those values can actually be obtained by a particular copula.

4.2.4.2 Tail Dependence

The idea of Kendall's tau and Spearman's rho is to measure the dependence of the copula on the event space (0, 1) (Nicklas, 2013). On the contrary, there are cases that measure the dependence between the variables in the upper tail or the lower tail of the bivariate distributions. This is called tail dependence, which measures the dependence of extreme events. Nelson (2007) defines tail dependence for a copula as follows. A random variable X and Y have marginal distribution functions $F_X(x)$, $F_Y(y)$ and the copula function C, the lower tail dependence coefficient is given by

$$\lambda_L = \lim_{u \downarrow 0} \Pr\left(Y \le F_Y^{-1}(u) \mid X \le F_X^{-1}(u)\right)$$
$$= \lim_{u \downarrow 0} \frac{C(u,v)}{u}$$

and dependence coefficient for the upper tail is given by

$$\lambda_U = \lim_{u \uparrow 1} \Pr\left(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)\right)$$
$$= 2 - \lim_{u \uparrow 1} \frac{1 - C(u,v)}{1-u}$$

Note that C has lower tail dependence if $\lambda_L \in (0,1)$ and no lower tail dependence if $\lambda_L = 0$. Similarly, C has upper tail dependence if $\lambda_U \in (0,1)$ and no upper tail dependence if $\lambda_U = 0$.

4.2.4.3 Dependencies Characteristics of Bivariate Copula Families

Previous sections have introduced Kendall’s tau as a measure of dependence. The parameters of the copula and their values of Kendall’s tau have determined relationships as presented in Table 4.1. The table also includes both the upper and lower tail dependence properties for each copula family.

The upper or lower tail dependence cannot be treated in the Gaussian as well as in the Frank copula. Nevertheless, the t copula treats both cases and it is represented by $\lambda_U = \lambda_L$. On the other hand the Clayton can be used to model lower tail dependence whereas the Gumbel copula for upper tail dependence.

Table 4. 1: Kendall's tau, upper and lower tail dependence for bivariate copula families (Dissmann, 2010).

Copula	Kendall's tau	Upper tail dependence	Lower tail dependence
Gauss	$\tau = \frac{2}{\pi} \arcsin e(\rho)$	$\lambda_u = 0$	$\lambda_L = 0$
T	$\tau = \frac{2}{\pi} \arcsin e(\rho)$	$\lambda_u = \lambda_L = t_{v+1}\left(-\sqrt{v+1}\sqrt{\frac{1-\rho}{1+\rho}} \right)$	
Clayton	$\tau = \frac{\delta}{\delta + 2}$	$\lambda_u = 0$	$\lambda_L = 2^{-1/\delta}$
Gumbel	$\tau = \frac{\delta - 1}{\delta}$	$\lambda_U = 2^{-1/\delta}$	$\lambda_L = 0$
Frank	$\tau = 1 - \frac{4}{\delta} + 4\frac{D_1(\delta)}{\delta}$	$\lambda_u = 0$	$\lambda_L = 0$
Ali-Mikhail-Haq	$\tau = \frac{3\theta - 2}{3\theta} - \frac{2(1-\theta)^2 \ln(1-\theta)}{3\theta^2}$		
$D_1(\delta) = \int_0^\delta \frac{x/\delta}{\exp(x) - 1} dx$, being the Debye function.			

4.3 A Pair Copula Construction Approach for Multivariate Ordinal Data

This section provides pair copula-based cumulative logit model for jointly modelling the dependence between availability, accessibility and utilisation of food security dimensions and their respective determinants. The quartile score computed for each dimension categorized the food security status for each dimension as "severe food in-secured", "mildly food in-secured", "moderately food in-secured", and "food secured". This computation resulted in three ordinal dependent variables, namely, availability, accessibility and utilisation. Therefore, for assessing and interpreting food security status and determinant factors, a well-defined conceptual framework is crucial. As a result, modelling the determinant factors of household food insecurity is the case of modelling multivariate ordinal data that can consider the dependency between the dimensions.

A pair copula construction approach was proposed to determine both the dependence between food security dimensions and their respective associated factors simultaneously. A nice feature of the PCC approach in this setting is measuring the dependency of the food security dimensions using the copula parameter and the associated determinant factors of household food security for each dimension using the parameters of the marginal distributions. The pair copula construction approach with D-vine is attractive since it allows pairwise positive dependence structures as the presented conceptual framework by FAO (2008) and has closed form cumulative distribution function (cdf). Moreover, no other copula family has both these properties. FAO (2008) indicates that availability contributes to accessibility, accessibility contributes to utilisation and given that accessibility, availability contributes to utilisation; this is what a D-vine assumes in PCC.

This section demonstrates how to model and estimate dependence and marginal parameters from multivariate ordinal data using pair copula constructions via ordinal logistic regression to our motivating problem. This thesis did not evaluate the performance of this approach through simulation studies because it was evaluated via Bernoulli and Poisson discrete distributions by (Panagiotelis et al., 2012) and found to be a good model. They have also implemented the model for longitudinal ordinal data via probit model. However, the scale of the logistic is greater than the normal and this made the interpretation easier for logistic version and popular in many fields (Choi, 2012). As far as the researcher review of literature is concerned, no work has been conducted on the ordinal logistic version so far. Hence this section concerned on implementing the developed discrete PCC model via ordinal logistic regression for modelling household food insecurity determinants.

Since the current study concerns on discrete aspect in particular multivariate ordinal data, we now briefly review some key concepts for vine PCCs in the continuous case before introducing discrete vine PCCs. The aim here is to highlight some important distinctions in modelling discrete and continuous data via a copula approach, and to provide background for the introduction of discrete D-vine PCC presented in detail in section 4.3.2.

4.3.1 PCCs in the Continuous Case

For a vector $Y = (Y_1, ..., Y_m)$ of continuous random variables with joint density function $f(y_1, ..., y_m)$, a PCC is derived by starting with the following decomposition

$$f(y_1,...,y_m) = f_{1|2,...,m}(y_1 \mid y_2,...,y_m)\, f_{2|3,...,m}(y_2 \mid y_3,...,y_m) ... f_m(y_m) \tag{4.13}$$

Recalling equation (2.4), we can simplify the bivariate case to

$$f(y_1, y_2) = c_{12}(F(y_1), F(y_2))\, f_1(y_1)\, f_2(y_2) \tag{4.14}$$

where $c_{12}(.,.)$ is the appropriate pair-copula density for the pair of transformed variables $F_1(y_1)\, and\, F_2(y_2)$.

Any transformation using the factorization of Equation (4.13) and Equation (4.14) different decomposition can be constructed. For example, the 3-dimensional case decomposition results in

$$f_{1,2,3}(y_1, y_2, y_3) = f_{1|2,3}(y_1 \mid y_2, y_3)\, f_{2|3}(y_2 \mid y_3)\, f_3(y_3) \tag{4.15}$$

Basic calculations give the conditional density of $Y_2\, and\, Y_3$

$$\begin{aligned} f_{2|3}(y_2 \mid y_3) &= \frac{f_{2,3}(y_2, y_3)}{f_3(y_3)} = \frac{c_{23}(F(y_2), F(y_3))\, f_2(y_2)\, f_3(y_3)}{f_3(y_3)}. \\ &= c_{23}(F(y_2), F(y_3))\, f_2(y_2) \end{aligned} \tag{4.16}$$

Similarly,

$$\begin{aligned} f_{1|2,3}(y_1 \mid y_2, y_3) &= \frac{f_{13|2}(y_1 \mid y_2,\, y_3 \mid y_2)}{f_{3|2}(y_3 \mid y_2)} \\ &= \frac{c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2))\, f_{1|2}(y_1 \mid y_2)\, f_{3|2}(y_3 \mid y_2)}{f_{3|2}(y_3 \mid y_2)} \\ &= c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2))\, f_{1\backslash 2}(y_1 \mid y_2) \end{aligned}$$

$$\begin{aligned} &= c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2))\, f_{1|2}(y_1 \mid y_2) \\ &= c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2)) . c_{12}(F(y_1), F(y_2))\, f_1(y_1) \end{aligned} \tag{4.17}$$

Using Equation (4.15), (4.16) and (4.17), the following decomposition is appears

$$\begin{aligned} f_{1,2,3}(y_1, y_2, y_3) = c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2)) . c_{12}(F(y_1), F(y_2))\, f_1(y_1). \\ c_{23}(F(y_2), F(y_3)).\, f_3(y_3) . f_2(y_2).\, f_1(y_1). \end{aligned} \tag{4.18}$$

This example illustrates the construction of a 3-dimensional density using the bivariate copula and the corresponding marginal distributions. Any other factor in Equation (4.13), the same procedure is possible using the general formula letting V_h be any scalar element of V and $V_{\backslash h}$ its complement, with Y_j not an element of V (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015);

$$
\begin{aligned}
f_{Y_j|V} &= \frac{c_{Y_j,V_h|V_{|h}}(F_{Y_j|V_{|h}}, F_{V_h|V_{|h}})\, f_{Y_j|V_{|h}} \cdot f_{V_h|V_{|h}}}{f_{V_h|V_{|h}}} \\
&= c_{Y_j,V_h|V_{|h}}(F_{Y_j|V_{|h}}, F_{V_h|V_{|h}})\, f_{Y_j|V_{|h}}
\end{aligned}
\tag{4.19}
$$

where $c_{Y_j,V_h|V_{|h}}$ denotes the pair copula density describing the dependence between Y_j and V_h conditional on $V_{|h} = v_{|h}$. If we assume that the conditional copulas depend on the conditioning set only through their arguments, the decomposition in Equation (4.19) can motivate a statistical model. Typically, parametric bivariate copulas such as among the Archimedean families (Clayton, Gumbel & Frank) and elliptical families (Gaussian and Student t) copulas can be chosen to model the pair copulas.

The arguments of the pair copulas are conditional distribution functions and can be evaluated using the following expression given by Joe (1996) and cited in (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015);

$$
F_{Y_j|V_h,V_{|h}}(y_j \mid v_h, V_{|h}) = \frac{\partial C_{Y_j,V_h|V_{|h}}(F_{Y_j|V_{|h}}(y_j \mid V_{|h}), F_{V_h|V_{|h}}(v_j \mid V_{|h}))}{\partial F_{V_h|V_{|h}}(v_j \mid V_{|h})}. \tag{4.20}
$$

To compute the density of a PCC, one can be found the algorithms that recursively compute PCC density (Aas et al., 2009).

In conclusion, under appropriate regularity conditions, a multivariate density can be expressed as a product of m*(m − 1)/2 bivariate copulas, acting recursively on several different conditional probability distributions using expression (4.19). This leads to a large number of possible pair-copulas constructions. To organize all possible decompositions, a graphical model called a regular vine has been introduced by Bedford and Cooke (2002). Regular vine decompositions are concentrated only on the D-vines and C-vines, the special cases of regular vines. Each vine gives a specific way of decomposing the density. These models can be specified as a nested set of trees.

Vines of Continuous Case

A vine is characterized by $m-1$ trees denoted T_j for $j=1,\dots,m-1$. The j^{th} tree is made up of nodes, denoted N_j and edges which join these nodes, denoted E_j (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015). "A regular vine tree is called D-vine tree if each node in T −1 has at most 2 edges whereas C-vine tree if each tree T_j has a unique node with $m-j$ edges." The node with $m-1$ edges in tree T_1 is called the root. Figure 4.1 shows a D-vine decomposition for a 5-dimensional density function and Figure 4.2 shows a canonical-vine (Panagiotelis et al., 2012).

In Figure 4.1, in the first tree of a D-vine, the edges simply join adjacent nodes yielding $E_1=\{12=c_{12},23=c_{23},34=c_{34},45=c_{45}\}$. The edges on the first tree become the nodes on the second tree and in general $N_{j+1}=E_j$. The edges of trees $T_2,\dots,T_{m-1}$ also connect adjacent nodes. Any element shared by two nodes will be in the conditioning set of the edge joining them. For example, the edge joining node 12 and 23 is 13|2 while the edge joining 24|3 and 35|4 will be 25|34 (Panagiotelis et al., 2012).

The pair copulas that make up the corresponding PCC are simply indicated by the edges of the entire vine $\{E_1,\dots,E_{m-1}\}$, so that the density for a 5-dimensional PCC is given by

$$f(y_1,\dots,y_5)=\left[\prod_{k=1}^{5} f_k(y_k)\right].c_{12}.c_{23}.c_{34}.c_{45}.c_{13|2}.c_{24|3}.c_{35|4}.c_{14|23}.c_{25|34}.c_{15|234}$$

where the arguments of the pair copulas and density functions have been dropped for ease of notation.

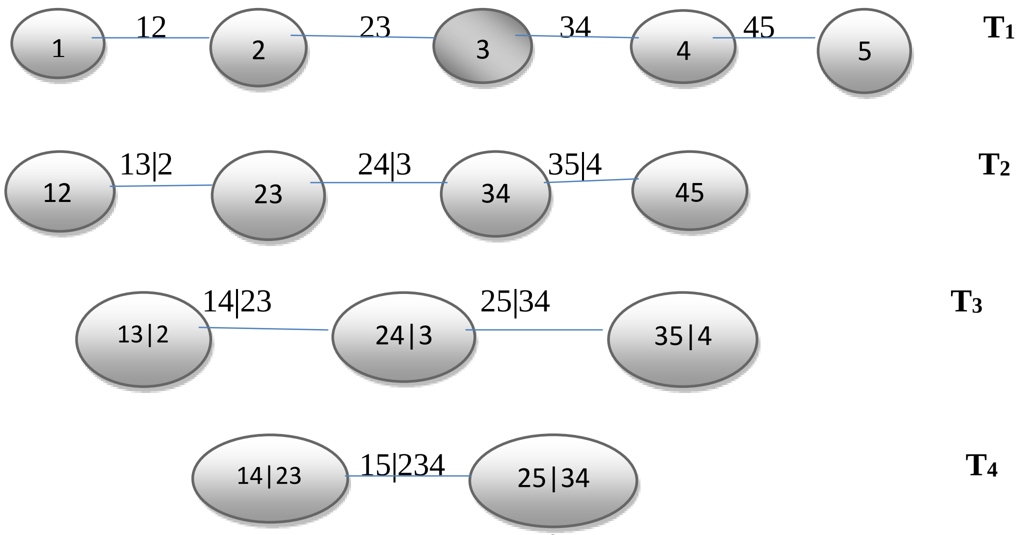


Figure 4. 1: A D-vine tree representation for m = 5.

The density $f(y_1,...,y_m)$ corresponding to a D-vine may be written using a general formula as

$$\prod_{k=1}^{m} f(y_k) \prod_{j=1}^{m-1} \prod_{i=1}^{m-j} c_{i,i+j|i+1,...,i+j-1}\{F(y_i \mid y_{i+1},...,y_{i+j-1}), F(y_{i+j} \mid y_{i+1},...,y_{i+j-1})\}, \quad (4.21)$$

where index j identifies the trees, while i runs over the edges in each tree. In a D-vine, no node in any tree T_j is connected to more than two edges.

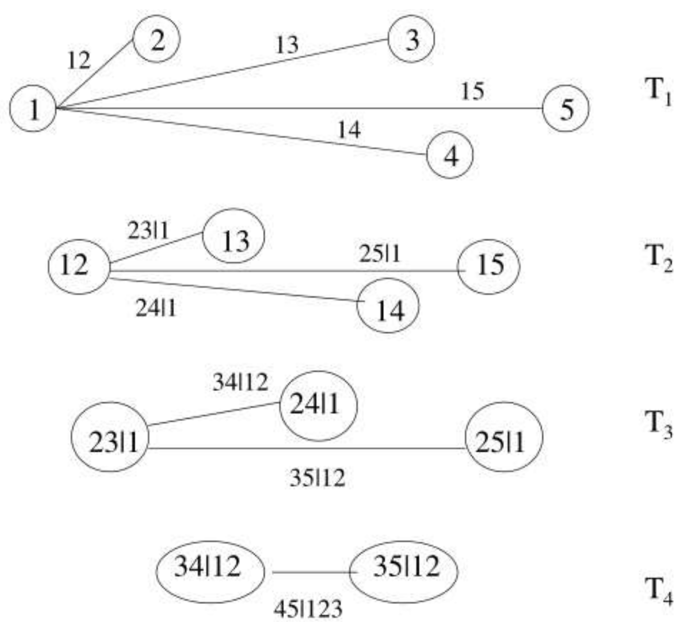


Figure 4. 2: A C-vine tree representation for m = 5.

Similarly, in Figure 4.2, in the first tree of a C-vine, the edges simply join adjacent nodes yielding $E_1 = \{12,13,14,15\}$. The edges on the first tree become the nodes on the second tree and in general $N_{j+1} = E_j$. The edges of trees $T_2,...,T_{m-1}$ also connect adjacent nodes. Any element shared by two nodes will be in the conditioning set of the edge joining them. For example, the edge joining node 12 and 13 is 23|1, while the edge joining 23|1 and 25|1, will be 35|12. The pair copulas that make up the corresponding PCC are simply indicated by the edges of the entire vine $\{E_1,...,E_{m-1}\}$, so that the density for a 5-dimensional PCC is given by

$$f(y_1,...,y_5) = \left[\prod_{k=1}^{5} f_k(y_k)\right].c_{12}.c_{13}.c_{14}.c_{15}.c_{23|1}.c_{24|1}.c_{25|4}.c_{34|12}.c_{35|12}.c_{45|123}$$

where the arguments of the pair copulas and density functions have been dropped for ease of notation. As a D-vine, a canonical vine for m-dimensional density is given by

$$\prod_{k=1}^{n} f(y_k)\prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{j,\,j+i|1,...,j-1}\{F(y_j \mid y_1,...,y_{j-1}), F(y_{i+j} \mid y_1,...,y_{j-1})\}, \qquad (4.22)$$

where index j identifies the trees, while i runs over the edges in each tree. In a canonical vine, each tree T_j has a unique node that is connected to $n-j$ edges.

Regular Vine Parameter Estimation

For estimation of regular vine, different scholars proposed non-standard methods and standard estimation methods. Stepwise and MLE, Inference Function for Margins (IFM) and Stepwise Semi-parametric Estimator (SSP) are the common standard estimation methods. MLE were considered for the first time by several researchers (Aas et al., 2009) IFM by (Joe, 1996), and SSP were by (Haff, 2012). These methods were designed for continuous data. However, we will not discuss here in detail since the current concern is on the discrete data. One can refer the references cited here for more detail. Just we now go to the PCC in discrete data.

4.3.2 PCCs in the Discrete Case

In the following sections, we are going to introduce vine PCCs for discrete margins that can be applicable for ordinal data. First, the discrete analogues to some important equations introduced in Section 4.3.2.1. Second, we discuss the D-vine decomposition with an illustration in full detail using 3-dimensional vine because the D-vine has certain advantages in applications where some intuitive ordering of the margins can be made and that it gives flexible models with parameters that can be estimated in a computationally and statistically efficient manner (Panagiotelis et al., 2012). Third, we discuss the selection of pair copula families of the D-vine and the parameter estimation.

4.3.2.1 Discrete PCCs

The aim here is to decompose a general multivariate probability mass function (pmf) into bivariate pair copula building blocks like continuous data. The joint pmf can be decomposed into a product of conditional probabilities using equation (4.13) for m discrete random variables $Y_1, Y_2, ..., Y_m$ as

$$\Pr(Y_1 = y_1,...,Y_m = y_m) = \Pr(Y_1 = y_1 \mid Y_2 = y_2,...,Y_m = y_m) \times \Pr(Y_2 = y_2 \mid Y_3 = y_3,...,Y_m = y_m) \times \\ ...\Pr(Y_m = y_m). \tag{4.23}$$

Now, this expression has terms of the form $\Pr(Y_j = y_j \mid Y_{|j} = y_{|j})$ where $Y_{|j}$ is the vector of random variables $Y_1, Y_2, ..., Y_m$ excluding Y_j and $y_{|j}$ is the same vector for the realized values of the random variables. Choosing another element from the vector of random variables, we can rewrite the discrete joint probability in a similar fashion to the continuous case as following:

$$\Pr(Y_j = y_j \mid Y_{|j} = y_{|j}) = \frac{\Pr(Y_j = y_j, Y_h = y_h \mid Y_{|j,h} = y_{|j,h})}{p(Y_h = y_h \mid Y_{|j,h} = y_{|j,h})} \tag{4.24}$$

Now, recalling the probability mass function and the multivariate copula function for discrete data in Equation (4.5), the bivariate conditional probability in the numerator can be expressed in terms of a copula giving

$$\Pr(Y_j = y_j \mid Y_{|j} = y_{|h})$$
$$= \frac{\sum_{i_j=0,1}\sum_{i_h=0,1} (-1)^{i_j+i_h} \Pr(Y_j \le y_j - i_j, Y_h \le y_h - i_h \mid Y_{|j.h} = y_{|j.h})}{\Pr(Y_h = y_h \mid Y_{|j.h} = y_{|j.h})}$$
$$= \frac{\sum_{i_j=0,1}\sum_{i_h=0,1} (-1)^{i_j+i_h} C_{Y_j.Y_h|y_{|j.h}}(F_{Y_j|Y_{|j.h}}(y_j - i_j), F_{Y_j|Y_{|j.h}}(y_h - i_h))}{\Pr(Y_h = y_h \mid Y_{|j.h} = y_{|j.h})} \quad (4.25)$$

The arguments in equation (4.25) of the copula functions are evaluated using the following (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015);

$$\begin{aligned} F_{Y_j|Y_h.Y_{|j.h}}(y_j \mid y_h, y_{|j.h}) = [&C_{Y_j.Y_h|Y_{|j.h}}(F_{Y_{j\backslash}|Y_{|j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h \mid y_{|j.h})) \\ &- C_{Y_j.Y_h|Y_{|j.h}}(F_{Y_{j\backslash}|Y_{\backslash j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h - 1 \mid y_{|j.h}))] \\ &/ \Pr(Y_h = y_h \mid Y_{|j.h} = y_{|j.h}) \end{aligned} \quad (4.26)$$

This vine PCC has nice feature than multivariate and Gaussian copulas functions in evaluating the probability mass function because the PCC requires $2m(m-1)$ evaluations whereas the multivariate and Gaussian copulas require 2^m evaluations (Panagiotelis et al., 2012).

4.3.2.2 D-vine in Discrete Data

For illustration purposes, we present in detail the 3-dimensional case for instance in the food security data, Y_1 = Availability, Y_2 = Accessibility and Y_3 = Utilisation. Therefore,

$$\begin{aligned} &\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) \\ &\quad = \Pr(Y_1 = y_1 \mid Y_2 = y_2, Y_3 = y_3) \times \Pr(Y_3 = y_3 \mid Y_2 = y_2) \times \Pr(Y_2 = y_2) \end{aligned} \quad (4.27)$$

Utilizing Equation (4.25) the right hand side of the first conditional probability can be rewritten as:

$$\Pr(Y_1 = y_1 \mid Y_2 = y_2, Y_3 = y_3) = \frac{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3} C_{13|2}(F_{1|2}(y_1 - i_1 \mid y_2), F_{3|2}(y_3 - i_3 \mid y_2))}{\Pr(Y_3 = y_3 \mid Y_2 = y_2)} \quad (4.28)$$

Similarly, utilizing Equation (4.25), the first argument of the copula function in the numerator of Equation (3.16) is given by;

$$F_{1|2}(y_1 - i_1 \mid y_2) = \frac{C_{12}(F_1(y_1 - i_1), F_2(y_2)) - C_{12}(F_1(y_1 - i_1), F_2(y_2 - 1))}{\Pr(Y_2 = y_2)}, \quad (4.29)$$

and the second argument can be expressed as

$$F_{3|2}(y_3 - i_3 \mid y_2) = \frac{C_{23}(F_2(y_2), F_3(y_3 - i_3)) - C_{23}(F_2(y_2 - 1), F_3(y_3 - i_3))}{\Pr(Y_2 = y_2)} \quad (4.30)$$

By cancelling terms and substituting, the probability mass function of the full expression for the 3-dimensional discrete D-vine is given by

$$\begin{aligned}\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) \\ = \{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3} C_{13|2}(\frac{C_{12}(F_1(y_1 - i_1), F_2(y_2)) - C_{12}(F_1(y_1 - i_1), F_2(y_2 - 1))}{F_2(y_2) - F_2(y_2 - 1)}, \\ \frac{C_{23}(F_2(y_2), F_3(y_3 - i_3)) - C_{23}(F_2(y_2 - 1), F_3(y_3 - i_3))}{F_2(y_2) - F_2(y_2 - 1)})\}[F_2(y_2) - F_2(y_2 - 1)].\end{aligned} \quad (4.31)$$

Somewhat confusing to write the general D-vine structure; however, the general dimension algorithm for computing the probability mass function of a D-vine was outlined by Panagiotelis et al. (2012). It is evident both from this algorithm and the 3-dimensional example above that each bivariate pair copula only needs to be evaluated 4 times, specifically $C_{Y_j, Y_h | Y_{|j.h}}$ must be evaluated at (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015);

$(F_{Y_{j\setminus}|Y_{|j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h \mid y_{|j.h}))$, $(F_{Y_{j\setminus}|Y_{|j.h}}(y_j - 1 \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h \mid y_{|j.h}))$, $(F_{Y_{j\setminus}|Y_{|j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h - 1 \mid y_{|j.h}))$ and $(F_{Y_{j\setminus}|Y_{|j.h}}(y_j - 1 \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h - 1 \mid y_{|j.h}))$.

In general, the evaluation of the probability mass function requires 2m(m − 1) evaluations of bivariate copula functions, even though the continuous vine PCC is composed of only m(m−1)/2 pair copulas. These vine PCCs have still greater potential in high-dimensional settings since the computational burden of evaluating the pmf in the elliptical copulas is 2^m. As pointed out by Panagiotelis et al. (2012) among the major advantage of D-vine PCCs, a wide variety of dependence structures can be modelled by selecting different copula families as building blocks. Gaussian, t, Clayton, Frank and Gumbel copulas are the commonly used parametric copula families as building blocks.

The marginal probabilities can be modelled by logistic, probit or Poisson models. If the discrete data have binary outcomes, binary logistic or probit models can be used, if ordinal, ordinal logistic or probit and if counted, Poisson regression model can be used. For the current study since the data are ordinal, cumulative logit model is used. A detail of the cumulative logit model is given below.

The marginal distribution of ordinal data via cumulative logistic regression model for single ordinal response variable Y that has C categories and labelled $1, 2, ..., C-1.$ is given by (Agresti, 2010);

$$\begin{aligned} \Pr(Y \le j) &= \frac{\Pr(Y \le j)}{1 - \Pr(Y \le j)} \\ &= \frac{\exp(\alpha_j + \beta' \mathrm{X})}{1 + \exp(\alpha_j + \beta' \mathrm{X})}, \qquad j = 1, 2, ..., C-1. \end{aligned} \tag{4.32}$$

This is the cumulative probabilities that an observation fall in category j or below, for C categories. Each of the cumulative logits is an ordinary binary logit indicating the probability of an outcome falling into either the first 1... j categories or the j+ 1..., C categories. Similarly, α_j the intercept for each cumulative probability of the c category and column vector ß of parameters that describes the effects of the explanatory variables (Agresti, 2010).

Hence, the arguments in equation (4.31) can take the marginal distribution function given in equation (4.32). Moreover, the joint probability mass function in question (4.27) is expressed in terms of the pair copula functions and ordinal marginal distributions. The newly constructed joint probability mass function can be called pair copula-based multivariate cumulative logit model. Finally, this function can be estimated using the appropriate parameter estimation technique and appropriate bivariate copula families as building blocks (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015).

4.3.2.2.1 Selection of Pair Copula Families and Parameter Estimation of the D-Vine

Before parameter estimations of the pair copula construction model, determining the order of the D-vine and selection of appropriate bivariate copula families for the model is the first task to be done.

For the structure of R-vine copula extensive works of several researchers (Aas et al., 2009, Czado et al., 2013, Dissmann et al., 2013, Sutkoff, 2014) were conducted. In particular, "the order in the trees corresponding to a D-vine copula Aas et. al (2009) put the strongest bivariate dependencies in the first tree of the D-vine tree specification. Strongest bivariate dependencies within the copula distribution might be measured by Kendall's τ or the tail dependence coefficient λ, which is a function of the chosen bivariate copula". For the current study on food security, the conceptual framework for food security dimension is given by FAO (2008) like Availability-Accessibility-Utilisation. We can use this order for the structure of D-vine for inference purposes.

Now the bivariate Copula Families of the vine distribution can be selected since the order of the D-vine tree specification is chosen. This part can be discussed as follows.

i. **Selection of Pair Copula Families**

Accordingly, as described above, we need to select a copula family for every pair of variables. Commonly used copula families that we consider in the later applications are Gaussian (N), t, Clayton (C), Gumbel (G) and Frank (F). The Clayton and Gumbel copulas are applicable only to model positive dependence. Hence, in case of negative dependence (i.e. negative values for Kendall's tau) we can reduce them. Further, if the degree of freedom of the MLE is higher than 30, we will not use a t copula.

After reducing the possible options further, we can decide which copula fits "best". To select jointly the vine structure and best fit copula families, (Panagiotelis et al., 2015) have developed an algorithm through adaptation from the algorithm developed for continuous data by Dissmann et al. (2013). We also customized this algorithm for our purpose to decide best fit copula families only as presented in Algorism I.

Algorithm I

Consider discrete random variables $Y = Y_1, Y_2, Y_3, \dots, Y_m$ with known marginal distribution functions $F_j(.)$, the steps to select copula families is as follows.

1. Generate the `pseudo data'$u_{ij}^{+} := F_j(y_{ij})$ and $u_{ij}^{+} := F_j(y_{ij} - 1)$ for $j = 1, 2, \dots, m$ and $i = 1, 2, \dots, n$, where $y_{ij}$the value of the response for the j[th] margin and the i[th] observation.
2. Consider a pair of two margins $I_1 \; and \; I_2 \subset \{1, 2, \dots, m\}$
   i. Fit the copula C^{θ^r} using the pseudo data for the margins $I_1 \; and \; I_2$ for each bivariate copula families as follows

$$\hat{\theta}^r = argmaxlnL_{I_1, I_2}^r(\theta^r) \qquad (4.33)$$

Where,

$$lnL_{I_1,I_2}^r(\theta^r) = \sum_{i=1}^{n} \ln \left(C^{\theta^r}(u_{iI_1}^{+}, u_{iI_2}^{+}) - C^{\theta^r}\left(u_{iI_1}^{+}, u_{iI_2}^{-}\right) - C^{\theta^r}\left(u_{iI_1}^{-}, u_{iI_2}^{+}\right) + C^{\theta^r}\left(u_{iI_1}^{-}, u_{iI_2}^{-}\right)\right)$$

   ii. Compute a modified Akaike Information Criterion (AIC), that removes the effect of the margins, given by

$$mAIC^r = -2lnL_{I_1,I_2}^r(\theta^r) - lnL_{I_1}^r - lnL_{I_2}^r + 2q_r \qquad (4.34)$$

Where q_r is the dimension ofθ^r,$lnL_{I_1}^r = \sum_{i=1}^{n} \ln \left(u_{iI_1}^{+} - u_{iI_1}^{-}\right)$,

$$nL_{I_2}^r = \sum_{i=1}^{n} \ln \left(u_{iI_2}^{+} - u_{iI_2}^{-}\right).$$

A smaller mAIC value indicates a better parametric model.

   iii. Compute new pseudo data for tree 2 means that conditional pseudo data as given by

$u_{i,h_1|h_2}^{+} := F_{h_1|h_2}(y_i h_1 | y_i h_2)$, $u_{i,h_1|h_2}^{-} := F_{h_1|h_2}(y_i h_1 - 1 | y_i h_2)$,

$u_{i,h_1|h_2}^{+} := F_{h_2|h_1}(y_i h_2 | y_i h_1)$, $u_{i,h_2-1|h_1}^{+} := F_{h_2|h_1}(y_i h_2 | y_i h_1)$,

3. Repeat step 2 for all pairs of the new pseudo data and corresponding pair copulas. Also compute new pseudo data in a similar fashion as step 2(iii).
4. Iterate to select the pair copulas.

Once the copula families selected for each edges and consecutive trees, the next step is parameter estimation.

ii. Parameter Estimation

For estimation of regular vine, stepwise and MLE were consider for the first time by Aas et al. (2009), Inference function for margins (IFM) by Joe (1996) and Stepwise semiparametric estimator (SSP) were by Haff (2013) in the continuous margins. Similarly, Panagiotelis et al. (2012) conducted MLE and IFM for the discrete margins.

a. Maximum Likelihood (ML) Estimator

Since the 3 dimensional D-vine is derived, hence the log-likelihood function of a 3 dimensional D-vine is given by

$$l(\beta,\theta;y)=\sum_{i=1}^{n}\log(\Pr(Y_1=y_1,Y_2=y_2,Y_3=y_3;\beta,\theta))$$

$$=\sum_{t=1}^{3}\log(\{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3}C_{13|2}(\frac{C_{12}(F_1(y_1-i_1;\beta_1),F_2(y_2;\beta_2);\theta_{12})-C_{12}(F_1(y_1-i_1;\beta_1),F_2(y_2-1;\beta_2);\theta_{12})}{F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)},$$

$$\frac{C_{23}(F_2(y_2;\beta_2),F_3(y_3-i_3;\beta_3);\theta_{23})-C_{23}(F_2(y_2-1;\beta_2),F_3(y_3-i_3;\beta_3);\theta_{23})}{F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)};\theta_{13|2})\}[F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)]) \quad (4.35)$$

Here $\beta=(\beta_1,\beta_2,\beta_3)$ and $\theta=(\theta_{12},\theta_{23},\theta_{13|2})$ where the marginal and copula parameters respectively. Then, the ML estimator $\hat{\theta}^{ML}$ is obtained by maximizing the above log-likelihood function over all parameters, β and θ, simultaneously.

For the general case, let the model for the j[th] margin imply a marginal distribution function $F_{ij}(y_{ij};\beta_j)$, where β_j are the marginal parameters and the subscript i denotes that we observe a sample $y_i=(y_{i1},y_{i2},...,y_{im})'$ for i = 1, 2, ..., n. Similarly, the copula parameters for m-dimensional dependence were given by $\theta_{i,i+j|i+1,...,i+j-1}$. Then, the ML estimator $\hat{\theta}^{ML}$ is obtained by maximizing the log-likelihood function over all parameters, β_j and $\theta_{i,i+j|i+1,...,i+j-1}$, simultaneously.

In this optimization, good starting values are required. Starting values for the marginal parameters are obtained by following the first step of the IFM approach (will be discussed next to this). Moreover, starting values for the copula parameters can be found by computing empirical Kendall's τ of bivariate copula function of the first tree which act as 'pseudo' data and then transformed back to the copula parameter using a known Bijection.

b. Inference Function for Margins (IFM) Estimator

Consider equation (3.18) the 3-dimensional marginal distribution and the general case, in the first step, maximum likelihood estimates of the marginal parameters $\hat{\beta}^{IFM}$ are estimated for all one at a time, ignoring dependence with the other margins. The resulting estimates $\hat{\beta}^{IFM}$ are plugged into the arguments of the marginal in the bivariate copula functions to estimate the pair copula parameter $\hat{\theta}^{IFM}$. In the second step, the copula parameters are estimated by maximum likelihood using $\hat{\beta}^{IFM}$ as an argument of the marginal (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015).

Panagiotelis et al. (2012) point out that joint ML estimates are generally of a higher quality than IFM, but only slightly so. On the other hand, IFM estimation is simpler and faster, particularly for more complicated marginal models. Hence, in the current study, the marginal models are only three; joint MLE can be implemented with the cumulative logit margin as discussed above.

4.4 Pair Copula Construction for Longitudinal Ordinal Data

In repeated or longitudinal outcomes, the dependency among outcomes must be accounted for in order to make valid inference. In this study, the households were surveyed three times at six-month interval for each of the three dimensions. A composite food security index that has four levels was computed from these three dimensions for each of the three round of data collection. The levels are "sever food in-secured", "mildly food in-secured", "moderately food in-secured" and "food secured" for each round of data. Therefore, three composite food security indexes were obtained from the three phases of data collection, resulted in longitudinal ordinal outcomes. Hence, in this regard, modelling the household food security status is the case of modelling longitudinal ordinal data that can take into consideration the dependency between consecutive time points.

In modelling the stability and the determinants of household food insecurity, a PCC model was proposed. A nice feature of the PCC approach in this setting is measuring the dependency of the consecutive food security status of the households using the copula parameter and the respective associated determinants using the parameters of the marginal distributions. The dependences of the consecutive food security statuses are one-dimensional. Hence, the pair copula construction approach with D-vine is attractive since it allows pairwise positive dependence structures and has closed form cumulative distribution function (cdf), no other copula family has both these properties.

This section demonstrated how to model and estimate dependence and marginal parameters from longitudinal ordinal data using pair copula constructions via ordinal logistic regression to our motivating problem. The thesis did not evaluate the performance of this approach through simulation studies because it was evaluated via Bernoulli and Poisson discrete distributions by (Panagiotelis et al., 2012) and found to be a good model. They have also implemented the model for longitudinal ordinal data via probit model. However, the scale of the logistic is greater than the normal and this made the interpretation easier for logistic version and popular in many fields (Choi, 2012). As far as the researcher review of literature is concerned, no work has been conducted on the ordinal logistic version so far. Hence, this thesis in this section implemented the developed discrete PCC model via ordinal logistic regression for modelling the stability and determinants of household food insecurity status.

Since the current study concerns on discrete aspect in particular longitudinal ordinal data, first we briefly review some key concepts for vine PCCs in the continuous longitudinal cases before introducing discrete vine PCCs for longitudinal ordinal margins.

4.4.1 Pair Copula Construction for Longitudinal Continuous Data

A continues univariate random variable repeatedly measured for T time points given by $Y = (Y_1, \ldots, Y_T)$, the joint density function $f(y_1, \ldots, y_T)$ is decomposed as follows;

$$f(y_1,...,y_T) = f(y_t \mid y_{t-1},...,y_1) * f(y_{t-1} \mid y_{t-2},...,y_1) * ... * f(y_1) = \prod_{t=2}^{T} f(y_t \mid y_{t-1},...,y_1) * f(y_1) \quad . \qquad (4.36)$$

In the Sklar's theorem the conditional bivariate densities, $f(y_t|y_{t-1}, y_{t-2}, \dots, y_1)$ in equation (4.36) for $t > s$ is given by

$$\begin{aligned} f(y_t \mid y_{t-1},...,y_1) &= \frac{f(y_t, y_s \mid y_{t-1},...,y_{s+1})}{f(y_s \mid y_{t-1},...,y_{s+1})} \\ &= c_{t,s \mid t-1,t-2,...,s+1}(F(y_t \mid y_{t-1},...,y_{s+1}), F(y_s \mid y_{t-1},...,y_{s+1}))\, f(y_t \mid y_{t-1},...,y_{s+1}) \end{aligned} \quad , \quad (4.37)$$

where $f(\bullet|\bullet)$ *and* $F(\bullet|\bullet)$ denotes the conditional density and cumulative density functions, respectively and t and s be any arbitrary distinct indices.

By setting s=1, the bivariate conditional density in equation (4.37) yields the following decomposition,

$$f(y_t \mid y_{t-1},...,y_1) = c_{t,1 \mid t-1,t-2,...,2}(F(y_t \mid y_{t-1},...,y_2), F(y_s \mid y_{t-1},...,y_2))\, f(y_t \mid y_{t-1},...,y_2) \quad . \quad (4.38)$$

Repeatedly, setting s = 2, 3, …, t-1, the conditional density in (4.3) leads the following conditional density decomposition,

$$f(y_t \mid y_{t-1},...,y_1) = \prod_{s=1}^{t-1} \left\{ c_{t,s \mid t-1,t-2,...,s+1}(F(y_t \mid y_{t-1},...,y_{s+1}), F(y_s \mid y_{t-1},...,y_{s+1})) \right\} * f(y_t)$$

(4.39)

Replacing equation (4.39) in equation (4.36), then the joint distribution function becomes

$$f(y_1,...,y_T) = \prod_{t=2}^{T} \left\{ \prod_{s=1}^{t-1} \left\{ c_{t,s|t-1,t-2,...,s+1}(F(y_t|y_{t-1},...,y_{s+1}), F(y_s|y_{t-1},...,y_{s+1})) \right\} * f(y_t) \right\} * f(y_1) \qquad (4.40)$$

Equation (4.40) is a product of $T(T-1)/2$ bivariate pair copula densities and T marginal densities (Ruscone and Osmetti, 2017, Smith et al., 2010). This leads to a large number of possible pair-copulas constructions. To organize all possible decompositions, a graphical model called a regular vine has been introduced by (Bedford and Cooke, 2002). Regular vine decompositions are concentrated only on the D-vines and C-vines, the special cases of regular vines. Equation (4.40) can be recognized as D-vine model. Detail of D-vine construction was displayed earlier in section 4.3.1.

D-Vine Parameter Estimation

For estimation of regular vine, different scholars proposed non-standard methods and standard estimation methods. Stepwise and MLE (MLE), Inference Function for Margins (IFM) and Stepwise Semi-parametric Estimator (SSP) are the common standard estimation methods. MLE were considered for the first time by (Aas et al., 2009), IFM by (Joe, 1996), and SSP were by (Haff, 2012). These methods are designed for continuous data. We will not discuss here in detail since the current concern is on the discrete data. One can refer the referees cited here for more detail. Just we now go to the PCC in discrete data.

4.4.2 Pair Copula Construction for Longitudinal Discrete Data

The aim here is to decompose the probability mass function (pmf) of longitudinal discrete data into bivariate pair copula building blocks like the decomposition of longitudinal continuous data. For T time ordered discrete random variables given by $Y_1, Y_2, \dots, Y_T$, the joint pmf can be decomposed into a product of conditional probabilities as

$$\Pr(Y_1 = y_1, \dots, Y_T = y_T) = \Pr(Y_t = y_t \mid Y_{t-1} = y_{t-1}, \dots, Y_1 = y_1) \times \Pr(Y_{t-1} = y_{t-1} \mid Y_{t-2} = y_{t-2}, \dots, Y_1 = y_1) \times \dots \Pr(Y_1 = y_1)$$
$$= \prod_{t=2}^{T} \left\{ \Pr(Y_t = y_t \mid Y_{t-1} = y_{t-1}, \dots, Y_1 = y_1) \right\} * \Pr(Y_1 = y_1). \quad (4.41)$$

Now, the expression $Pr(Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2,} = y_{t-2}, \dots, Y_1 = y_1)$ can be written as the form $Pr(Y_t = y_t | Y_{|t} = y_{|t})$, where $Y_{|t}$ is the vector of random variables $Y_1, Y_2, \dots, Y_T$ excluding Y_t and $y_{|t}$ is the same vector for the realized values of the random variables. Choosing another element s from the vector of random variables, we can rewrite the discrete joint probability as following:

$$\Pr(Y_t = y_t \mid Y_{|t} = y_{|t}) = \frac{\Pr(Y_t = y_t, Y_s = y_s \mid Y_{|t,s} = y_{|t,s})}{\Pr(Y_s = y_s \mid Y_{|t,s} = y_{|t,s})} \quad (4.42)$$

Recalling equation (4.5), the bivariate conditional probability in the numerator of equation (4.42) can be expressed in terms of a copula giving,

$$\Pr(Y_t = y_t \mid Y_{|t} = y_{|t}) = \frac{\sum_{i_t=0,1} \sum_{i_s=0,1} (-1)^{i_t+i_s} \Pr(Y_t \le y_t - i_t, Y_s \le y_s - i_s \mid Y_{|t.s} = y_{|t.s})}{\Pr(Y_s = y_s \mid Y_{|j.s} = y_{|j.s})}$$

$$= \frac{\sum_{i_t=0,1} \sum_{i_s=0,1} (-1)^{i_t+i_s} C_{Y_t.Y_s|y_{|t.s}} (F_{Y_t|Y_{|t.s}}(y_t - i_t), F_{Y_s|Y_{|t.s}}(y_s - i_s))}{\Pr(Y_s = y_s \mid Y_{|t.s} = y_{|t.s})} \tag{4.43}$$

Inserting equation (4.43) in equation (4.41), now we get the following decomposed joint probability mass function for longitudinal discrete data.

$$\Pr(Y_1 = y_1, \ldots, Y_T = y_T) = \prod_{t=2}^{T} \left\{ \frac{\sum_{i_t=0,1} \sum_{i_s=0,1} (-1)^{i_t+i_s} C_{Y_t.Y_s|y_{|t.s}} (F_{Y_t|Y_{|t.s}}(y_t - i_t), F_{Y_t|Y_{|t.s}}(y_s - i_s))}{\Pr(Y_s = y_s \mid Y_{|t.s} = y_{|t.s})} \right\} * \Pr(Y_1 = y_1). \tag{4.44}$$

Equation (4.44) can be recognized as general D-vine pair copula construction model. This vine PCC requires $2T(T-1)$ evaluations for evaluating the probability mass function whereas the multivariate and Gaussian copulas require 2^T evaluations (Panagiotelis et al., 2012).

For illustration purposes, we present in detail the 3-dimensional longitudinal case. Therefore,

$$\begin{aligned}\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) &\\ &= \Pr(Y_3 = y_3 \mid Y_1 = y_1, Y_2 = y_2) \times \Pr(Y_2 = y_2 \mid Y_1 = y_1) \times \Pr(Y_1 = y_1) \\ &= \Pr(Y_3 = y_3 \mid Y_1 = y_1, Y_2 = y_2) \times \Pr(Y_1 = y_1 \mid Y_2 = y_2) \times \Pr(Y_2 = y_2)\end{aligned} \tag{4.45}$$

Utilizing Equation (4.43) the conditional probability, $Pr(Y_3 = y_3 | Y_1 = y_1, Y_2 = y_2)$ can be rewritten as:

$$\Pr(Y_3 = y_3 \mid Y_1 = y_1, Y_2 = y_2) = \frac{\sum_{i_1=0,1} \sum_{i_3=0,1} (-1)^{i_1+i_3} C_{13|2}(F(y_1 - i_1 \mid y_2), F(y_3 - i_3 \mid y_2))}{\Pr(Y_3 = y_3 \mid Y_2 = y_2)} \tag{4.46}$$

Similarly utilizing Equation (4.8), the first argument of the copula function in the numerator of Equation (4.45), $F(y_1 - i_1 | y_2)$ is given by

$$F(y_1 - i_1 \mid y_2) = \frac{C_{12}(F(y_1 - i_1), F(y_2)) - C_{12}(F(y_1 - i_1), F(y_2 - 1))}{\Pr(Y_2 = y_2)}. \tag{4.47}$$

And the second argument, $F(y_3 - i_3 \mid y_2)$ can be expressed as

$$F(y_3 - i_3 \mid y_2) = \frac{C_{23}(F(y_2), F(y_3 - i_3)) - C_{23}(F(y_2 - 1), F(y_3 - i_3))}{\Pr(Y_2 = y_2)} . \qquad (4.49)$$

By cancelling similar terms and substituting, the probability mass function of the full expression for the 3-dimensional longitudinal discrete D-vine is given by

$$\begin{aligned}
&\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) \\
&\quad = \{\sum_{i_1=0,1} \sum_{i_3=0,1} (-1)^{i_1+i_3} C_{13|2}(\frac{C_{12}(F(y_1 - i_1), F(y_2)) - C_{12}(F(y_1 - i_1), F(y_2 - 1))}{F(y_2) - F(y_2 - 1)}, \qquad (4.50) \\
&\quad \frac{C_{23}(F(y_2), F(y_3 - i_3)) - C_{23}(F(y_2 - 1), F(y_3 - i_3))}{F(y_2) - F(y_2 - 1)})\}[F(y_2) - F(y_2 - 1)].
\end{aligned}$$

Equation (4.50) is a D-vine model for 3-dimentional longitudinal discrete case. It is evident from the 3-dimensional example above that each bivariate pair copula only needs to be evaluated four times, specifically $C_{Y_t, Y_s | Y_{|t.s}}$ must be evaluated at

$(F(y_t \mid y_{|t.s}), F(y_s \mid y_{|t.s}))$, $(F(y_t - 1 \mid y_{|t.s}),\ F(y_s \mid y_{|t.s}))$, $(F(y_t \mid y_{|t.s}),\ F(y_s - 1 \mid y_{|t.s}))$ and $(F(y_t - 1 \mid y_{|t.s}),\ F(y_s - 1 \mid y_{|t.s}))$.

In general, the evaluation of the probability mass function requires 2T(T − 1) evaluations of bivariate copula functions, even though the continuous vine PCC is composed of only T(T−1)/2 pair copulas. These vine PCCs have still greater potential in high-dimensional settings since the computational burden of evaluating the pmf in the elliptical copulas is 2^T. As pointed out by Panagiotelis et al. (2012) among the major advantage of D-vine PCCs, a wide variety of dependence structures can be modelled by selecting different copula families as building blocks. Among these, Gaussian, t, AMH Clayton, Frank and Gumbel copulas are the commonly used parametric copula families as building blocks.

The marginal probabilities in equation (4.50) can be modelled either of among discrete probability distributions. For the current study, since the data are ordinal, cumulative logit model is used. Details of the cumulative logit model were given in equation (4.32).

Hence, the arguments in equation (4.44 or 4.50) can take the marginal distribution function given in equation (4.32). Moreover, the joint probability mass function in question (4.45) is expressed in terms of the pair copula functions and ordinal marginal distributions. The newly constructed joint probability mass function can be called pair copula-based longitudinal cumulative logit model. Finally, this function can be estimated using the appropriate parameter estimation technique and appropriate bivariate copula families as building blocks.

4.4.3 Selection of Pair Copula Families and Parameter Estimation of the D-Vine

Before the estimation of the parameters for the pair copula construction model determining the order of the D vine and choose appropriate bivariate copula families for the model is the first task to be done.

For the structure of R-vine, copula extensive approaches were reviewed in chapter three of this paper. Moreover, the current study concerned on time ordered or longitudinal case, the time order by itself can be taken as D-vine structure for inference purposes being stars from newest to the oldest. Now the bivariate Copula Families of the vine distribution can be selected since the order of the D-vine tree specification is chosen. This part can be discussed as follows.

4.4.3.1 Selection of Pair Copula Families

Accordingly as described above, we need to select a copula family for every pair of variables. Commonly used copula families that we consider in the later applications are Gaussian (N), Ali-Mikhail-Haq (AMH), Clayton (C), Gumbel (G) and Frank (F). The Clayton and Gumbel copulas are applicable only to model positive dependence. Hence, in case of negative dependence (i.e., negative values for Kendall’s tau), we can reduce them. Further, if the degree of freedom of the MLE is higher than 30, we will not use a t copula.

After reducing the possible options further, we can decide which copula fits “best”. To select jointly the vine structure and best fit copula families, (Panagiotelis et al., 2015) have developed an algorithm through adaptation from the algorithm developed for continuous data by Dissmann et al. (2013). We also customized this algorithm for our purpose to decide best fit copula families only as presented in Algorism II of this chapter.

Algorithm II

Consider T time ordered discrete random variables $Y = Y_1, Y_2, Y_3, \dots, Y_T$ with known marginal distribution functions $F_t(.)$, the steps to select copula families is as follows.

1. Generate the `pseudo data' $u_{it}^{+} := F_t(y_{it})$ and $u_{it}^{+} := F_t(y_{it} - 1)$ for $t = 1, 2, \dots, T$ and $i = 1, 2, \dots, n$, where y_{it} the value of the response for the t[th] time ordered margin and the i[th] observation.
2. Consider a pair of two time ordered margins I_1 and $I_2 \subset \{1, 2, \dots, T\}$
   i. Fit the copula C^{θ^r} using the pseudo data for the margins I_1 and I_2 for each bivariate copula families as follows:

$$\hat{\theta}^r = argmax lnL_{I_1,I_2}^r(\theta^r) \tag{4.51}$$

   Where,

$$lnL_{I_1,I_2}^r(\theta^r) = \sum_{i=1}^{n} \ln\left(C^{\theta^r}(u_{iI_1}^{+}, u_{iI_2}^{+}) - C^{\theta^r}\left(u_{iI_1}^{+}, u_{iI_2}^{-}\right) - C^{\theta^r}\left(u_{iI_1}^{-}, u_{iI_2}^{+}\right) + C^{\theta^r}\left(u_{iI_1}^{-}, u_{iI_2}^{-}\right)\right)$$

   ii. Compute a modified Akaike Information Criterion (AIC), that removes the effect of the margins, given by

$$mAIC^r = -2lnL_{I_1,I_2}^r(\theta^r) - lnL_{I_1}^r - lnL_{I_2}^r + 2q_r \tag{4.52}$$

   Where q_r is the dimension of θ^r,$lnL_{I_1}^r = \sum_{i=1}^{n} \ln(u_{iI_1}^{+} - u_{iI_1}^{-})$,
   $lnL_{I_2}^r = \sum_{i=1}^{n} \ln(u_{iI_2}^{+} - u_{iI_2}^{-})$.

   A smaller mAIC value indicates a better parametric model.

   iii. Compute new pseudo data for tree 2 means that conditional pseudo data as given by
   $u_{i,h_1|h_2}^{+} := F_{h_1|h_2}(y_i h_1 | y_i h_2)$, $u_{i,h_1|h_2}^{-} := F_{h_1|h_2}(y_i h_1 - 1 | y_i h_2)$,
   $u_{i,h_1|h_2}^{+} := F_{h_2|h_1}(y_i h_2 | y_i h_1)$, $u_{i,h_2-1|h_1}^{+} := F_{h_2|h_1}(y_i h_2 | y_i h_1)$,
3. Repeat step 2 for all pairs of the new pseudo data and corresponding pair copulas. Also compute new pseudo data in a similar fashion as step 2(iii).
4. Iterate to select the pair copulas.

Once the copula families selected for each edges and consecutive trees, the next step is parameter estimation.

4.4.3.2 Parameter Estimation

For estimation of regular vine, details of different estimation techniques were assessed in Chapter 3. Here, we only concerned on MLE that will be applied for this application area.

Since the 3-dimensional D-vine is derived, hence the log-likelihood function of a 3 dimensional D-vine is given by

$$l(\beta,\theta;y)=\sum_{i=1}^{n}\log(\Pr(Y_1=y_1,Y_2=y_2,Y_3=y_3;\beta,\theta))$$
$$=\sum_{n=1}^{3}\log(\{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3}C_{13|2}(\frac{C_{12}(F_1(y_1-i_1;\beta_1),F_2(y_2;\beta_2);\theta_{12})-C_{12}(F_1(y_1-i_1;\beta_1),F_2(y_2-1;\beta_2);\theta_{12})}{F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)},$$
$$\frac{C_{23}(F_2(y_2;\beta_2),F_3(y_3-i_3;\beta_3);\theta_{23})-C_{23}(F_2(y_2-1;\beta_2),F_3(y_3-i_3;\beta_3);\theta_{23})}{F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)};\theta_{13|2})\}[F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)]) \quad (4.53)$$

Here $\beta=(\beta_1,\beta_2,\beta_3)$ and $\theta=(\theta_{12},\theta_{23},\theta_{13|2})$ where the marginal and copula parameters respectively. Then, the ML estimator $\hat{\theta}^{ML}$ is obtained by maximizing the above log-likelihood function over all parameters, β and θ, simultaneously.

For the general case, let the model for the t[th] time ordered margin imply a marginal distribution function $F_{it}(y_{it};\beta_t)$, where β_t are the marginal parameters and the subscript i denotes that we observe a sample $y_i=(y_{i1},y_{i2},...,y_{iT})'$ for i = 1, 2, ..., n. Similarly, the copula parameters for t-dimensional dependence are given by $\theta_{i,i+t|i+1,\ldots,i+t-1}$. Then, the ML estimator $\hat{\theta}^{ML}$ is obtained by maximizing the log-likelihood function over all parameters, β_t and $\theta_{i,i+t|i+1,\ldots,i+t-1}$, simultaneously.

In this optimization, good starting values are required. Starting values for the marginal parameters are obtained by maximum likelihood estimates of the marginal parameters estimated for all one at a time. And starting values for the copula parameters can be found by computing empirical Kendall's τ of bivariate copula function of the first tree which act as 'pseudo' data and then transformed back to the copula parameter using a known Bijection.

4.5 Pair Copula Construction for Multivariate Longitudinal Ordinal Data

In many fields of specializations including clinical trials, medicine, public health, social sciences, education, economics, psychometric and pharmacokinetics, multiple outcomes measured repeatedly over time from the same sets of study participants to analyse the changes over time, resulting in multivariate longitudinal data. The statistical analysis of this type of data are for studying changes across time by reducing the dimension of the multivariate longitudinal data to univariate longitudinal data using some kind of summary measures (Asar and İlk, 2014, Laffont et al., 2014), or jointly addressing the associations/dependencies across multivariate covariates and the changes across time points (Abegaz et al., 2015, Jiang, 2012, Verbeke et al., 2017). In multivariate longitudinal outcomes, the dependence among outcomes and changes over time must be accounted for in order to make valid inference. Our motivated example was household food security which has similar feature with multivariate longitudinal outcomes.

As we have discussed so far in the previous section of this chapter, the availability, accessibility and utilisation dimensions took ordinal levels based on the quartile score. For incorporating the fourth dimension, the stability of the three dimensions over time, the dimensions were repeatedly measured three times at six months interval. Hence, three longitudinal ordinal outcomes were obtained, resulting in multivariate longitudinal ordinal outcomes. The dimensions have pair-wise dependence between them at the same time point and each have dependence between consecutive time points (Capaldo et al., 2010, FAO, 2008). Therefore, modelling the stability and determinants of household food insecurity is the case of modelling multivariate longitudinal ordinal data that can consider the dependence between the dimensions and the dependence of each dimension between consecutive time points.

In modelling the stability and determinants of each household food security dimensions, the PCC model was proposed. A nice feature of the PCC approach in this setting is measuring the dependency of the three dimensions using the copula parameters, the parameter of the consecutive food security status of the households and the associated determinants of household food security for each dimension using the parameters of the marginal distributions. Hence, the pair copula construction approach with D-vine is attractive since it allows pairwise positive dependence structures and has closed form cumulative distribution function (cdf), no other copula family has both these properties.

In sum, this section proposes a model for estimating dependence and marginal parameters from multivariate longitudinal ordinal data using pair copula constructions via marginal model of the longitudinal ordinal logistic regression to our motivating problem. The performance of pair copula construction for discrete data was evaluated via Bernoulli and Poisson discrete distributions by (Panagiotelis et al., 2012) and found to be a good model. In our work, we follow their approach of constructing pair copula construction to model multivariate longitudinal ordinal data. As far as the researcher review of literature is concerned, no work has been conducted on multivariate longitudinal ordinal outcomes in the ordinal logistic version so far using pair copula construction. Hence, this study is concerned on implementing the developed discrete PCC model via marginal model of ordinal logistic regression for modelling household food insecurity status and determinant factors.

Even thought our aim is concerned on multivariate longitudinal ordinal data, let us review the continuous case and will continue to discrete one.

4.5.1 Pair Copula Construction for Multivariate Longitudinal Continuous Data

Consider an M-dimension multivariate continues random variables repeatedly measured for T time points of the i^{th} individual given by $Y_{ji} = (Y_{ji1}, Y_{ji2}, \dots, Y_{jiT})$ where $j = 1, 2, \dots, M$ and $i = 1, 2, \dots, n$. Smith (2015) re-ordered the observations of the multivariate series into the univariate outcomes of dimensions $N = T * n$ given by $Y = (Y_1, Y_2, \dots, Y_M)$, where $Y_1 = (y_{11}, y_{21}, \dots, y_{N1})'$, $Y_2 = (y_{12}, y_{22}, \dots, y_{N2})'$ and $Y_M = (y_{1m}, y_{2m}, ., y_{Nm})'$ (Smith, 2015).

Hence, the joint density function $f(y_1, y_2, \dots, y_m)$ is decomposed as follows:

$$f(y_1, \dots, y_m) = f_{1|2,\dots,m}(y_1 \mid y_2, \dots, y_m)\, f_{2|3,\dots,m}(y_2 \mid y_3, \dots, y_m) \dots f_m(y_m)\,. \qquad (4.56)$$

Recalling equations (4.4 and 4.12), we can simplify the bivariate case to

$$f(y_1, y_2) = c_{12}(F(y_1), F(y_2))\, f_1(y_1)\, f_2(y_2) \qquad (4.57)$$

where $c_{12}(.,.)$ is the appropriate pair-copula density for the pair of transformed variables $F_1(y_1)$ and $F_2(y_2)$.

Any transformation using the factorization of Equation (4.56), Equation (4.57) and Equation (4.4) different decomposition can be constructed. For example, the 3-dimensional case decomposition results in

$$f_{1,2,3}(y_1, y_2, y_3) = f_{1|2,3}(y_1 \mid y_2, y_3)\, f_{2|3}(y_2 \mid y_3)\, f_3(y_3) \qquad (4.58)$$

Using the Sklar's theorem, the conditional density of Y_2 *and* Y_3 in equation (4.58) is given by:

$$f_{2|3}(y_2 \mid y_3) = \frac{f_{2,3}(y_2, y_3)}{f_3(y_3)} = \frac{c_{23}(F(y_2), F(y_3))\, f_2(y_2)\, f_3(y_3)}{f_3(y_3)}$$
$$= c_{23}(F(y_2), F(y_3))\, f_2(y_2) \quad . \tag{4.59}$$

Similarly,

$$f_{1|2,3}(y_1 \mid y_2, y_3) = \frac{f_{13|2}(y_1 \mid y_2,\, y_3 \mid y_2)}{f_{3|2}(y_3 \mid y_2)}$$
$$= \frac{c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2))\, f_{1|2}(y_1 \mid y_2)\, f_{3|2}(y_3 \mid y_2)}{f_{3|2}(y_3 \mid y_2)}$$
$$= c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2))\, f_{1\backslash 2}(y_1 \mid y_2)$$

$$= c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2))\, f_{1|2}(y_1 \mid y_2)$$
$$= c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2)).c_{12}(F(y_1), F(y_2))\, f_1(y_1) \tag{4.60}$$

Using Equation (4.58), (4.59) and (4.60), the following decomposition appears

$$f_{1,2,3}(y_1, y_2, y_3) = c_{13|2}(F_{1|2}(y_1 \mid y_2),\, F_{3|2}(y_3 \mid y_2)).c_{12}(F(y_1), F(y_2))\, f_1(y_1).$$
$$c_{23}(F(y_2), F(y_3)).\, f_3(y_3).\, f_2(y_2).\, f_1(y_1). \tag{4.61}$$

This example illustrates the construction of a 3-dimensional density using the bivariate copula and the corresponding marginal distributions.

Similarly, for any other factor in Equation (4.56), the same procedure is possible using the general formula as follows:

$$f_{i|jk}(y_i|y_{j,k}) = c_{jk|k}\left(F_{i|k}(y_i|y_k), F_{i|k}(y_i|y_k)\right) f_{i|k}(y_i|y_k), \tag{4.62}$$

where k can be empty, a single index or multiple indices (Lennon, 2016).

For instance, the joint density of a four-dimension varieties can be decomposed into bivariate pair copulas using (4.62) as,

$$f(y_1, y_2, y_3\,, y_4) = f_{4|321}(y_4|y_3, y_2, y_1) f_{3|21}(y_3|y_2, y_1) f_{2|1}(y_2|y_1) f_1(y_1),$$
$$= c_{14|23} f_{4|23}(y_4|y_2,\ y_3).\, c_{13|2} f_{3|2}(y_3|y_2).\, c_{12} f_2(y_2) f_1(y_1),$$

$$= c_{14|23}c_{24|3}f_{4|2}(y_4|y_2)c_{13|2}c_{23}f_3(y_3)c_{12}f_2(y_2)f_1(y_1),$$
$$= c_{14|23}c_{24|3}c_{24}f_4(y_4)c_{13|2}c_{23}f_3(y_3)c_{12}f_2(y_2)f_1(y_1),$$
$$= c_{14|23}c_{13|2}c_{24|3}c_{34}c_{23}c_{12}f_4(y_4)f_3(y_3)f_2(y_2)f_1(y_1), \tag{4.63}$$

Hence, the decomposition in (4.63) can be written with full expression as:

$$f(y_1, y_2, y_3, y_4) = c_{12}(F_1(y_1), F_2(y_2))c_{23}(F_2(y_2), F_3(y_3))c_{34}(F_3(y_3), F_4(y_4)) \times c_{13|2}\left(F_{1|2}(y_1|y_2), F_{3|2}(y_3|y_2)\right)c_{14|2}(F_{1|2}(y_1|y_2), F_{4|2}(y_4|y_2)) \times c_{34|12}\left(F_{3|12}(y_3|y_1, y_2), F_{4|12}(y_4|y_1, y_2)\right)f_4(y_4)f_3(y_3)f_2(y_2)f_1(y_1)$$

Hence, based on the general form of equation 4.62, the decomposition of $f(y_1, y_2, \ldots, y_m)$ according to the D-vine pair copula construction can be written as

$$f(\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_\mathrm{m}) = \prod_{k=1}^{m} f(y_k)\prod_{j=1}^{m-1}\prod_{i=1}^{m-j} c_{i,i+j|i+1,\ldots,i+j-1}\{F(y_i \mid y_{i+1},\ldots,y_{i+j-1}), F(y_{i+j} \mid y_{i+1},\ldots,y_{i+j-1})\}, \tag{4.64}$$

Equation (4.64) is a product of $m * (m-1)/2$ bivariate pair copula densities and m marginal densities (Aas et al., 2009, Czado, 2010, Lennon, 2016, Ruscone and Osmetti, 2017, Smith et al., 2010, Smith, 2015). This leads to a large number of possible pair-copulas constructions. To organize all possible decompositions, a graphical model called a regular vine has been introduced by (Bedford and Cooke, 2002). Regular vine decompositions are concentrated only on the D-vines and C-vines, the special cases of regular vines. Detail of D-vine construction was displayed earlier in section 4.2.3.

D-Vine Parameter Estimation

For estimation of regular vine, different scholars proposed non-standard methods and standard estimation methods. Stepwise and MLE, IFM and SSP are the common standard estimation methods. MLE were considered for the first time by (Aas et al., 2009), IFM by (Joe, 1996), and SSP were by (Haff, 2012). These methods are designed for continuous data. We will not discuss here in detail since the current concern is on the discrete data. One can refer the references cited here for more detail. Just we now go to the PCC in discrete data.

4.5.2 Pair Copula Construction for Multivariate Longitudinal Discrete Data

Like the continuous case, consider an M-dimension multivariate discrete random variable repeatedly measured for T time points of the i^{th} individual given by $Y_{ji} = (Y_{ji1}, Y_{ji2}, \ldots, Y_{jiT})$ where $j = 1, 2, \ldots, M\ and\ i = 1, 2, \ldots, n$. We re-ordered the observations of the multivariate series into the univariate outcomes of dimensions $N = T * n$ given by $Y = (Y_1, Y_2, \ldots, Y_M)$, where $Y_1 = (y_{11}, y_{21}, \ldots, y_{N1})',\ Y_2 = (y_{12}, y_{22}, \ldots, y_{N2})' and\ Y_M = (y_{1m}, y_{2m}, .., y_{Nm})'$.
Hence, the joint probability mass function $Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_m = y_m)$ is decomposed as follows:

$$\Pr(Y_1, Y_2, \ldots, Y_m) = \Pr(Y_1 = y_1 | Y_2 = y_2, \ldots, Y_m = y_m) \times \Pr(Y_2 = y_2 | Y_3 = y_3, \ldots, Y_m = y_m) \times \ldots \times \Pr(Y_m = y_m) \tag{4.65}$$

The bivariate cumulative distribution function of (Y_1, Y_2) is given by the standard notation as

$$F_{ij}(y_i, y_j) = \Pr\ (Y_i \leq y_i, Y_j \leq y_j) \tag{4.66}$$

Similarly, the conditional cumulative distribution is given as follows:

$$F_{ij|k}(y_i, y_j | y_k) = \Pr\ (Y_i \leq y_i, Y_j \leq y_j | Y_k = y_k)$$

(4.67)

The expression in equation (4.65) has terms of the form $\Pr(Y_j = y_j \,|\, Y_{|j} = y_{|j})$ where $Y_{|j}$ is the vector of random variables $Y_1, Y_2, \ldots, Y_m$ excluding Y_j and $y_{|j}$ is the same vector for the realized values of the random variables. Choosing another element h from the vector of random variables, we can re-write the discrete joint probability in a similar fashion to the continuous case as following:

$$\Pr(Y_j = y_j \,|\, Y_{|j} = y_{|j}) = \frac{\Pr(Y_j = y_j, Y_h = y_h \,|\, Y_{|j,h} = y_{|j,h})}{\Pr(Y_h = y_h \,|\, Y_{|j,h} = y_{|j,h})} \tag{4.68}$$

Now, recalling the probability mass function and the multivariate copula function for discrete data in Equation (4.5), the bivariate conditional probability in the numerator can be expressed in terms of a copula giving

$$\Pr(Y_j = y_j \mid Y_{|j} = y_{|h})$$

$$= \frac{\sum_{i_j=0,1}\sum_{i_h=0,1} (-1)^{i_j+i_h} \Pr(Y_j \le y_j - i_j, Y_h \le y_h - i_h \mid Y_{|j.h} = y_{|j.h})}{\Pr(Y_h = y_h \mid Y_{|j.h} = y_{|j.h})}$$

$$= \frac{\sum_{i_j=0,1}\sum_{i_h=0,1} (-1)^{i_j+i_h} C_{Y_j.Y_h|y_{|j.h}}(F_{Y_j|Y_{|j.h}}(y_j - i_j), F_{Y_j|Y_{|j.h}}(y_h - i_h))}{\Pr(Y_h = y_h \mid Y_{|j.h} = y_{|j.h})} \tag{4.69}$$

The arguments in equation (4.69) of the copula functions are evaluated using the following (Nicklas, 2013, Panagiotelis et al., 2012, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015);

$$\begin{aligned} F_{Y_j|Y_h.Y_{|j.h}}(y_j \mid y_h, y_{|j.h}) = [&C_{Y_j.Y_h|Y_{|j.h}}(F_{Y_{j\backslash}|Y_{|j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h \mid y_{|j.h})) \\ &- C_{Y_j.Y_h|Y_{|j.h}}(F_{Y_{j\backslash}|Y_{\backslash j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h - 1 \mid y_{|j.h}))] \\ &/ \Pr(Y_h = y_h \mid Y_{|j.h} = y_{|j.h}) \end{aligned} \tag{4.70}$$

This vine PCC has nice feature than multivariate as well as Gaussian copulas functions in evaluating the probability mass function because the PCC requires $2m(m-1)$ evaluations whereas the multivariate and Gaussian copulas require 2^m evaluations (Panagiotelis et al., 2012).

4.5.2.1 D vine in Multivariate Longitudinal Discrete Data

For illustration purposes, we present in detail the 3-dimensional case. Therefore,

$$\begin{aligned} &\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) \\ &= \Pr(Y_1 = y_1 \mid Y_2 = y_2, Y_3 = y_3) \times \Pr(Y_3 = y_3 \mid Y_2 = y_2) \times \Pr(Y_2 = y_2) \end{aligned} \tag{4.71}$$

Utilizing Equation (4.71) the right-hand side of the first conditional probability can be rewritten as:

$$\Pr(Y_1 = y_1 \mid Y_2 = y_2, Y_3 = y_3) = \frac{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3} C_{13|2}(F_{1|2}(y_1 - i_1 \mid y_2), F_{3|2}(y_3 - i_3 \mid y_2))}{\Pr(Y_3 = y_3 \mid Y_2 = y_2)} \quad (4.72)$$

Similarly utilizing Equation (4.69), the first argument of the copula function in the numerator of Equation (4.72) is given by

$$F_{1|2}(y_1 - i_1 \mid y_2) = \frac{C_{12}(F_1(y_1 - i_1), F_2(y_2)) - C_{12}(F_1(y_1 - i_1), F_2(y_2 - 1))}{\Pr(Y_2 = y_2)}, \quad (4.73)$$

and the second argument can be expressed as

$$F_{3|2}(y_3 - i_3 \mid y_2) = \frac{C_{23}(F_2(y_2), F_3(y_3 - i_3)) - C_{23}(F_2(y_2 - 1), F_3(y_3 - i_3))}{\Pr(Y_2 = y_2)} \quad (4.74)$$

By cancelling terms and substituting, the probability mass function of the full expression for the 3-dimensional discrete D-vine is given by

$$\begin{aligned}\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) &= \{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3} C_{13|2}(\frac{C_{12}(F_1(y_1 - i_1), F_2(y_2)) - C_{12}(F_1(y_1 - i_1), F_2(y_2 - 1))}{F_2(y_2) - F_2(y_2 - 1)}, \quad (4.75)\\ &\frac{C_{23}(F_2(y_2), F_3(y_3 - i_3)) - C_{23}(F_2(y_2 - 1), F_3(y_3 - i_3))}{F_2(y_2) - F_2(y_2 - 1)})\}[F_2(y_2) - F_2(y_2 - 1)].\end{aligned}$$

Somewhat confusing to write the general D-vine structure; however, the general dimension algorithm for computing the probability mass function of a D-vine was outlined by Panagiotelis et al. (2012). It is evident both from this algorithm and the 3-dimensional example above that each bivariate pair copula only needs to be evaluated 4 times, specifically $C_{Y_j.Y_h|Y_{|j.h}}$ must be evaluated (Panagiotelis et al., 2012, Nicklas, 2013, Sirisrisakulchai and Sriboonchitta, 2014, Stöber et al., 2015);

$(F_{Y_j\backslash|Y_{|j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h \mid y_{|j.h}))$, $(F_{Y_j\backslash|Y_{|j.h}}(y_j - 1 \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h \mid y_{|j.h}))$, $(F_{Y_j\backslash|Y_{|j.h}}(y_j \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h - 1 \mid y_{|j.h}))$ and $(F_{Y_j\backslash|Y_{|j.h}}(y_j - 1 \mid y_{|j.h}), F_{Y_h|Y_{|j.h}}(y_h - 1 \mid y_{|j.h}))$.

In general, evaluation of the probability mass function requires 2m(m − 1) evaluations of bivariate copula functions, even though the continuous vine PCC is composed of only m(m−1)/2 pair copulas. These vine PCCs have still greater potential in high-dimensional settings since the computational burden of evaluating the pmf in the elliptical copulas is 2^m. As pointed out by Panagiotelis et al. (2012) among the major advantage of D-vine PCCs, a wide variety of dependence structures can be modelled by selecting different copula families as building blocks. Among these, Gaussian, t, AMH, Clayton, Frank and Gumbel copulas are the commonly used parametric copula families as building blocks.

The marginal probabilities in equation (4.75) can be modelled either of among discrete probability distributions. For the current study since the data are multivariate longitudinal ordinal data, the marginal model for univariate longitudinal ordinal data via cumulative logit model is used. Details of this model are given below.

Marginal Model of longitudinal ordinal outcomes

Let Y denotes ordinal response variable observed over T time points such that the response variable has C categories and labelled $(1,2,\ldots,C-1)$ and Y_{it} is the i^{th} individual at time t for $i=1,2,\ldots,n$ and $t=1,2,\ldots,T$. The marginal model for univariate longitudinal ordinal outcome via the cumulative logit model is given by:

$$logitP\left(Y_{it} \leq c \middle| X, \alpha_{c,} \beta\right) = \alpha_c + X'\beta \; for \; c = 1,2,\ldots,C-1 \tag{4.76}$$

where X is a vector of fixed or time varying covariates and β is a vector of unknown regression coefficients (Abegaz et al., 2015). Equation 4.76 implies

$$logit\boldsymbol{P}\left(Y_{it} \leq c \middle| X, \alpha_{c,} \beta\right) = \frac{\exp(\alpha_c + X'\beta)}{1+\exp(\alpha_c + X'\beta)} - \frac{\exp(\alpha_{c-1} + X'\beta)}{1+\exp(\alpha_{c-1} + X'\beta)} \tag{4.77}$$

Equation 4.76 or 4.77 is the marginal model of the proportional odds univariate ordinal logistic regression model. The parameters of this model are obtained by maximizing the likelihood function defined in 4.77. The regression coefficients in this model have simple interpretations in the population-average interpretation fashion in-terms of odds ratios. Equation (4.77) serves for the pair copula construction as a marginal distribution of the multivariate longitudinal ordinal cumulative logistic regression model.

Hence, the arguments in equation (4.75) can take the marginal distribution function given in equation (4.77). Moreover, the joint probability mass function in question (4.71) is expressed in terms of the pair copula functions and longitudinal ordinal marginal distributions. The newly constructed joint probability mass function can be called pair copula based multivariate longitudinal (marginal) cumulative logit model. Finally, this function can be estimated using the appropriate parameter estimation technique and appropriate bivariate copula families as building blocks.

4.5.2.2 Selection of Pair Copula Families

Before parameter estimations of the pair copula construction model, determining the order of the D-vine and the appropriate bivariate copula families are required. For the current study on food security, the conceptual framework for food security dimensions is given by FAO (2008) like Availability-Accessibility-Utilisation as it was discussed in section 4.3 of this thesis. We can use this order for the structure of D-vine for inference purpose. The difference in this chapter is that each dimension of food security consists of $N = T * n$ length of dataset. However, in section 4.3 the length of each dimension was only n, where $T \text{ and } n$ are the number of data collection phases and sample size, respectively. The reason that the length of the dataset became $N = T * n$ was because of reordering the multivariate time series based on the time order into one vector for each dimension.

The next step was selection of bivariate copula families of the vine distribution using the D-vine for every pair of variables. In Chapter 3, we customized the algorithm (Algorism I) developed for discrete pair copula bivariate copula family selection by Panagiotelis et al. (2015). We followed the same fashion of this algorithm for this chapter. For this chapter we have used marginal model via cumulative logit model to compute the “Pseudo” data. However, in Chapter 3, we have used simply cumulative logit model. “Pseudo” data were computed from the marginal model via cumulative logit model using Algorithm I to select appropriate bivariate pair families that fit the model best which have the smallest modified Akaike Information Criterion (mAIC).

4.5.2.3 Parameter Estimation of the D-Vine

For estimation of regular vine, details of different estimation techniques were assessed in Chapter 3. Here, we only concerned on MLE that will be applied for this application area.

Since the 3-dimensional D-vine is derived, hence the log-likelihood function of a 3-dimensional D-vine is given by

$$l(\beta,\theta;y)=\sum_{i=1}^{n}\log(\Pr(Y_1=y_1,Y_2=y_2,Y_3=y_3;\beta,\theta))$$
$$=\sum_{n=1}^{3}\log(\{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3}C_{13|2}(\frac{C_{12}(F_1(y_1-i_1;\beta_1),F_2(y_2;\beta_2);\theta_{12})-C_{12}(F_1(y_1-i_1;\beta_1),F_2(y_2-1;\beta_2);\theta_{12})}{F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)},$$
$$\frac{C_{23}(F_2(y_2;\beta_2),F_3(y_3-i_3;\beta_3);\theta_{23})-C_{23}(F_2(y_2-1;\beta_2),F_3(y_3-i_3;\beta_3);\theta_{23})}{F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)};\theta_{13|2})\}[F_2(y_2;\beta_2)-F_2(y_2-1;\beta_2)]) \quad (4.78)$$

Here $\beta=(\beta_1,\beta_2,\beta_3)$ and $\theta=(\theta_{12},\theta_{23},\theta_{13|2})$ where the marginal and copula parameters respectively. Then, the ML estimator $\hat{\theta}^{ML}$ is obtained by maximizing the above log-likelihood function over all parameters, β and θ, simultaneously.

For the general case, let the model for the j^th^ margin of the t^th^ time order imply a marginal distribution function $F_{ij}(y_{ij},\beta_j)$, where β_j are the marginal parameters and the subscript i denotes that we observe a sample $y_i=(y_{i1},y_{i2},\dots,y_{iM})'$ for $i=1,2,\dots,T*n$. Similarly, the copula parameters for M-dimensional dependence are given by $\theta_{i,i+j|i+1,\dots,i+m-1}$. Then, the ML estimator $\hat{\theta}^{ML}$ is obtained by maximizing the log-likelihood function over all parameters, β_j and $\theta_{i,i+j|i+1,\dots,i+m-1}$, simultaneously.

In this optimization, good starting values are required. Starting values for the marginal parameters are obtained by maximum likelihood estimates of the marginal parameters estimated from the marginal model via cumulative logit model for all dimensions. In addition, starting values for the copula parameters can be found by computing empirical Kendall's τ of bivariate copula function of the first tree which act as 'pseudo' data and then transformed back to the copula parameter using a known Bijection like we have computed for sections 4.3 and 4.4.

4.6 Statistical Software

"VineCopula" R package was used to compute the modified Akaike Information Criterion (mAIC) for selecting the best fitted bivariate copula families (Schepsmeier et al., 2015). Since the package did not include for t-bivariate copula owing to computational challenges, the t-copula is not proposed as a candidate in this thesis.

An "Alabama" R package was implemented to jointly estimate the marginal and copula parameters and their respective standard errors. Alabama is Augmented Lagrangian Minimization Algorithm for optimising smooth nonlinear objective functions with constraints. It allows both for Linear or nonlinear equality and inequality constraints (Varadhan and Grothendieck, 2011). It optimises using the nonlinear optimization with constraints. The package addressed this issue using "auglag" optimization. We wrote our own R code using "auglag" optimisation R package to estimate the parameters of the copula and the marginal distribution functions. Details of the R code were displayed in Appendix C I-III for all the three models, respectively.

4.7 Variable Selection

Variables that had p-value less than 0.2 in the preliminary analysis of the univariate analysis were incorporated for the final model of the univariate model. The significant variables in the final model were selected using forward wald method with significance level of 0.05. Variables that were significance with 0.05 p-values were incorporated for the multivariate, longitudinal and multivariate longitudinal ordinal models.

All statistical significant variables in each of the marginal model via the cumulative logit model were incorporated in the model for each dimension that hopes to be helpful as additional information with the existing knowledge in this area. In some of the tables of the result section, the blank space indicated that particular variable was not statistically significant for that dimension in the marginal model and hence that particular variable was not included only for that particular dimension during the estimation of the final joint model.

Chapter Five

5. Analysis of the Household Food Data

5.1 Internal Validity Assessment of the Data Collection Tools

Internal consistency of the data collection instrument for household food security in terms of the three dimensions was assessed using Cronbach's Alpha. The Cronbach's Alpha value computed as 0.735 which is in the acceptable internal consistence level since it is above 0.7. The Cronbach's Alpha value of a particular item deleted was assessed for each item. The detail summary statistics was displayed in appendix B.

5.2 Characteristics of the Study Participants

Three phases of data collection were conducted from the same household head at six months interval. The study included a total of 630 households after the removal of 2.5 % (16) of the respondents because they have dropped out at least one of the data collection phases. Since the study was longitudinal and have both time varying and fixed covariates, for the time varying variables we used all responses from the three follow-up interviews and arranged the data into a vector according to the time points while for fixed covariates we used the response of one time point. The summary statistics is displayed in Tables 5.1 and 5.2 for the fixed and time-varying covariates, respectively.

Household Head Characteristics

Table 5.1 showed that husbands headed majorities of the household heads (more than 85%). Almost 69% of the households did not have formal education of which 38% were unable to read and write and 31% can read and write. Only 31% of the household heads attended formal education of which 6% was secondary school or above completion. Two thirds of the household heads were married and living together. Twelve percent of them were cohabitating but not married and 7% were widowed.

Table 5.1 also showed majorities of the households consisted of four to six family members which accounts 65% of the total respondents. Nineteen percent of the respondents had more than six family members whereas 18% of them have less than four family members.

Table 5. 1: Summary measures of non-time varying variables of the households

Variables	Frequency	Percent
Household Head (HH)		
Husband	548	86.9
Wife	71	11.3
Son/daughter	11	1.7
Highest level of education attained by HH		
Unable to read and write	241	38.2
Can read and write	192	30.5
Regular Primary education (1-8)	157	24.9
Secondary education and above	40	6.4
HH current marital status		
Never married	13	2
Cohabiting	74	11.7
Married	475	75.4
Divorced	23	3.7
Widowed	45	7.2
Family Size of HH		
Three and Less	110	17.5
Four to Six	400	63.5
Seven and Above	120	19.0
Study Site		
Kutaber	204	32.5
Kalu	234	37.2
Tehuledere	192	30.5
Total farmland size in Hectare		
<=0.5 Hectare	383	60.8
> 0.5 Hectare	247	39.2

Table 5. 1: Summary measure of non-time varying variables of the households … continued

Variables	Frequency	Percent
Quality or fertility of land ploughed		
Fertile	75	11.9
Medium fertile	422	66.9
Less fertile	133	21
HH Income Source		
Only Farming	261	41.4
Both Farming and Off-farming	369	58.6
Types of Cereal Crops Cultivated		
One type	169	26.8
two type	187	29.7
Three and more types	274	43.5
Types of livestock		
One and less type	156	24.8
Two to Three types	275	43.7
Four or more types	198	31.5
Agro-ecology of study site		
Hot (Kolla)	146	23.2
Medium (Weinadega)	340	54.0
Cold (Dega)	144	22.9

Income Source of the Household

As Table 5.1 showed, 44% of the households obtained their income from farming only whereas 59% of them practice off-farming activities to get their income. About 73% of the households had harvested more than one type of cereal crops. In addition to farming, the households participated in different agricultural activities, among these two thirds of the households participated in more than one type of livestock activities.

Land Size and Fertility Characteristics

Based on the information in Table 5.1, around 60% of the households ploughed not greater than 0.5 hectares and 67% of them had medium fertile land for cultivation purposes. Only 11% households had fertile land for plough and 21% less fertile. Among the total households in this sample, 74% of them cultivated once per year whereas 26% of them cultivated twice or triple per year.

Table 5. 2: Summary measures of the time-varying variables of the households

Characteristic variables	Data collection phases					
	Phase one		Phase two		Phase three	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Shortage of rainfall						
No	117	18.6	378	60.0	92	14.6
Yes	513	81.4	252	40.0	538	85.4
Crop Disease						
No	422	67.0	328	52.1	421	66.8
Yes	208	33.0	302	47.9	209	33.2
Increase in market price						
No	385	61.1	476	75.6	452	71.7
Yes	245	38.9	154	24.4	178	28.3
Use of Pesticides						
No	501	79.5	332	52.7	501	79.5
Yes	129	20.5	298	47.3	129	20.5
Presence of Pests						
No	259	41.1	239	37.9	285	45.2
Yes	371	58.9	391	62.1	345	54.8

Agro-Ecology of the Study Site

The study was conducted in three selected South Wollo Zone Woredas, namely, Kutaber, Kalu and Tehuledere. As Table 5.1 revealed that around 33, 37 and 31 percent of the households were from Kutaber, Kalu and Tehuledere Woredas respectively. Moreover, almost half (54%) of the study areas have medium (Weinadega) agro-ecology and the rest proportion was covered by hot (Kola) and cold (Dega) agro-ecology with almost equal proportion.

Time-Varying Characteristics

Table 5.2 shows the distribution of some of the time-varying covariates during the three-phase data collection. Based on Table 5.2, the presence of rainfall, disease of the cultivation, presence of market price increase, use of pesticides, and presence of pests vary with time in the study area. More than 80% of rainfall shortage was observed in the first and third phases, and smallest 40% during the second phase.

Moreover, Table 5.2 showed that 48% of cultivation disease was reported in the second phase but 33% each in the first and third phase. The highest market price increase was observed in phase one followed by phase three and two, 39%, 28%, and 24% respectively. The presence of pests had highest proportion in phase two followed by phase one and three, 62, 59 and 55 percent, respectively.

5.3 Household Food Security Analysis

The household food security status was computed for each dimension at each data collection phase using the quartile score as presented in Table 5.3. Table 5.3 showed that food secured households was smaller in phase three for food availability and utilisation dimensions than phases one and two. Conversely, the proportion of food secured households was higher in phase three for accessibility dimension than the other phases.

Based on the information in Table 5.3, the proportion of chronically food in-secured households decreased from 19% to 2% and from 3% to 0% in accessibility and utilisation dimensions respectively when one goes from phases one and two to phase three. However, in availability dimension, the proportion of chronically food in-secured was almost similarly distributed between the three data collection phases.

Table 5.3 also showed the highest proportion of mildly food in-secured (60%) was found in food utilisation during the third phase followed by 42% in accessibility. The highest proportion (51%) and (50%) of moderately food in-secured households were observed in utilisation of phase one and two respectively followed by 45% in availability of phase three.

Furthermore, the household food security status was determined using the composite index method combining the household food security status of the three data collection phases into one vector for each dimension. The result of this computation was summarized in Table 5.4.

Table 5. 3: Household food security statuses of the three food security dimensions in all data collection rounds (*f* stands for frequency and *%* for percent)

Food Security Dimensions	Household food security Status	Data collection phases					
		Phase I		Phase II		Phase III	
		f	*%*	*F*	*%*	*F*	*%*
Availability	Chronically food in-secured	32	5.1	33	5.2	41	6.5
	Moderately food in-secured	215	34.1	218	34.6	284	45.1
	Mildly food in-secured	254	40.3	254	40.3	209	33.2
	Food secured	129	20.5	125	19.8	96	15.2
	Total	630	100.0	630	100.0	630	100.0
Accessibility	Chronically food in-secured	119	18.9	118	18.7	14	2.2
	Moderately food in-secured	219	34.8	216	34.3	230	36.5
	Mildly food in-secured	222	35.2	223	35.4	263	41.7
	Food secured	70	11.1	73	11.6	123	19.5
	Total	630	100.0	630	100.0	630	100.0
Utilisation	Chronically food in-secured	20	3.2	20	3.2	1	.2
	Moderately food in-secured	318	50.5	314	49.8	228	36.2
	Mildly food in-secured	213	33.8	216	34.3	379	60.2
	Food secured	79	12.5	80	12.7	22	3.5
	Total	630	100.0	630	100.0	630	100.0

Based on the composite index summary statistics presented in Table 5.4, the highest proportion of chronically food in-secured was observed in accessibility dimension (13.3%) followed by in availability and utilisation (5.6% and 2.2%) respectively. Similarly, in-terms of moderately food in-secured and mildly food in-secured classification, the highest proportion were observed in utilisation followed by availability and then accessibility. Moreover, the highest proportion of food secured households was observed in availability followed by accessibility and then utilisation.

Table 5. 4: The Composite Food Security Status of households in the three data collection rounds

Food security status	Availability		Accessibility		Utilisation	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Chronically food in-secured	106	5.6	251	13.3	41	2.2
Moderately food in-secured	717	37.9	665	35.2	860	45.5
Mildly food in-secured	717	37.9	708	37.5	808	42.8
Food secured	350	18.5	266	14.1	181	9.6
Total	1890	100.0	1890	100.0	1890	100.0

5.4 A Pair Copula Construction Approach for Multivariate Ordinal Data with Application to Household Food Insecurity Data

5.4.1 Kendall's tau Correlation Coefficient

The bivariate correlation coefficients between the food security dimensions were computed through the non-parametric correlations called "Kendall's tau_b" using the raw discrete food security data and presented in Table 5.5.

Similarly, the Kendall's tau was also computed from the pseudo data of food security dimensions (Availability, Accessibility and Utilisation) using cumulative logit marginal distributions as presented in Table 5.6 and this result is further used for preliminary analysis to select appropriate copula families for the copula cumulative logit model.

Table 5. 5: Nonparametric correlation coefficients and their significant α values in brackets of the household food security status of the dimensions

	Availability	Accessibility	Utilisation
Availability	1	-0.051 (0.128)	.092 (0.011)
Accessibility		1	-.199 (.000)
Utilisation			1

Based on the results in Table 5.5 and 5.6, one can observe that almost all similar and consistent results were found and conclude that food utilisation had positive and significant correlation with food availability but negative with accessibility.

Table 5. 6: The Kendall's tau for the pseudo data computed from the cumulative logit marginal distribution for the application data availability, accessibility and utilisation.

Dimensions	Availability	Accessibility	Utilisation
Availability	1.000000	-0.05132	0.091865
Accessibility	-0.05132	1.000000	-0.19932
Utilisation	0.091865	-0.19932	1.000000

5.4.2 PCC Selection

The preliminary analysis computed in Table 5.7 reduces the number of bivariate copula families and the result showed that the Clayton and Gumbel copulas are not applicable for measuring the dependence between availability, accessibility and accessibility and utilisation because their Kendall's tau showed negative dependence.

For those satisfied the preliminary analysis, Algorithm I was employed and their corresponding modified Akaike Information Criterion (mAIC) were computed as presented in Table 5.7. The D-vine structure of household food security in the first tree was (availability, accessibility) and (accessibility, utilisation). Similarly, in the second tree the structure is (availability | accessibility and utilisation | accessibility). From Table 5.7, one can deduce that the best fitted copula families on the first tree were AMH for edge 1 since the mAIC is smaller than the other copula families. However, for edge 2 the Gaussian and Frank copulas have negligible mAIC difference between them. As a result, we set additional criterion in-terms of parsimonious that the model provides. The Gaussian copula models are a natural choice for integer-valued covariates with interpretable parameters (Lennon, 2016). A Frank copula can capture a wide range of dependence including positive and negative dependence and belong to the Archimedean family with a closed form of distribution functions and benefits of easy computation (Yang et al., 2020). Frank bivariate copula is the ideal candidate in this study. Hence, AMH and Frank copula families were selected as best fit bivariate copula that serves for parameter estimation in the full maximum likelihood parameter estimation for the corresponding vine structure they fitted best.

Table 5. 7: The summary of copula families by applying Algorithm 1 to the data and estimating modified Akaike Information Criterion (mAIC) to select the best fit copula families.

Copula	Modified Akaki Information Criterion (mAIC)		
	Av, Ac	Ac, Ut	Av |Ac, Ut |Ac
Gaussian	4736.193	4465.663	4281.878
Clayton	NA	NA	4290.001
Gumbel	NA	NA	4281.459
Frank	4736.951	**4465.641**	**4280.713**
AMH	**4736.076**	4469.087	4282.067
Independent	4739.37	4498.055	4285.219

where Av = Availability, Ac = Accessibility and Ut = Utilisation

5.4.3 Estimation of the Copula and Marginal Parameters

Table 5.7 showed the dependence between availability and accessibility was expressed by AMH copula. On the other hand, Frank expressed the dependence between accessibility and utilisation dimensions. Similarly, like the pervious one the dependence between availability given accessibility and utilisation given accessibility was also expressed by Frank copula. Here, the conditioning of accessibility is assumed to be not affecting the dependence between availability and utilisation.

Our main purpose in this chapter was to estimate jointly the dependence between food security dimensions and their corresponding predictor factors at household level. We applied MLE using the selected bivariate copula families and the cumulative logit marginal distribution functions. Derivation of the likelihood and log-likelihood function was made as well (see Appendix B I).

We wrote our own R code using “auglag” optimization R package to estimate the parameters of the copula and the marginal distribution functions. Details of the R code were displayed in Appendix C I. The estimated values of the parameters and the corresponding standard errors for the copula functions were displayed in Table 5.8 and for marginal distributions in Table 5.9.

Copula Parameter

Table 5.8 summarizes the results of the estimated dependence parameters of the selected bivariate copula families. The result showed positive dependence was observed between all dimensions of household food security statuses. AMH copula measures the dependence between availability and accessibility household food security status, Frank measures the accessibility and utilisation and Frank availability and utilisation given that the accessibility dimension has been happened and found that positive dependence was observed. Moreover, these dependences were observed as statistically significant.

Table 5. 8: The estimates of dependence parameters using the selected pair copula for the application data of multivariate ordinal household food security status.

Tree	**Copula family**	**Estimated Parameter**	**Estimated SE**	**Bijection tau**
I	AMH	0.999	0.2612	0.3333267
	Frank	1.4053	0.2425	0.1531605
II	Frank	1.18358	0.3152	0.1297094

Marginal Parameter

Several variables were incorporated in the model that hopes to be helpful as additional information with the existing knowledge in this area. Among the incorporated variables, the presence of crops/vegetables/fruits disease, shortage of rainfall, cultivating once a year and small land size cultivated were identified the potential statistically significant variables that lead the household to be chronically, mildly and moderately food in-secured in all dimensions.

On the other hand, Table 5.9 showed that study site contributes on the status of household food insecurity in availability and utilisation dimensions. Moreover, cold (dega) agro-ecology leads household to be chronically, mildly, and moderately food in-secured in availability and accessibility dimensions. Similarly, the presence of market price increase and household head headed by son/daughter were more likely to be chronically, mildly and moderately food in-secured than headed by husband at availability and utilisation dimensions respectively.

Table 5. 9: Summary results of pair copula based cumulative logit model parameters estimates for the household food security data.

.		**Availability**		**Accessibility**		**Utilisation**	
Variable	Categories	Estimates	S.E	Estimates	S.E	Estimates	S.E
	1|2	**-4.3413***	1.123	**-4.6699***	1.085	-1.58125	1.189
Intercept	2|3	-1.4085	1.105	**-2.6982***	1.075	2.16749	1.198
	3|4	0.901668	1.104	-0.61065	1.070	**4.2333***	1.202
Household Head (HH)							
	Husband	-0.41414	.608	-0.64717	.594	**-1.2361***	.606
	Wife	-0.69952	.638	-0.73125	.621	-0.9262	.633
	Sibling						
Education Level (HH)							
	Regularly educated	-0.16271	.184	-0.16329	.176	-0.17596	.185
	Regularly uneducated						
Marital Status (HH)							
	Never married or Cohabiting	**-1.0803***	.370	**0.67283***	.253	-0.56257	.367
	Married	-0.14025	.315	0.380646	.301	-0.3404	.310
	Divorced or Widowed						
Study Site (Woredas)							
	Kutaber	**-1.2178***	.476	-0.13779	.457	**1.04154***	.427
	Tehuledere	-0.39089	.452	0.183894	.437	**0.81636***	.408
	Kalu						
Land Size Cultivated							
	Less than 0.5 Hectare	**0.45521***	.161	**0.06726***	.029	**0.24435***	.110
	Above 0.5 Hectare						
Land Fertility							
	Fertile	-0.49066	.311	-0.0784	.297	-0.58645	.322
	Medium fertile	-0.4183	.208	-0.07936	.197	-0.08722	.205
	Less fertile						

* indicates significant at 5% level of significance

Table 5.9 continued……………. (* indicates significant at 5% level of significance)

		Availability		Accessibility		Utilisation	
Variable	Categories	Estimates	S.E	Estimates	S.E	Estimates	S.E
Cultivate time							
	Yearly	**0.94121***	.438	**-1.3246***	.427	**1.3303***	.496
	Biannual and more						
Shortage of Rainfall							
	Yes	**0.64373***	.235	**0.56948***	.225	**0.34071***	.144
	No						
Crops/Vegetables Disease							
	Yes	**1.6343***	.187	**0.20649***	.068	**0.24712***	.076
	No						
Market Price Increase							
	Yes	**-0.5452***	.171	0.217151	.163	-0.06202	.172
	No						
Agro-ecology of study site							
	Cold (Dega)	**-1.5768***	.356	**-0.8357***	.337	-0.04717	.351
	Medium (Wenadega)	**-1.1473***	.285	-0.48742	.271	0.08621	.282
	Hot (Kolla)						

5.4.4 Effects of PCC on the Univariate Cumulative Logit Model

To assess the effect of the PCC model on the usual univariate cumulative logit model, the univariate estimates for each dimension was displayed in Table 5.10. The following comparisons were made between the marginal parameters of the PCC-Based cumulative logit model presented in Table 5.9 and univariate cumulative logit model presented in Table 5.10. The PCC- based cumulative logit model identified more significant determinants for households to be food in-secured in all of the three dimensions over the univariate cumulative logit model.

In availability dimension, both PCC and univariate cumulative logit models identified marital status classified under never married or cohabitated, study site, small cultivable land, shortage of rainfall, presences of crop disease, presences of market price increase, and clod agro-ecology as positive predictors for households to be severe to mildly food in-secured compared with their counterparts. In addition to these determinants, the PCC model identified medium agro-ecology and yearly once cultivation season as positive predictors for household to be food in-secured. Conversely, the univariate model identified less fertile cultivable land as positive predictor for food insecurity.

A yearly based cultivation of agricultural activities and cold agro-ecology were the determinants of household food insecurity in the accessibility dimension both in the PCC and univariate cumulative logit model. However, the PCC model identified additional determinants for households to be food in-secured include never married or cohabitated household headed, small cultivable land, shortage of rainfall, and occurrences of crop disease.

The study site, small cultivable land and yearly once cultivation activities were identified as significant determents for households to be severe to mildly food in-secured in the utilisation dimension both in the two models. Like accessibility dimension, the PCC identified marital status grouped under never married or cohabitated and occurrences of cultivation diseases as additional determinants for households to be food in-secured.

The majority of the estimates obtained through PCC model were overestimated in availability and accessibility dimension. However, in utilisation dimension, the PCC model underestimated almost all of the determinants of household food insecurity. Since PCC identified more significant determinants in all of the three food security dimensions, interpretation and discussion of determinants were made using the estimates of the PCC model throughout this topic.

Table 5. 10: Summary results of the univariate cumulative logit model parameters estimates for the household food security data.

.		Availability		Accessibility		Utilisation	
Variable	Categories	Estimates	S.E	Estimates	S.E	Estimates	S.E
	1|2	**-5.861***	.810	-1.037	.761	**-6.562***	0.893
Intercept	2|3	**-3.057***	.785	0.763	.763	**-2.793***	0.837
	3|4	-0.696	.774	**2.842***	.768	-0.896	0.833
Household Head (HH)							
	Husband	-.514	.590	-0.865	.584	-1.065	.615
	Wife	-.698	.619	-0.751	.611	0.908	645
	Sibling						
Education Level (HH)							
	Regularly uneducated	.192	.183	0.242	.177	0.112	.185
	Regularly educated						
Marital Status (HH)							
	Never married or Cohabiting	**-1.090***	.374	0.647	.359	-0.544	.370
	Married	-.231	.311	0.185	.299	-0.184	.313
	Divorced or Widowed		.				
Study Site (Woredas)							
	Kutaber	**-1.258***	.467	-0.223	.447	**1.117***	.532
	Kalu	-.445	.443	0.058	.426	-0.910	.508
	Tehuledere						
Land Size Cultivated							
	Less than 0.5 Hectare	**.432***	.161	-0.050	.154	**0.349***	.161
	Above 0.5 Hectare						
Land Fertility							
	Fertile	-.492	.307	-0.086	.309	-0.582	.316
	Medium fertile	**-.503***	.210	-0.267	.200	0.066	.203
	Less fertile						

* indicates significant at 5% level of significance

Table 5.10 continued……………. (* indicates significant at 5% level of significance)

		Availability		Accessibility		Utilisation	
Variable	Categories	Estimates	S.E	Estimates	S.E	Estimates	S.E
Cultivate time							
	Yearly	.850	.432	**-1.586***	0.421	**1.542***	0.495
	Biannual and more		.				
Shortage of Rainfall							
	Yes	**.586***	.232	0.388	0.227	**0.492***	0.241
	No						
Crop/vegetable Disease							
	Yes	**1.650***	.187	0.159	0.170	0.290	0.174
	No						
Market Price Increase							
	Yes	**-.578***	.174	0.132	0.165	-0.010	0.170
	No						
Agro-ecology of study site							
	Cold (Dega)	**1.622***	.235	**0.920***	0.340	-0.012	0.354
	Medium (Wenadeg)	.4308	.360	0.331	0.214	0.158	0.224
	Hot (Kolla)						

5.5 A Pair Copula Construction Approach for Longitudinal Ordinal Data with Application to Household Food Insecurity Data

5.5.1 Kendall's tau Correlation Coefficient

The bivariate correlation coefficients between the successive time point’s food security statuses were computed through the non-parametric correlations called “Kendall's tau_b” using the raw discrete food security data and presented in Table 5.11.

The Kendall’s tau was computed from the pseudo data of food security statuses of the three-phase using cumulative logit marginal distributions as presented in Table 5.12 and this result further used for preliminary analysis to select appropriate copula families for the copula cumulative logit model.

Table 5. 11: Nonparametric correlation coefficients of the household food security states of the three data collection phases

	Phase I	**Phase II**	**Phase III**
Phase I	1	.486**	.214**
Phase II	.486**	1	.253**
Phase III	.214**	.253**	1

** indicates significant at 0.01 level of significance.

Table 5. 12: The Kendall's tau for the pseudo data computed from the cumulative logit marginal distribution for the application of three-phase longitudinal data.

	Phase I	**Phase II**	**Phase III**
Phase I	1.000	0.487	0.211
Phase II	0.487	1.000	.250
Phase III	0.211	.250	1.000

Based on the results in Tables 5.11 and 5.12, one can observe that similar and consistent results were found, and it can be said that positive and significant correlation were observed between the successive food security phases.

5.5.2 PCC Selection

The preliminary analysis computed in Table 5.12 showed that the Kendall's tau of the pseudo data computed from the cumulative logit model was positively correlated. The correlation between successive time points of food security status was positive so that all bivariate copula families listed in this thesis can be the candidate for measuring the dependence between successive food security statuses of household. Algorithm II was used to select the best fit bivariate copula for this application data. For the time being, for minimizing computational challenges, t copula is not used. The result of pair copula selection process is presented in Table

5.13. Algorithm II was employed and their corresponding modified Akaike Information Criterion (mAIC) were computed as presented in Table 5.13.

Table 5. 13; Summary of copula families by applying Algorithm II to the data and estimated modified Akaike Information Criterion (mAIC) to select the best fit copula families.

	Modified Akaike Information Criterion (mAIC)		
Copula	(Y_1, Y_2)	(Y_2, Y_3)	$(Y_1 \| Y_2, Y_3 \| Y_2)$
Gaussian	4011.59	4140.945	3720.277
Clayton	4098.848	4140.945	3721.848
Gumbel	**3995.368**	**4132.143**	**3715.798**
Frank	4009.531	4158.39	3720.474
AMH	4227.585	4180.113	3720.715
Independent	4406.431	4213.82	3733.431

Where; Y_1= household food security status at the first 12 months considered as baseline food security status

Y_2= household food security status at the middle six months

Y_3= household food security status at the last six months

The D-vine structure of the longitudinal household food security status in the first tree was (Y_1, Y_2) and (Y_2, Y_3). Similarly, in the second tree the structure is $(Y_1|Y_2, Y_3|Y_2)$. From Table 5.13, one can deduce that the best fitted copula families on the first tree were Clayton and Gumbel for edge 1 and 2 respectively while the second tree has AMH copula family since the mAIC is smaller than the other copula families. Hence, Clayton, Gumbel and AMH copula families were selected as best fit bivariate copula that serves for parameter estimation in the full maximum likelihood parameter estimation for the corresponding vine structure they fitted best.

5.5.3 Estimation of the Copula and Marginal Parameters

The best fitted bivariate copulas were selected using Algorithm II as presented in Table 5.13. Gumbel bivariate copula was selected to measure the dependence between first and second household food security status, and second and third phase. Similarly, it was also selected for the first and the third phase given that the second household food security status. The corresponding marginal distribution for the application data was cumulative logit model.

The main purpose of this study was to address the dependence of household food security status over time and their corresponding predictor factors at household level. MLE approach was used to compute jointly the dependence parameters using the selected bivariate copula families and the marginal parameters using cumulative logit model. Derivation of the likelihood and log-likelihood function was made and presented in Appendix B II.

We wrote our own R code using “auglag” optimization R package to estimate the parameters of the copula and the marginal distribution functions. Details of the R code were displayed in Appendix B II. The estimated values of the parameters and the corresponding standard errors both for the copula and marginal distributions were displayed in Table 5.14 and 5.15 respectively.

Copula Parameter

Table 5.14 summarizes the results of the estimated dependence parameters of the selected bivariate copula families. The result shows positive dependence was observed between all phases of household food security statuses. Gumbel copula measures the pairwise dependence between all of the three phases of the individual household food security status. Moreover, these dependences were observed as statistically significant. As a result, this leads to the conclusion that individual household food security status varies with time. Therefore, the household food security status in the study area is not stable over time.

Table 5. 14: The estimates of dependence parameters using the selected pair copula for the application data of longitudinal household food security status.

Tree	**Copula family**	**Estimated Parameter**	**Estimated SE**	**Bijection tau**
I	Gumbel	1.2511	0.4452	0.2007034
	Gumbel	1.1987	0.4625	0.1657629
II	Gumbel	1.057081	0.4531	0.0539987

Table 5. 15: Summary results of the marginal parameters of pair copula based longitudinal cumulative logit model for the household food security data.

		Y_1		Y_2		Y_3	
Variable	**Categories**	**Estimates**	**S.E**	**Estimates**	**S.E**	**Estimates**	**S.E**
	1|2	**-2.55073**	0.601	**-2.1034**	0.59882	**-3.24248**	0.6156
Intercept	2|3	0.166815	0.583	-0.06289	0.58211	-0.15301	0.57222
	3|4	**2.93809**	0.599	**3.29915**	0.60151	**2.42231**	0.57948
Times of cultivate within a year							
	Yearly	0.018707	0.186	**-0.5623***	0.1868	-0.01574	0.18434
	Biannual and more						
Crop disease							
	Yes	**0.748778***	0.177	**0.596795***	0.17873	**0.731031***	0.17284
	No						
Increase in market price							
	Yes	**-0.33019***	0.169	**-0.61814***	0.16923	**-0.53993***	0.16799
	No						
Weathering condition of the village							
	Cold (Dega)	0.21021	0.212	-0.04505	0.16923	**-0.4704**	0.21033
	Medium (Weinadega)	**0.761084***	0.293	**0.557457***	0.2133	**-0.70407***	0.28844
	Hot (Kolla)						
Availability of rain							
	Little	-0.39891	0.533	-0.40582	0.53077	0.446841	0.52183
	Enough	-0.53664	0.544	-0.73143	0.54211	0.960943	0.53419
	High						

Marginal Parameter

Five time-vary covariates were included in this study to assess the effect of these variables on the household food security status over time. Summary statistics were computed in Table 5.15. The result shows a statistically significant difference in the marginal parameter between the presence and the absence of crop disease, the presence of market price increase and not, and hot weathering condition and medium weathering condition in all time points of the household food security status. Areas where crop disease happened are more likely to lead households to be chronically, moderately and mildly food in-secured than areas not crop disease happened. In addition, hot weather conditions are more likely to lead households to be chronically, moderately and mildly food in-secured than medium weather conditions in all time points. Likewise, increased market price is another factor to lead households chronically, moderately and mildly food in-secured than stable market price in all phases household data.

Moreover, households cultivating once a year are more likely chronically, moderately and mildly food in-secured than those cultivating two or more time a year in the second phase of household data. In contrast, in this study, the only time-varying covariate that did not affect the household food security status in all time points was availability of rainfall.

5.5.4 Effects of PCC Model on the Univariate Cumulative Logit

To compare the effects of PCC model on the cumulative logit model in fitting longitudinal ordinal data, the finding of both the PCC model and univariate model in the cumulative version were presented in Table 5.15 and 5.16 respectively. Both the PCC and univariate models identified the presence of cultivation disease and cold agro-ecology as positive and significant determinants for households to be food in-secured in the first round. Furthermore, the PCC model identified the presence of market price as determinant of household food insecurity. Likewise, both models identified cultivation once a year, crop disease and cold agro-ecology determinants for households to be food in-secured in the second round. Moreover, the PCC model identified the presence of market price as determinant of household food insecurity.

In the third round, the univariate model identified one more predictor for household food insecurity than the PCC model. Crop disease, market price increase and hot and medium agro-ecology were identified as predictors of household food insecurity through both the PCC and univariate models. The PCC model drops the yearly once cultivation season while it was significant in the univariate model.

Table 5. 16: Summary results of the marginal parameters of the univariate cumulative logit model for the household food security data

		Y_1		Y_2		Y_3	
Variable	**Categories**	**Estimates**	**S.E**	**Estimates**	**S.E**	**Estimates**	**S.E**
	1|2	**-3.328***	0.628	**-1.594***	0.410	**-4.677***	0.699
Intercept	2|3	-0.717	0.606	**0.962***	0.394	**-1.442***	0.653
	3|4	**2.122***	0.613	**3.792***	0.426	1.098	0.652
Times of cultivate within a year							
	Yearly	0.284	0.193	**-0.423***	0.178	**0.789***	0.176
	Biannual and more						
Crop disease							
	Yes	**1.033***	0.177	**0.426***	0.190	**0.989***	0.199
	No						
Increase in market price							
	Yes	-0.257	0.169	-0.213	0.201	**-0.830***	0.195
	No						
Agro-Ecology of Study Site							
	Cold (Dega)	**0.901***	0.300	0.277	0.266	**-0.620***	0.226
	Medium (Weinadega)	0.088	0.208	0.433	0.222	**-0.668***	0.289
	Hot (Kolla)						
Availability of rain							
	Little	-0.657	0.551	**1.448***	0.414	-0.150	0.613
	Enough	-0.537	0.544	0.573	0.306	0.310	0.601
	High						

In all of the significant determinants in the availability dimension, the PCC model underestimated the parameters of the marginal model. In the accessibility dimension, 50% of the predictors were overestimated and the rest of them were underestimated. Similarly, in the utilisation dimension, some of the significant predictors were overestimated and some of them were underestimated. The interpretation and discussion of the finding were made using the finding of the PCC model.

5.6 A Pair Copula Construction Approach for Multivariate Longitudinal Ordinal Data with Application to Household Food Insecurity Data

5.6.1 Kendall's tau Correlation Coefficient

The bivariate correlation coefficients between the successive food security dimensions for the combined data were computed through the non-parametric correlations called “Kendall's tau_b” using the raw discrete food security data and presented in Table 5.17.

The Kendall’s tau was also computed from the pseudo data of food security statuses of the three dimensions using marginal model of the cumulative logit marginal distributions presented in section 4.5 equation (4.77) as presented in Table 5.18. This result is further used for preliminary analysis to select appropriate copula families for the copula marginal model via cumulative logit model.

Table 5. 17: Nonparametric correlation coefficients of the household food security states of the three combined dimensions

	Availability	**Accessibility**	**Utilisation**
Availability	1	0.0234(0.24)	082** (0.000)
Accessibility	0.0234(0.24)	1	-.096** (0.000)
Utilisation	082** (0.000)	-.096** (0.000)	1

** indicates significant at 0.01 level of significance.

Table 5. 18: The Kendall’s tau for the pseudo data computed from the marginal model of the cumulative logit marginal distribution for the multivariate longitudinal data.

	Availability	Accessibility	Utilisation
Availability	1.00000000	0.02336231	0.08231387
Accessibility	0.02336231	1.00000000	-0.09637995
Utilisation	0.08231387	-0.09637995	1.00000000

Based on the results in Tables 5.17 and 5.18, one can observe that similar and consistent results were found. It can be said that negative and significant correlation was observed between food access and utilisation. In contrast, positive but not statistically significant correlation was observed between food access and availability.

5.6.2 PCC Selection

The preliminary analysis computed in Table 5.18 reduces the number of bivariate copula families and the result showed that the Clayton and Gumbel copulas are not applicable for measuring the dependence between accessibility and utilisation because their Kendall's tau shows negative dependence. Moreover, the Archimedean and AMH copulas are preferable for longitudinal data to incorporate determinant factors in the sense that the dataset is large. Hence, we removed the Gaussian and t copula from the lists of bivariate copulas. For those that satisfied the preliminary analysis, Algorithm I was employed and their corresponding modified Akaike Information Criterion (mAIC) were computed as presented in Table 5.19.

Table 5. 19: Summary of copula families using Algorithm I to the data and estimated modified Akaike Information Criterion (mAIC) to select the best fit copula families.

	Modified Akaike Information Criterion (mAIC)		
Bivariate Copula	(Y_1, Y_2)	(Y_2, Y_3)	$(Y_1 \vert Y_2, Y_3 \vert Y_2)$
Clayton	**14115.8**	NA	NA
Gumbel	14121.48	NA	NA
Frank	14117.95	13078.27	12667.74
AMH	14117.78	**13077.34**	12668.16
Independent	14119.07	13100.01	**12663.28**

where Y_1 = Availability, Y_2 = Accessibility and Y_3 = Utilisation

The D-vine structure of household food security in the first tree was (availability (Y_1), accessibility(Y_2)) and (accessibility (Y_2), utilisation (Y_3)). Similarly, in the second tree, the structure is ($Y_1|Y_2$ and $Y_3|Y_2$). The bivariate correlation between $Y_1|Y_2$ and $Y_3|Y_2$ was computed as negative. Hence, the Clayton and Gumbel copulas are not still applicable for measuring the dependence between them. From Table 5.19, one can deduce that the best fitted copula families on the first tree were Clayton and AMH for edge 1 and 2 respectively while the second tree has independent copula family since the mAIC is smaller than the other copula families. Hence, Clayton, AMH and Independent copula families were selected as best fit bivariate copula that serves for parameter estimation in the full maximum likelihood parameter estimation for the corresponding vine structure they fitted best.

5.6.3 Estimation of the Copula and Marginal Parameters

The best fitted bivariate copulas were selected using Algorithm I as presented in Table 5.19. Clayton bivariate copula was selected to measure the dependence of household food security statuses between availability and accessibility dimensions. Similarly, to measure between accessibility and utilisation, AMH was selected. Moreover, independent copula was for availability and utilisation given that accessibility household food security status has already appeared. The corresponding marginal distribution for the application data was marginal model via cumulative logit model.

The likelihood function was computed using these selected bivariate copula families and the marginal distribution. MLE approach was used to jointly compute the dependence parameters using the selected bivariate copula families and the marginal parameters using marginal model of the cumulative logit model. Derivation of the likelihood and log-likelihood function was computed and presented in Appendix B III.

We wrote an R code using “auglag” optimization R package to estimate the parameters of the copula and the marginal distribution functions. Detail of the R code is displayed in Appendix C III. The estimated values of the parameters and the corresponding standard errors both for the copula and marginal distributions were shown in Table 5.20 and 5.21 respectively.

Copula Parameter

Table 5.20 summarises the results of the estimated dependence parameters of the selected bivariate copula families. Clayton copula measures the dependence between availability and accessibility household food security status, AMH copula measures the accessibility and the utilisation and independent the availability and the utilisation given that accessibility happened and found that positive dependence were observed. Moreover, these dependences were observed as statistically significant. This leads to the conclusion that household food security status dependences to each other.

Table 5. 20: The estimates of dependence parameters using the selected pair copula for the application data of multivariate longitudinal household food security data.

Tree	Copula family	Estimated Parameter	Estimated SE	Bijection tau
I	Clayton	1.5	0.6351	0.4285714
	AMH	0.99999	0.3811	0.3333267
II	Independent			

Marginal Parameter

All statistically significant variables in each of the marginal model via the cumulative logit model were incorporated in the model for each dimension that hopes to be helpful as additional information with the existing knowledge in this area. Summary statistics of the final model is displayed in Table 5.21. In Table 5.21, the blank space indicated that particular variable was not statistically significant for that dimension in the marginal model and hence, that particular variable was not included only for that particular dimension during the estimation of the final joint model.

Among the variables incorporated in the model, the follow-up time point and total farmland size in hectare were identified as the potential statistically significant variables for the household food in-security status in all dimensions. In each dimension, household food security status varies with time. Therefore, the household food security status in each dimension in the study area is not stable over time. Small land size ploughed (<= 0.5 hectare) is more likely to lead households to be chronically, moderately and mildly food in-secured than those ploughed greater than 0.5 hectare.

Table 5. 21: Summary results of pair copula based marginal model via cumulative logit model parameters estimates for the household food security data.

Variable	Categories	Availability		Accessibility		Utilisation	
		Estimates	S.E	Estimates	S.E	Estimates	S.E
Intercept	1|2	**-4.10884**	0.244	**-2.76063**	.585	**-5.02558**	.626
	2|3	**-1.06692**	0.221	-0.60402	.583	-1.06226	.606
	3|4	**1.20519**	0.221	**1.56259**	.583	**1.33108**	.606
Time		**0.127028**	0.01	**0.043073**	.009	**0.101924**	.009
Household Head (HH)							
	Husband			-1.06369	.564	-1.095	.590
	Wife			-1.19779	.575	**-1.41323**	.598
	Son/daughter						
Marital Status (HH)							
	Never married/Cohabiting	-0.33469	0.16			**-0.42594**	.199
	Married	-0.30143	0.146			0.03492	.189
	Divorced/Widowed						
Study Site (Woredas)							
	Kutaber	**-0.98267**	0.131	0.24066	.124		
	Tehuledere	**-1.36491**	0.133	**-0.46358**	.123		
	Kalu						
Total farmland size in Hectare							
	<=0.5 Hectare	**0.39055**	0.092	**0.276432**	.087	**0.362342**	.091
	> 0.5 Hectare						
Types of Cereal Crops cultivated							
	One type	**0.355914**	0.129				
	two type	**0.452018**	0.116				
	Three and more types						
Time of cultivate within a year							
	Yearly	**-0.71082**	0.117				
	Biannual and more						

Table 5.21 Summary results of pair copula based marginal model via cumulative logit model parameters estimates for the household food security data … **continued**

Variable	Categories	Availability		Accessibility		Utilisation	
		Estimates	S.E	Estimates	S.E	Estimates	S.E
Types of livestock							
	One and less type			**0.29618**	.121		
	Two to Three types			0.118352	.104		
	Four or more types						
Presence of Pests							
	Yes					-0.17848	.092
	No						
Shortage of rainfall							
	Yes	**0.33973**	0.1			**0.444659**	.098
	No						
Crop Disease							
	Yes	**1.00354**	0.103				
	No						
Increase in market price							
	Yes	**0.47612**	0.102	**0.239897**	.097	-0.148	.099
	No						
Use of Pesticides							
	Yes	**0.838011**	0.113				
	No						
Agro-Ecology of Study Site							
	Hot (Kolla)			**-0.74647**	.179		
	Medium (Weinadega)			**-0.59454**	.125		
	Cold (Dega)						

Similarly, among the fixed covariates, study site was identified as a significant influencing variable for household food security status both in availability and accessibility dimensions. Households cultivated less than three types of cereal crops and cultivating once a year were more likely to be chronically, moderately and mildly food in-secured than those who cultivated more than two types and two or more times per year in availability dimension respectively. Moreover, households headed by women who have widowed or divorced marital status are more likely to be chronically, moderately and mildly food in-secured than headed by son/daughter and never married or cohabiting in utilisation dimensions.

On the other hand, among time varying covariates, the shortage of rainfall is identified as significant variables that lead households to be chronically, moderately and mildly food in-secured both in availability and utilisation dimensions in all aggregated time points. Areas used pesticides, market price increase and crop disease happened are more likely to lead households to be chronically, moderately and mildly food in-secured than areas not used pesticides and crop disease happened in availability dimension in all time points. Similarly, hot (Kolla) or medium (Weinadega) agro-ecology, less than two types of livestock and stability of market price are more likely to lead households to be chronically, moderately and mildly food in-secured than cold (Dega), more than three types of livestock and the absence of market price increase in accessibility dimension in all aggregated time points.

5.6.4 Effects of PCC Model on the Univariate Marginal Cumulative Logit Model

The univariate and PCC population-average cumulative logit model identified almost equal significant predictors for household food insecurity in the availability and accessibility dimensions. However, in the utilisation dimension, the univariate model identified more significant predictors for household food insecurity over the PCC model. For comparison purpose, the univariate marginal model outputs both for fixed and time varying covariates were displayed in Table 5.22. Similarly, the PCC population-average cumulative model outputs both for fixed and time-varying covariates were presented as well in Tables 5.21.

Table 5. 22: Summary results of the univariate marginal model via cumulative logit model parameters estimates for the household food security data.

Variable	Categories	Availability		Accessibility		Utilisation	
		Estimates	S.E	Estimates	S.E	Estimates	S.E
Intercept	1|2	**-3.923***	0.235	**-2.827***	.575	**-5.392***	.675
	2|3	**-1.054***	0.209	-0.929	.572	**-1.571***	.656
	3|4	**1.182***	0.210	1.006	.571	0.853	.656
Time		**-0.063***	0.010	0.073	.009	**0.039***	.009
Household Head (HH)							
	Husband			-1.095	.552	-1.228	.638
	Wife			**-1.207***	.564	**-1.429***	.645
	Son/daughter						
Marital Status (HH)							
	Never married/Cohabiting	**-0.412***	0.158			-0.**539***	.203
	Married	**-0.337 ***	0.144			-0.056	.193
	Divorced/Widowed						
Study Site (Woredas)							
	Kutaber	**-0.804***	0.139	**0.268***	.124		
	Kalu	**-1.070***	0.141	**0.442***	.123		
	Tehuledere						
Total farmland size in Hectare							
	<=0.5 Hectare	**-0.414***	0.091	**0.250***	.088	**-0.332**	.091
	> 0.5 Hectare						
Types of Cereal Crops cultivated							
	One type	**0.539***	0.123				
	two type	**0.670***	0.110				
	Three and more types						
Time of cultivate within a year							
	Yearly	-0.209	0.110				
	Biannual and more						

*significant at 5% level of significance

Table 5.22 Summary results of the univariate marginal model via cumulative logit model parameters estimates for the household food security data **…Continued**

Variable	Categories	Availability		Accessibility		Utilisation	
		Estimates	S.E	Estimates	S.E	Estimates	S.E
Types of livestock							
	One and less type			**0.294***	.121		
	Two to Three types			0.091	.104		
	Four or more types						
Presence of Pests							
	Yes					**-0.341***	.091
	No						
Shortage of rainfall							
	Yes	**0.323***	0.10			**0.362***	.092
	No						
Disease of cultivation							
	Yes	**1.060***	0.104				
	No						
Increase in market price							
	Yes	**0.651***	0.101	**0.274**	.097	**-0.363***	.099
	No						
Use of Pesticides							
	Yes	**0.898***	0.114				
	No						
Weathering condition of the cite							
	Hot (Kolla)			**-0.766***	.177		
	Medium (Weinadega)			**-0.634***	.125		
	Cold (Dega)						

*significant at 5% level of significance

Based on the finding presented in Tables 5.21 and 5.22, among the fixed covariates, the PCC population-average cumulative model identified a yearly once cultivation activity was the positive determinant for households to be severe to mildly food in-secured compared with counterparts while it was not significant in the univariate model in the availability dimension. Conversely, the univariate model identified the marital status of the household head and types of household head as a predictor for household food insecurity in availability and accessibility dimensions respectively while the PCC model has dropped them out.

The PCC model underestimated the almost all significant predictors of household food insecurity in the availability dimension. On the other hand, the PCC model overestimated in almost all significant predictors of household food insecurity in the accessibility dimension. Similarly, the PCC model in the utilisation dimension dropped out predictors like presences of pests and market price increase while these were significant determinants for household food insecurity in the univariate population-average cumulative model. The PCC model overestimated all of the significant predictors in the utilisation dimensions. The interpretation and discussions for the finding were intended on the findings of the PCC population-based cumulative logit model.

Chapter Six

6. Discussions

In this study, pair copula construction (PCC) approach was implemented for analysing multivariate, longitudinal and multivariate longitudinal ordinal data applied to household food insecurity collected from selected Woredas of South Wollo Zone. The practical implementation of PCC model for the three types of data was separately discussed in the subsequent sections.

6.1 Discussions for the Findings of the Pair Copula Based Multivariate Ordinal Model Application to Food Security Data

In this section, we applied a pair copula construction based cumulative logit regression model to jointly determine the dependence between the three food security dimensions and the respective determinants (Olaomi and Yimam, 2019). Prior to the application, we conducted selection of appropriate bivariate copula families that best fit to examine the dependence of food security dimensions and estimate their corresponding parameter using the pseudo data. For this purpose, the algorithm relevant for bivariate copula selection was developed. Among the candidate bivariate copula families, the Frank copula selected as best fitted to express the dependence between accessibility and utilisation, and availability | accessibility and utilisation | accessibility. Moreover, AMH copula was the best fit for measuring the dependence of availability and accessibility. Furthermore, the cumulative logit model was used as marginal distribution to estimate the marginal parameters. Finally, full MLE method was implemented for jointly estimate the dependence between the three dimensions and their respective determinants.

Overall, this model provided a very good description of the data and estimated all the dependence parameter and marginal parameters as needed. The nice feature of this model in this setting was that it allows for the estimation of the effect of the covariates both on the marginal parameters and on the dependence of the outcomes over the other multivariate ordinal models. This model depicted clearly the effects of the covariates on the dependence parameter of the three dimensions. The Kendall's tau for the pseudo data computed from the cumulative logit and Bijection tau computed from the copula parameter showed large difference. The Bijection tau was computed from the copula parameter, which was estimated incorporating the effect of covariates in the model. The Bijection tau more relied with the literature in the food security analysis. Furthermore, to assess the effects of PCC model on the cumulative logit model, the univariate cumulative logit estimates were also fitted for each dimension. The PCC model identified additional significant determinants of household food insecurity in all dimensions. On the other hand, the PCC model overestimated majority determinants of household food insecurity in availability and accessibility dimensions while underestimated in the utilisation dimension.

In PCC, the copula parameters captured the pair-wise non-normal relation between the food security dimensions. Food availability, accessibility and utilisation have pair-wise positive relationship. Moreover, the D-vine PCC determined the direction of the relationship as availability contributes to accessibility, accessibility contributes to utilisation and given that accessibility, availability contributes to utilisation. This finding is consistent with the framework developed by FAO (FAO, 2008). The finding implies that the households being food secured in availability dimension, the likelihood of food insecurity trap declines in food accessibility and then in utilisation dimensions. Likewise, the households that are food secured in accessibility dimension, the likelihood of food insecurity trap declines in food utilisation dimension.

Determinants of Availability: The findings of marginal parameters revealed that households with higher agro-ecology (study site), less ploughed land, shortage of rainfall, cultivating once a year, market price increases, hot agro-ecology and presence of disease on the cultivated land were more likely to be chronically to mildly food in-secured. In contrast, households headed by divorced or widowed marital status were less likely to be chronically to mildly food in-secured (Olaomi and Yimam, 2019).

The study site with lower agro-ecology has positive effects on food in-security compared with the study site with higher agro-ecology. This finding may relate with hot agro-ecology of the study site because one of the factors that positively affect food insecurity was hot agro-ecology. This finding is consistent with the result of meta-analysis conducted in Ethiopia by (Bashir and Schilizzi, 2012). Similarly, a household with small land ownership is more likely to be food in-secured and this finding consistent with the study conducted in Ethiopia by (Bashir and Schilizzi, 2012). Households' agricultural activities have also a positive effect on the households' food insecurity status. Households cultivated agricultural produced once on the yearly base were more likely to be food in-secured than those obtained twice or more per year.

Among climate change related factors, the availability of limited amount of annual rainfall has positive effect on the household food insecurity. A similar finding was observed in a research conducted in Ethiopia by (Abegaz, 2017). Moreover, recurrent disease that occurred on the cultivated land was positively affects the household food insecurity status. Market price increase is also another factor that positively influences the status of household food insecurity. This finding is similar with findings in the meta-analysis conducted by (Bashir and Schilizzi, 2012) in Ethiopia and a research conducted by (Ahmed et al., 2017) in Pakistan.

Moreover, among household characteristics household headed by never married or cohabiting marital status were less likely to be food in-secured compared to households headed by divorced or widowed.

Determinants of Accessibility: The findings of the marginal parameter of the pair copula based cumulative logit model revealed that households headed by never married or cohabiting marital status, small farmland size, shortage of rain fall, cultivating once a year, hot weathering condition and presence of disease on the cultivated land contribute to making the households to be chronically to mildly food in-secured.

Among demographic factors for household food insecurity status, households with never married or cohabiting were more likely food in-secured compared with divorced or widowed households. This finding is in contrast with finding in availability dimension of this study. On the other hand this finding is similar with the finding pointed out by Mensah et al. (2013) in Sekyere-Afram Plains District of Ghana. However, the finding by (Magaña-Lemus et al., 2016) in Mexico household headed by single, widowed or divorced women were more likely to be food in-secured compared with the married one.

Moreover, the amount of land ownership has effects on the household food security status. Households with small farmland size were more likely to be food in-secured than who have large farm size. The results of this study provides consistent result with the finding in some part of Ethiopia by (Astemir, 2015), (Shone et al., 2017), (Feyisa, 2018) and (Moroda et al., 2018) and in part of Ghana by Mensah et al. (2013). Similarly, agricultural activities have also a positive effect on the households' food insecurity status. Households those harvested once yearly were more likely to be food in-secured than those harvested twice or more per year.

On the other hand, among climate change and environmental factors, the availability of limited amount of annual rainfall has a positive effect on the household food insecurity and this result resonates with finding in the research conducted by (Abafita and Kim, 2014) and (Abegaz, 2017) in Ethiopia and in rural Zambia by (Wineman, 2016). Households living in hot agro-ecology were more exposed to chronic to moderate food in-secured compared with households living in cold agro-ecology. Moreover, recurrent disease occurred on the cultivated land was also positively affects the household food insecurity status.

Determinants of Utilisation: The marginal estimates of the pair copula based cumulative logit model of the utilisation dimensions revealed that household headed by sibling, higher agro-ecology (study site), small farmland size, shortage of rainfall, cultivating once a year, and presence of disease on the cultivated land were positively affect households to be chronically to mildly food in-secured.

Among the demographic variables statistically significant factors for food insecurity status is the household headed by siblings. Households headed by siblings were more likely to be food in-secured than headed by husbands. Households living in higher agro-ecology environment were more likely to be food insecurity than living in lower agro-ecology environment. This finding is in contrast with the finding in the availability dimension of this study. Moreover, households that cultivated large farmland size have the potential to be food secured than cultivated small farmland size. This is because households have the potential to harvested different food groups by their own other than purchase from the local market. This finding is in line with the finding carried out by (Moroda et al., 2018) in Ethiopia. Households cultivating once per year were more likely to be food in-secured than cultivated twice or more per year.

Among climate change and environmental variables, household obtained small amount of annual rainfall were more likely to be food in-secured. This finding is consistent with the finding carried out by (Moroda et al., 2018) in Ethiopia. Moreover, recurrent disease that occurred on the cultivated land was also positively affects the household food insecurity status. This finding is also harmony with the finds by (Abegaz, 2017).

Among the variables incorporated in the final model, four of them were obtained as common determinant factors for all of the three food insecurity dimensions. These are small farmland size, shortage of annual rainfall, cultivating once a year and presence of disease on the cultivated land.

6.2 Discussions for the Findings of the Pair Copula Based Longitudinal Ordinal Model Application to Food Security Data

In this section, we applied a pair copula construction based cumulative logit regression model to jointly determine the stability of household food security over time and the respective determinants (Olaomi and Yimam, 2019). Prior to the application, we conducted the selection of appropriate bivariate copula families that best fits to examine the dependence of food security statuses over time and estimate their corresponding parameters using the pseudo data. For this purpose, the algorithm relevant for bivariate copula selection was developed. Among the bivariate copula families, the Gumbel copula selected as best fitted to express the dependence between the first and second, the second and the third time points. Moreover, the Gumbel copula was the best fit for measuring the dependence of first and third phase food security status given that the second phase has already happened. Furthermore, the cumulative logit model was used as marginal distribution to estimate the marginal parameters. Finally, full MLE method was implemented for jointly estimate the dependence between the three consecutive time point household food security status and their respective determinants.

Overall, this model provided a very good description of the data and estimated all the dependence parameter and marginal parameter as needed. The PCC model estimated the dependence between food security status among the successive time points and the effects of the time varying covariates on the dependence of food security statuses using the estimated bivariate copula parameters and the effect of the time varying covariates on each food security status of the successive time points using the estimated parameter of the cumulative logit model. The nice feature of this model in this setting was that it allows estimation of the effect of the covariates both on the marginal parameters and on the dependence of the outcomes over the other longitudinal ordinal models. On the top of this, the model depicted also the recurrent covariates that affect the household food security status over time the other longitudinal ordinal models did not have. Furthermore, to assess the effects of PCC model on the cumulative logit model, the univariate cumulative logit estimates was also fitted for each time points. The PCC model identified additional significant determinants of household food insecurity in the first and second rounds. The univariate model identified more significant predictors than the PCC model in third rounds. Moreover, the PCC model underestimated majority of determinants of household food insecurity in the first round. In the second and third rounds in some of the predictors, the PCC model underestimated and in some of them overestimated.

The findings in the copula parameter showed that there were statistically significant differences between the pair-wise dependence in all successive time points. The copula parameters showed positive dependence between successive time points. In the longitudinal analysis, positive correlation is expected between successive time points of the individual response. This model is concordant with the expectation of longitudinal analysis. Again, in the longitudinal analysis, strong correlation resulted in stability over time. Likewise, large dependence copula parameter resulted in strong correlation and subsequently stability over time. Therefore, the current study indicates that significant but small dependence parameter leads to instability over time. Hence, the individual household food security status is not stable over time. This means that household food security status varied from time-to-time.

The findings in the marginal parameter showed that presence of crop disease, market price increase and medium (Weinadega) agro-ecology were significant and recurrent factors for households to be chronically to mildly food in-secured. The findings of this study indicate that availability of adequate amount of annual rainfall is crucial for household food security. This finding is consistent with the study conducted in Ethiopia by (Abafita and Kim, 2014) and (Mbolanyi et al., 2017) and in Uganda by (Mbolanyi et al., 2017). An increase of market price positively influences the status of household food insecurity. This finding resonates with the finding conducted in Pakistan by (Ahmed et al., 2017) and a meta-analysis in Ethiopia by (Bashir and Schilizzi, 2012). Moreover, the fluctuation of agro-ecology over time affects the status of household food security status. The finding is similar with finding conducted Ethiopia by (Mbolanyi et al., 2017).

One-time cultivation per year is the significant covariate that leads households to be chronically to mildly food in-secured in the second time points compared with those cultivated more than one cultivation season. For the time being, in this study, the availability of rainfall was not significant factor for household food security.

6.3 Discussions for the Findings of the Pair Copula Based Multivariate Longitudinal Ordinal Model Application to Food Security Data

In this study, the pair copula construction approach was extended for multivariate longitudinal ordinal data via the marginal model of the cumulative logit model. Marginal model is one of the statistical models commonly used in the univariate longitudinal ordinal data analysis using the cumulative logit version. Therefore, we proposed population-average based PCC model for multivariate longitudinal ordinal data. This model has the potential to accommodate jointly the dependence between multivariate ordinal outcomes, the covariate and follow-up time effects of the ordinal outcomes both on the dependence measures and the marginal probabilities. However, previous recent works on this area lacked to accommodate the aforementioned important information jointly in a single model (Abegaz et al., 2015, Laffont et al., 2014). Our model filed the population-average gap of the random effect models developed by (Laffont et al., 2014) and the computational challenge of population-average multivariate t-copula models developed by (Abegaz et al., 2015). Furthermore, the additional nice feature of this model is it allows estimation of the effect of the covariates and follow-up time points both on the marginal parameters and the dependency of the outcomes and allows estimation of the dependence between multivariate ordinal outcomes.

Our model was applied in household food security data. The dependence between food security status among the three dimensions and the effects of the covariates and time components on the dependence of food security status of the three dimensions was computed using the estimated bivariate copula parameters while the effect of the covariates and the follow-up time components on each food security dimension status was computed using the estimated parameter of the marginal model of the cumulative logit model.

Prior to the application, we conducted the selection of appropriate bivariate copula families that fit best to examine the dependence of food security statuses over time in each of the three dimensions and estimate their corresponding parameter using the pseudo data. For this purpose, the Algorithm relevant for bivariate copula selection was developed. Among the bivariate copula families, the Clayton copula selected as best to express the dependence between the availability and accessibility, the AMH copula for measuring the dependence of accessibility and utilisation food security status in all of the aggregated time points for the particular application data.

Moreover, the independent copula was the best fit for measuring the dependence of availability and utilisation food security status given that the accessibility dimension has already conducted.

Based on the selected bivariate copula families, we fitted the pair copula multivariate longitudinal cumulative logit model to our data. Moreover, we focused on estimation of the dependence between food security dimensions using the copula parameter, the stability over time using the significance of the time component and their predictor variables using the marginal distribution parameters of the marginal model. Ten covariates for availability, seven for accessibility and eight for utilisation were incorporated in this model, which are significant at 5 percent significance level in the marginal model. The full MLE method was employed to estimate both the copula and marginal parameters simultaneously.

In line with the PCC estimation, the univariate population based cumulative logit was fitted for each dimension to see the effect of the PCC model on the univariate one. Both the univariate and PCC population-average cumulative logit model identified almost equal significant predictors for household food insecurity in the availability and accessibility dimensions. However, in the utilisation dimension, the univariate model identified more significant predictors for household food insecurity over the PCC model. The PCC model underestimated almost all the significant predictors of household food insecurity in the availability dimension. The PCC model overestimated almost all of the significant predictors in the accessibility and utilisation dimensions.

The findings of the copula parameter showed that the copula parameters captured the pair-wise dependence between the food security dimensions. Food availability, accessibility and utilisation have pair-wise positive dependence. Moreover, the D-vine PCC determined the direction of the relationship as availability contributes to accessibility, accessibility contributes to utilisation and given that accessibility, availability contributes to utilisation. This finding is consistent with the framework developed by FAO (FAO, 2008). The finding implies that the households being food secured in availability, the likelihood of food insecurity trap declines in food accessibility and then in utilisation. Likewise, the households being food secured in accessibility, the likelihood of food insecurity trap declines in food utilisation. Hence, food security dimension specific intervention might reduce the likelihood of food insecurity at household level.

Determinants of availability: The marginal estimate of the population-average cumulative model in the pair copula-based model carried out so far produced different significant associated determinant factors for household food security status in the availability dimension. The finding of this study revealed that the household food security status changes over time. Likewise, population-average household food security status is not stable over time. Similarly, the finding showed that lower agro-ecology, shortage of rainfall, the presence of cultivation disease, increased market price, use of pesticides, cultivating smaller types of cereal crops and cultivating once per year positively affect the household food in-security (Olaomi and Yimam, 2019).

Households living in lower agro-ecology are more likely to be chronically to mildly food in-secured compared with living in higher agro-ecology. In contrast, the findings were observed in the research conducted by (Motbainor et al., 2016) in Ethiopia. Moreover, households who ploughed smaller farmland size (less or equal to half hectare) were more likely to be chronically to mildly food in-secured than ploughed greater than half hectare. (Bashir and Schilizzi, 2012) identified similar findings in their meta-analysis in Ethiopia. Harvesting different types of cereal crops also affects the status of household food security. Households harvesting fewer types of cereal crops (less than three types) were more likely chronically to mildly food in-secured than harvesting three or more types. Moreover, households living in a village that has one period of cultivation season were more suspected to be food in-secured than cultivating two or more times in one-year period.

Among the time varying covariates, during the three consecutive follow-up interviews, the presence of small amount of annual rainfall positively affects household food insecurity status. This finding resonates with that of the studies conducted by (Wineman, 2016) in rural Zambia and in Ethiopia by (Abafita and Kim, 2014) and (Abegaz, 2017). Similarly, the instability of market price positively affects the household food insecurity status and the finding is similar with finding in Pakistan carried out by (Ahmed et al., 2017). Moreover, households affected by cultivation disease and used pesticides were more likely to be food in-secured compared with those not affected by cultivation disease and not used pesticide. This is owing to either the disease destroying much number of products or the households may have invested much amount of money for pesticides. This automatically leads the households to be food in-secured.

Determinants of accessibility: The marginal estimate of the population-average cumulative model in the pair copula-based model carried out so far produced different significant associated determinant factors for household food security status in the accessibility dimension. The finding of this study revealed that the household food security status changes over time. Likewise, population-average household food security status is not stable over time. Similarly, the finding showed that lower agro-ecology, increased market price, herbing small number of livestock, hot agro-ecology and small farmland size positively affected the household food in-security (Olaomi and Yimam, 2019).

Like in the availability dimensions, households living in lower agro-ecology are more likely to be chronically to mildly food in-secured compared with those living in medium agro-ecology. This is may be owing to hot agro-ecology of the study site because cold agro-ecology is one of the factors that positively affect household food insecurity status. In contrast, the findings were observed in the research conducted by (Motbainor et al., 2016) in Ethiopia. Moreover, households which ploughed smaller farmland size (less or equal to half hectare) were more likely to be chronically to mildly food in-secured than ploughed greater than half hectare. (Bashir and Schilizzi, 2012) and (Shone et al., 2017) in Ethiopia and (Mensah et al., 2013) in Pakistan identified similar findings. Households herding fewer types of livestock (less than two types) were more likely to be chronically to mildly food in-secured than herding two or more types. (Motbainor et al., 2016) and (Habyarimana, 2015) also discovered similar findings in Ethiopia and in Rwanda respectively.

Among the time varying covariates, during the three consecutive follow-up interviews, the presence of small amount of annual rainfall positively affects household food insecurity status. Similar findings were observed in Ethiopia conducted by (Abafita and Kim, 2014), (Abegaz, 2017) and (Agidew and Singh, 2018) and also in rural Zambia by (Wineman, 2016). Similarly, the instability of market price positively affects the household food insecurity status and the finding is similar with finding of a study conducted in Pakistan carried out by (Ahmed et al., 2017).

Determinants of utilisation: Like in availability and accessibility, the marginal estimate of the population-average cumulative model in the pair copula-based model carried out so far produced different significant associated determinant factors for household food security status in the utilisation dimension. The findings of this study revealed that the household food security status changes over time. Likewise, population-average household food security status is not stable over time. Similarly, the results showed that households headed by a woman, divorced/widowed marital status of the household head, shortage of rainfall and small farmland size positively affected the household food in-security.

The findings suggested that households headed by women positively affect the utilisation food insecurity status of households. Moreover, divorced/widowed households were more likely food in-secured compared to single or cohabitation household. On the other hand, households which ploughed smaller farmland size (less or equal to half hectare) were more likely to be chronically to mildly food in-secured than those that ploughed greater than half hectare. The finding is similar to (Moroda et al., 2018) study conducted in Ethiopia is similar with the current study.

Among the time varying covariates, during the three consecutive follow-up interviews, the presence of small amount of annual rainfall positively affects household food insecurity status. Similar findings were observed in Ethiopia conducted by (Abafita and Kim, 2014) and (Abegaz, 2017) and also in rural Zambia by (Wineman, 2016).

This study revealed the determinants for all of the three dimensions and tried to compare with the work of the others. However, owing to lack of available literature in the availability and utilisation dimensions for some of the determinants the current study limited to compare with other works.

6.4 Potential Policy Implementing Strategies

Household food security dimensions are correlated to each other. This implies that for the households being food secured in availability, the likelihood of food insecurity trap declines in food accessibility and then in utilisation. Likewise, the households being food secured in accessibility, the likelihood of food insecurity trap declines in food utilisation. This type of modelling assists the food aid agents, planners or policy makers that in which dimension a household is highly affected and which dimension let household’s food in-secured. This implies food security dimension specific intervention might reduce the likelihood of food insecurity at household level.

Hot agro-ecology areas were highly vulnerable to food insecurity. This implies that agro-ecology or area specific intervention could alleviate the risk of food in-secured in the rural households. This suggests that adaption strategies like harvest in hot agro-ecology. In response to this strategy, farmers can use seeds that are resistant to short rain fall season and conserves water in a hot area to increase the productivity.

Small cultivable land size increases the likelihood of households being food in-secured. This implies that households with large land ownership could produce more food or may generate income from it to purchase food for consumption. Hence, alternative income generating mechanism should be set for rural households to reduce the pressure of cultivable land in addition to encourage maximum yield from a given holdings through investing in land improvements and soil conservation. Alternatively, to increase cultivable land size, a strong policy intervention may be needed to relocate the population where settlement is densely populated to the sites where it is not. This enables, at least some group, to share from land holdings and any entitlements to resources which can lead to ensuring food security.

The number of cultivation season increases on the yearly base, the likelihood of households being food in-secured trap declines. This implies that households cultivating two or three times per year could produce more food or getting more income to purchase food for consumption than cultivation once per year. Hence, ways should be sought through promoting irrigation activity in order to increase cultivation season.

Households which experienced lower rainfall level were more likely to remain food in-secured. The majority of the Ethiopian rural households are rain-dependent for their agricultural production, resulting in persistent food insecurity. Hence, careful promotion of investment in infrastructure to support irrigation and water resources development is one aspect worth considering. On the other hand, climatic adaption strategies should be insight like selecting appropriate crop varieties that can be planted in low amount of rainfall.

Crop disease positively affects the household food insecurity status. Hence, the development agents (DAs) working in the area should provide immediate response either by providing medicines or consulting the community for alleviating the problem in the early stage. The government should also advance the agricultural strategies in a way that provides immediate responses for this and other related agricultural problems.

The price of food increases, purchasing power goes down, dietary quality and total energy intake are reduced, the likelihood of household being food in-secured increased. Hence, strong market price policy to make stable the market price as well as increase the supply of food for consumption is worth considering.

More livestock was kept by households that were more food secured. As the households' livestock possessions were increased, their food security status would inevitably also respond positive. More importantly, livestock possession enables the households to be food secure either through the income earned or by direct consumption. This implies that the availability of greater number of livestock permit households enhances their economic wellbeing in general and their food entitlement in particular. Hence, careful promotion of investment in livestock project supported by scientific methods is one aspect worth considering.

Chapter Seven

7. Conclusions, Recommendations and Future Works

7.1 Conclusions

The pair copula based multivariate ordinal model with the cumulative logit version successfully captured the non-normal relationship between ordinal outcomes and their respective determinants simultaneously. Allowing the estimation of the effect of the covariates both on the marginal parameters and on the non-normal correlation of the ordinal outcomes strengthens the estimation performance of this model over previous multivariate ordinal models. The copula parameters in the food security data revealed that pair-wise positive dependence was observed between food availability, accessibility and utilisation dimensions. The marginal parameters of this model depicted that small cultivable land, shortage of rainfall, cultivating once a year and presence crop disease were positively influences household food insecurity in all the three dimensions. Moreover, lower agro-ecology and market price increase positively affects household food insecurity in availability dimension. Similarly, hot agro-ecology positively affects household food security in accessibility. Moreover, lower agro-ecology positively affects household food security in utilisation.

A pair copula based longitudinal ordinal model with cumulative logit version jointly estimated the stability of ordinal outcome over time and the respective determinants. The result of the copula parameter in this model pointed out the food security status at household level is not stable over time. Estimating determinants for each longitudinal ordinal outcome broadens the inclusion of recurrent determinants over the longitudinal periods does the previous longitudinal models lacked. The marginal parameters of this model revealed that the presence of crop disease, market price increase, and medium (Weinadega) agro-ecology were significant and recurrent factors for households' food insecurity over the three time periods.

Population-average based pair copula multivariate longitudinal ordinal model with cumulative logit version jointly estimated the dependence between multivariate ordinal outcomes, and the covariate and follow-up time effects of the ordinal outcomes both on the dependence measures and the marginal probabilities using the full MLE method. This model successfully reduced the population-average lack of the random effects model and the computational challenge of the multivariate copula models of the multivariate longitudinal ordinal data analysis. This study provides a good measure of dependence between food security dimensions using the copula parameter and also the stability over time and the determinants of household food using the marginal model parameters for all dimensions simultaneously. The findings of the copula parameter showed that positive and statistically significant dependence were observed between availability and accessibility and accessibility and utilisation. The marginal model of the cumulative logit model was used to measure the parameters of the marginal distributions. The findings of the model reveal that household food security was unstable over time for each dimension. Small land size and shortage of rainfall were the common predictors of household food insecurity in all dimensions. Moreover, lower agro-ecology and instability of market price were the common predictors of household food insecurity in availability and accessibility dimensions.

7.2 Recommendations

Food security dimensions depended to each other. The rap of one dimension affects the other dimensions. Therefore, it is critically important to consider the common factors to provide immediate intervention for severely food in-secured households. Moreover, great attention also required to lookup which dimension is leading households to food in-secured.

Households' food security status either in the individual food security dimensions or in the composite food security is not stable over time. So great attention is required for granting households to be food secured taking valuable intervention for the identified recurrent determinants as well as other climate change and environment factors. Moreover, climatic adaption strategies should be insight like selecting appropriate crop varieties that can be resistant to short rain fall season and conserves water in a hot area to increase the productivity.

Likewise, systematic investment in infrastructure to support irrigation and water resources development to increase cultivation season and alternative income generating mechanism to reduce the pressure of cultivable land from a given holdings through investing in land improvements and soil conservation are critical viewpoints to reduce the likelihood of food insecurity at household level.

The government should design strong market price policy to make stable the market price as well as increase the supply of food for consumption; careful promotion of investment in livestock project supported by scientific methods; and advance the agricultural strategies in a way that provides immediate responses for crop disease.

The pair copula based multivariate and longitudinal ordinal model provided easily interpretable and understandable outputs. Therefore, we suggest the model for any multivariate and longitudinal discrete data analysis.

The population-average based pair copula multivariate longitudinal ordinal model addressed all of the food security dimensions simultaneously and the model found computationally effective for not large set of data. Therefore, we suggest this model to apply for other application areas for not extremely large number of outcomes and covariates.

7.3 Limitations and Weaknesses of the Study

In this thesis, the applied pair copula model for multivariate longitudinal ordinal data used a three-stage sampling procedure to get the application data. In three-stage sampling, each random selection may introduce a random effect. It is due to the contributions of the different stages to the variance of an estimator. However, the current model didn't take into account this during parameter estimation.

The second limitation is, in the notion of the concept of food security we used three round data collection to address the stability of the other three dimensions over time. Three rounds of data may not be providing a realistic estimator to oversee the entire stability. It would have been great if measurements were obtained for three to five seasons. However, we believe that some kind of longitudinal data is better than cross-sectional data to study the household food security situation.

The study tried to address all the four food security dimensions. Due to limited data at the national level to address all the dimensions, the study forced to conduct primary data from selected Woredas of one region of Amhara, Ethiopia. Hence the study did not represent the food security situation of the Amhara region.

7.4 Future Works

The pair copula-based regression models have applied throughout this thesis allowed specifying the effect of covariates of the marginal distributions on the dependence and marginal structure. We proposed a population-average based pair copula construction models for multivariate longitudinal ordinal outcomes using marginal model of the cumulative logit marginal distribution. Although we feel that our contribution is a major step forward in the modeling of multivariate longitudinal ordinal outcomes via pair copula construction, we discuss three important open questions in more detail.

The first open question is in the multivariate longitudinal ordinal outcomes, we have M ordinal outcomes repeatedly measured T times. During our model development we re-ordered the observations of the multivariate series into the univariate outcomes of dimensions $N = T{*}M$ given by $Y = (Y_1, Y_2, \dots, Y_M)$, where $Y_1 = (y_{11}, y_{21}, \dots, y_{N1})'$, $Y_2 = (y_{12}, y_{22}, \dots, y_{N2})'and\ Y_M = (y_{1m}, y_{2m}, ., y_{Nm})'$. As a result, the joint probability mass function $Pr(Y_1 = y_1, Y_2 = y_2, \dots, Y_m = y_m)$ is decomposed as follows:

$$\Pr(Y_1, Y_2, \dots, Y_m) = \Pr(Y_1 = y_1 | Y_2 = y_2, \dots, Y_m = y_m) \times \Pr(Y_2 = y_2 | Y_3 = y_3, \dots, Y_m = y_m) \times \dots \times \Pr(Y_m = y_m).$$

The parameters of the joint distribution of the entire dependence were estimated by a D-vine copula of dimension T $*$ M. This model reduced the multivariate longitudinal dimensions into a univariate longitudinal series based on the time point for each multivariate ordinal outcome. This model loses the dependence between successive time points and the effect of covariates on the dependence measure of the successive time point outcomes. This is not the issue of population-average longitudinal data analysis but for other applications to consider the individual change over time one can extend this model for multivariate longitudinal data (T repeated observations of M dimensional vectors for the sample of n subjects) by using a different D-vine copula approach as follows.

$\boldsymbol{Y} = (\boldsymbol{Y^1}, \boldsymbol{Y^2}, \dots, \boldsymbol{Y^M})$, where $\boldsymbol{Y^1} = (Y_1, Y_2, \dots, Y_T), \boldsymbol{Y^2} = (Y_1, Y_2, \dots, Y_M), \dots, \boldsymbol{Y^M} = (Y_1, Y_2, \dots, Y_T)$, $Y_1 = (y_{11}, y_{21}, \dots, y_{t1})'$, $Y_2 = (y_{12}, y_{22}, \dots, y_{t2})'and\ Y_M = (y_{1m}, y_{2m}, ., y_{tm})'$.

Hence, the joint probability mass function $Pr(Y_1 = y_1, Y_2 = y_2, \dots, Y_m = y_m)$ is decomposed as follows.

$$\boldsymbol{Pr(Y^1 = y^1, Y^2 = y^2, \dots, Y^M = y^m) = Pr(Y^1 = y^1 | \, Y^2 = y^2, \dots, Y^M = y^m) \times Pr(Y^2 = y^2 | \, Y^M = y^m) \times \dots \times Pr(Y^M = y^m)}.$$

Taking the above expression into account one can apply the usual pair copula construction. This extension can consider three different levels of analysis. At first, a pair copula describes the relations of the responses observed at a specific time. Second each longitudinal series, corresponding to a given response over time, is modeled separately using a pair copula decomposition to relate the distributions of the variables describing the observation given in different times. Finally, the marginal distribution relates the associated factors for each responses and longitudinal time components.

The second open question is extending our model in in a Bayesian framework. Since the Bayesian approaches have many advantages in modeling multivariate as well as multivariate longitudinal outcomes, we believe that there may be significant advantages to estimating our model in a Bayesian framework. Furthermore, the modular nature of the MCMC in the Gibbs sampler may facilitate the development of more advanced multivariate longitudinal models. However, the Bayesian approach requires a good proposal of prior information or distribution for the marginal distribution during the construction of PCC. This implies that the selection of prior distribution in the ordinal outcomes setting requires intensive work. As a result, this thesis is concerned itself on the implementation of the PCC through the Frequentist paradigm and the likelihood for our PCC is fast to compute. However, in the presence of proper prior distribution, we believe that joint estimation of marginal and copula parameters for multivariate longitudinal ordinal model could be easier to develop in a Bayesian context.

The third open question of this thesis is considering non-ignorable missing values in the analysis of multivariate longitudinal outcomes via PCC. In the univariate longitudinal context, Cui et al., (2016) implemented the Peter and Clark (PC) algorithm for both the discrete and continuous data assumed to be drawn from a Gaussian copula models. Furthermore, Cui et al., (2019) extended the Gaussian copula models to Copula PC algorithm for incomplete data for mixed data with missing values. Likewise, Gomes et al. (2019) further extended Gaussian copula for non-

Gaussian responses that are missing not at random using copula selection models. This implies that in the univariate case, intensive works have been conducted in the copula context for discrete data. However, in the multivariate longitudinal discrete context, since PCC models in multivariate longitudinal data analysis are in the growing stage, it needs further research to handle missing-ness. In our PCC model we did not consider handling missing data since we were considering compensation of none-response rate and the missing-ness in our data were less than the none-response rate considered during sample size determination. However, we believe that considering PCC models in the multivariate longitudinal ordinal outcomes that can treat non-ignorable missing data could provide valid estimates both in the marginal and copula estimates of the final model extending one of the above methods that was implemented in the univariate case.

References

AAS, K., CZADO, C., FRIGESSI, A. & BAKKEN, H. (2009) Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44, 182-198.

ABAFITA, J. & KIM, K.-R. (2014) Determinants of household food security in rural Ethiopia: An empirical analysis. *Journal of Rural Development/Nongchon-Gyeongje*, 37, 129.

ABDU, J., KAHSSAY, M. & GEBREMEDHIN, M. (2018) Household Food Insecurity, Underweight Status, and Associated Characteristics among Women of Reproductive Age Group in Assayita District, Afar Regional State, Ethiopia. *Journal of environmental and public health*, 2018.

ABEGAZ, F., NOORAEE, N., VAN DEN HEUVEL, E. & WIT, E. (2015) Logistic Regression for Multivariate Longitudinal Ordinal Data. *University of Groningen*.

ABEGAZ, K. H. (2017) Determinants of food security: evidence from Ethiopian Rural Household Survey (ERHS) using pooled cross-sectional study. *Agriculture & Food Security*, 6, 70.

AGIDEW, A.-M. A. & SINGH, K. (2018) Determinants of food insecurity in the rural farm households in South Wollo Zone of Ethiopia: the case of the Teleyayen sub-watershed. *Agricultural and Food Economics*, 6, 10.

AGRESTI, A. (2010) *Analysis of ordinal categorical data*, John Wiley & Sons.

AHMED, U. I., YING, L., BASHIR, M. K., ABID, M. & ZULFIQAR, F. (2017) Status and determinants of small farming households' food security and role of market access in enhancing food security in rural Pakistan. *PloS one*, 12, e0185466.

ALI, M. M., MIKHAIL, N. & HAQ, M. S. (1978) A class of bivariate distributions including the bivariate logistic. *Journal of multivariate analysis*, 8, 405-412.

ASAR, Ö. & İLK, Ö. (2014) Flexible multivariate marginal models for analyzing multivariate longitudinal data, with applications in R. *Computer methods and programs in biomedicine*, 115, 135-146.

ASMAMAW, T., BUDUSA, M. & TESHAGER, M. (2015) Analysis of Vulnerability to Food Insecurity in the Case of Sayint District, Ethiopia. *Asian Journal of rural development*, 5, 1-11.

ASPELUND, T. (2002) *Non-linear association-marginal models for multivariate categorical data with application to ordinal receiver operating characteristic analysis*, UMI Dissertation Services.

ASSEFA, B. T. (2015) *Food price changes and food security in developing countries: a multidimensional analysis*, Sl: sn.

ASTEMIR, A. (2015) Determinants of food security in rural farm households in Ethiopia. *The Hague, Netherlands*.

BALLARD, T., COATES, J., SWINDALE, A. & DEITCHLER, M. (2011) Household hunger scale: indicator definition and measurement guide. *Washington, DC: Food and Nutrition Technical Assistance II Project, AED*.

BANDYOPADHYAY, S., GANGULI, B. & CHATTERJEE, A. (2011) A review of multivariate longitudinal data analysis. *Statistical methods in medical research*, 20, 299-330.

BASHIR, M. K. & SCHILIZZI, S. (2012) Measuring food security: Definitional sensitivity and implications. *Contributed Paper Prepared for Presentation at the 56th AARES Annual Conference, Fremantle, Western Australia, Paper*.

BEDFORD, T. & COOKE, R. M. (2001) Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32, 245-268.

BEDFORD, T. & COOKE, R. M. (2002) Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 1031-1068.

BHATNAGAR, S., ATHERTON, J. & BENEDETTI, A. (2015) Comparing alternating logistic regressions to other approaches to modelling correlated binary data. *Journal of Statistical Computation and Simulation*, 85, 2059-2071.

BIESALSKI, H. K., DREWNOWSKI, A., DWYER, J. T., STRAIN, J., WEBER, P. & EGGERSDORFER, M. (2017) *Sustainable nutrition in a changing world*, Springer.

BILINSKY, P. & SWINDALE, A. (2007) *Months of adequate household food provisioning (MAHFP) for measurement of household food access: indicator guide*, Food and Nutritional Technical Assistance Project, Academy for Educational ….

BIRHANE, T., SHIFERAW, S., HAGOS, S. & MOHINDRA, K. S. (2014) Urban food insecurity in the context of high food prices: a community based cross sectional study in Addis Ababa, Ethiopia. *BMC public health*, 14, 680.

CAGNONE, S., MOUSTAKI, I. & VASDEKIS, V. (2009) Latent variable models for multivariate longitudinal ordinal responses. *British journal of mathematical and statistical psychology,* 62, 401-415.

CAPALDO, J., KARFAKIS, P., KNOWLES, M. & SMULDERS, M. (2010) A Model of Vulnerability to Food Insecurity.

CARLETTO, C., ZEZZA, A. & BANERJEE, R. (2013) Towards better measurement of household food security: Harmonizing indicators and the role of household surveys. *Global food security,* 2, 30-40.

CASTRO, A. P. (2000) Food Security and Resource Access: A Final Report on the Community Assessments in South Wello and Oromiya Zones of Amhara Region, Ethiopia.

CHAGANTY, N. R. & NAIK, D. N. (2002) Analysis of multivariate longitudinal data using quasi-least squares. *Journal of Statistical Planning and Inference,* 103, 421-436.

CHOI, J. (2012) Prediction in the joint modeling of mixed types of multivariate longitudinal outcomes and a time-to-event outcome. University of Pittsburgh.

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

COATES, J. (2013) Build it back better: Deconstructing food security for improved measurement and action. *Global Food Security,* 2, 188-194.

COCHRAN, W. G. (2007) *Sampling techniques*, John Wiley & Sons.

COHEN, M. J. & GARRETT, J. L. (2010) The food price crisis and urban food (in) security. *Environment and Urbanization,* 22, 467-482.

CUI, R., GROOT, P. & HESKES, T. (2016) Copula PC algorithm for causal discovery from mixed data. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases.* Springer.

CUI, R., GROOT, P. & HESKES, T. (2019) Learning causal structure from mixed data with missing values using Gaussian copula models. *Statistics and Computing,* 29, 311-333.

CZADO, C. (2010) Pair-copula constructions of multivariate copulas. *Copula theory and its applications.* Springer.

CZADO, C., JESKE, S. & HOFMANN, M. (2013) Selection strategies for regular vine copulae. *Journal de la Société Française de Statistique*, 154, 174-191.

DARDANONI, V. & FORCINA, A. (2008) Multivariate ordered regression. mimeo.

DE BRUIN, E. & GRESSE, A. (2018) Dietary diversity amongst adults who buy at shopping malls in the Nelson Mandela Bay area. *Journal of Consumer Sciences*.

DISSMANN, J. (2010) Statistical inference for regular vines and application. *Diploma the-sis, Technische Universitat Miinchen*.

DISSMANN, J., BRECHMANN, E. C., CZADO, C. & KUROWICKA, D. (2013) Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis,* 59, 52-69.

ENDALE, W., MENGESHA, Z. B., ATINAFU, A. & ADANE, A. A. (2014) Food Insecurity in Farta District, Northwest Ethiopia: a community based cross–sectional study. *BMC research notes,* 7, 130.

ENDALEW, B., MUCHE, M. & TADESSE, S. (2015) Assessment of food security situation in Ethiopia. *World Journal of Dairy & Food Sciences,* 10, 37-43.

ETANA, D. & TOLOSSA, D. (2017) Unemployment and food insecurity in urban Ethiopia. *African Development Review,* 29, 56-68.

FABER, M., SCHWABE, C. & DRIMIE, S. (2009) Dietary diversity in relation to other household food security indicators. *International Journal of Food Safety, Nutrition and Public Health,* 2, 1-15.

FAN, C.-L. (2009) *Contributions to the theory of copula*, WASHINGTON UNIVERSITY IN ST. LOUIS.

FANG, H.-B., FANG, K.-T. & KOTZ, S. (2002) The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis,* 82, 1-16.

FAO, A. (2008) Introduction to the Basic Concepts of Food Security. *Food Security Information for Action, Rome*.

FAO. (2014) *State of Food Insecurity in the World 2013: The Multiple Dimensions of Food Security*, FAO.

FEYISA, M. N. (2018) Determinants of food insecurity among rural households of South Western Ethiopia. *Journal of Development and Agricultural Economics,* 10, 404-412.

FISHER, N. I. (1997) Copulas. *Encyclopedia of statistical sciences*.

FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. & MOLENBERGHS, G. (2009) Longitudinal Data Analysis. *International Statistical Review,* 77, 147-165.

FITZMAURICE, G. M. & LAIRD, N. M. (1993) A likelihood-based method for analysing longitudinal binary responses. *Biometrika,* 80, 141-151.

FRAHM, G., JUNKER, M. & SZIMAYER, A. (2003) Elliptical copulas: applicability and limitations. *Statistics & Probability Letters,* 63, 275-286.

GANGE, S. J. (1994) *Multivariate ordinal responses in clinical trials*, University of Wisconsin--Madison.

GENEST, C. & NEŠLEHOVÁ, J. (2007) A primer on copulas for count data. *Astin Bulletin,* 37, 475-515.

GODFRAY, H. C. J., BEDDINGTON, J. R., CRUTE, I. R., HADDAD, L., LAWRENCE, D., MUIR, J. F., PRETTY, J., ROBINSON, S., THOMAS, S. M. & TOULMIN, C. (2010) Food security: the challenge of feeding 9 billion people. *science,* 327, 812-818.

GOMES, M., RADICE, R., CAMARENA BRENES, J. & MARRA, G. (2019) Copula selection models for nonâ€• Gaussian outcomes that are missing not at random. *Statistics in medicine,* 38, 480-496.

GRAY, S. M. & BROOKMEYER, R. (2000) Multidimensional longitudinal data: estimating a treatment effect from continuous, discrete, or time-to-event response variables. *Journal of the American Statistical Association,* 95, 396-406.

GUEORGUIEVA, R. (2001) A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling,* 1, 177-193.

GUMBEL, E. J. (1960) Bivariate exponential distributions. *Journal of the American Statistical Association,* 55, 698-707.

HABYARIMANA, J. B. (2015) Determinants of household food insecurity in developing countries evidences from a probit model for the case of rural households in Rwanda. *Sustainable Agriculture Research,* 4.

HAFF, I. H. (2012) Comparison of estimators for pair-copula constructions. *Journal of Multivariate Analysis,* 110, 91-105.

HARVILLE, D. A. & MEE, R. W. (1984) A mixed-model procedure for analyzing ordered categorical data. *Biometrics,* 393-408.

HEAGERTY, P. J. & ZEGER, S. L. (1996) Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association,* 91, 1024-1036.

HIRK, R., HORNIK, K. & VANA, L. (2018) Multivariate ordinal regression models: an analysis of corporate credit ratings. *Statistical Methods & Applications*, 28, 507-539.

HUANG, G. H., BANDEENÂ€• ROCHE, K. & RUBIN, G. S. (2002) Building marginal models for multiple ordinal measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51, 37-57.

HUNNES, D. E. (2013) Nutrition and Food Security in a Changing Climate: Methods for predicting household coping strategy use and Food Security in the Ethiopian Context. UCLA.

HUYNH, V.-N., KREINOVICH, V. & SRIBOONCHITTA, S. (2014) *Modeling dependence in econometrics*, Springer.

JANSEN, J. (1990) On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, 75-84.

JIANG, Z. (2012) Joint Modeling of Multivariate Ordinal Longitudinal Outcome. University of Pittsburgh.

JOE, H. (1996) Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, 120-141.

JOE, H. (1997) *Multivariate models and multivariate dependence concepts*, CRC Press.

KELLY, J. L. & PEMBERTON, C. (2016) An assessment of the household food security status and local foods grown in rural Bahamas. *Farm and Business-The Journal of the Caribbean Agro-Economic Society*, 8, 82.

KENNE PAGUI, E. C. & CANALE, A. (2016) Pairwise likelihood inference for multivariate ordinal responses with applications to customer satisfaction. *Applied Stochastic Models in Business and Industry*, 32, 273-282.

KISI, M. A., TAMIRU, D., TESHOME, M. S., TAMIRU, M. & FEYISSA, G. T. (2018) Household food insecurity and coping strategies among pensioners in Jimma Town, South West Ethiopia. *BMC public health*, 18, 1373.

KOPER, N. & MANSEAU, M. (2009) Generalized estimating equations and generalized linear mixed-effects models for modelling resource selection. *Journal of Applied Ecology*, 590-599.

KURADA, R. R. (2011) *Modeling and analysis of repeated ordinal data using copula based likelihoods and estimating equation methods*, Old Dominion University.

LAFFONT, C. M., VANDEMEULEBROECKE, M. & CONCORDET, D. (2014) Multivariate analysis of longitudinal ordinal data with mixed effects models, with application to clinical outcomes in osteoarthritis. *Journal of the American Statistical Association,* 109, 955-966.

LEE, K., DANIELS, M. J. & JOO, Y. (2013) Flexible marginalized models for bivariate longitudinal ordinal data. *Biostatistics,* 14, 462-476.

LENNON, H. (2016) Gaussian copula modelling for integer-valued time series. University of Manchester.

LIANG, K.-Y. & ZEGER, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, 13-22.

LIPSITZ, S. R., KIM, K. & ZHAO, L. (1994) Analysis of repeated categorical data using generalized estimating equations. *Statistics in medicine,* 13, 1149-1163.

LIU, J. (2007) Multivariate ordinal data analysis with pairwise likelihood and its extension to SEM. University of California Los Angeles.

LIU, L. C. (2008) A model for incomplete longitudinal multivariate ordinal data. *Statistics in medicine,* 27, 6299-6309.

LIU, L. C. & HEDEKER, D. (2006) A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics,* 62, 261-268.

MAGAÑA-LEMUS, D., ISHDORJ, A., ROSSON, C. P. & LARA-ÁLVAREZ, J. (2016) Determinants of household food insecurity in Mexico. *Agricultural and Food Economics,* 4, 10.

MBOLANYI, B., EGERU, A. & MFITUMUKIZA, D. (2017) Determinants of household food security in a rangeland area of Uganda. *African Journal of Rural Development,* 2, 213-223.

MENSAH, O., JAMES, A. & TUFFOUR, T. (2013) Determinants of household food security in the Sekyere-Afram plains district of Ghana. *Global Advanced Research Journal of Agricultural Science,* 2, 34-40.

MÉTHOT, J. & BENNETT, E. M. (2018) Reconsidering non-traditional export agriculture and household food security: A case study in rural Guatemala. *PloS one,* 13, e0198113.

MOLENBERGHS, G. & VERBEKE, G. (2006) *Models for discrete longitudinal data*, Springer Science & Business Media.

MORODA, G. T., TOLOSSA, D. & SEMIE, N. (2018) Food insecurity of rural households in Boset district of Ethiopia: a suite of indicators analysis. *Agriculture & Food Security,* 7, 65.

MOTBAINOR, A., WORKU, A. & KUMIE, A. (2016) Level and determinants of food insecurity in East and West Gojjam zones of Amhara Region, Ethiopia: a community based comparative cross-sectional study. *BMC public health,* 16, 503.

NAPIER, C., OLDEWAGE-THERON, W. & MAKHAYE, B. (2018) Predictors of food insecurity and coping strategies of women asylum seekers and refugees in Durban, South Africa. *Agriculture & Food Security,* 7, 67.

NAPOLI, M., DE MURO, P. & MAZZIOTTA, M. (2011) Towards a food insecurity Multidimensional Index (FIMI). *Master in Human Development and Food Security*.

NEGATU, W. (2004) Reasons for food insecurity of farm households in South Wollo, Ethiopia: explanations at grassroots.

NELSEN, R. B. (2007) *An introduction to copulas,* Springer Science & Business Media.

NGEMA, P., SIBANDA, M. & MUSEMWA, L. (2018) Household Food Security Status and Its Determinants in Maphumulo Local Municipality, South Africa. *Sustainability,* 10, 3307.

NICKLAS, S. (2013) Pair Constructions for High-Dimensional Dependence Models in Discrete and Continuous Time. Universität zu Köln.

NIGUSSIE, Z. & ALEMAYEHU, G. (2013) Levels of household food insecurity in rural areas of Guraghe zone, Southern Ethiopia. *Wudpecker J. Agric. Res,* 2, 8-14.

NIKOLOULOPOULOS, A. K. (2017) Weighted scores method for longitudinal ordinal data. *arXiv preprint arXiv:1510.07376*.

NOORAEE, N., ABEGAZ, F., ORMEL, J., WIT, E. & VAN DEN HEUVEL, E. R. (2016) An approximate marginal logistic distribution for the analysis of longitudinal ordinal data. *Biometrics,* 72, 253-261.

NOORAEEA, N. (2015) *Statistical methods for marginal inference from multivariate ordinal data*, Doctoral Thesis. 205p, University of Groningen.

OLAOMI, J. & YIMAM, J. A. (2019) Modeling the Stability and Determinant Factors of Household Food Insecurity: A Pair Copula Construction Approach. *Joint Conference of the Sub-Saharan African Network (SUSAN) of the International Biometrics Society (IBS)* *and DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics(SSACAB)*.

PANAGIOTELIS, A., CZADO, C. & JOE, H. (2012) Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107, 1063-1072.

PANAGIOTELIS, A., CZADO, C., JOE, H. & STÖBER, J. (2015) Model selection for discrete regular vine copulas.

PERIN, J. (2009) Improved generalized estimating equations for incomplete longitudinal binary data, covariance estimation in small samples, and ordinal data.

PINSTRUP-ANDERSEN, P. (2009) Food security: definition and measurement. *Food security*, 1, 5-7.

PIRKTL, V. (2007) Copula models and dependence concepts. UNIVERSITY OF SOUTHERN CALIFORNIA.

PITT, M., CHAN, D. & KOHN, R. (2006) Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, 93, 537-554.

ROCHON, J. (1996) Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics*, 740-750.

RUSCONE, M. N. & OSMETTI, S. A. (2017) Modelling the dependence in multivariate longitudinal data by pair copula decomposition. *Soft Methods for Data Science.* Springer.

RYU, J.-H. & BARTFELD, J. S. (2012) Household food insecurity during childhood and subsequent health status: the early childhood longitudinal study—kindergarten cohort. *American Journal of Public Health*, 102, e50-e55.

SAVU, C. & TREDE, M. (2006) Hierarchical Archimedean Copulas: International Conference on High Frequency Finance. *Konstanz, Germany*.

SCHEPSMEIER, U., STOEBER, J., BRECHMANN, E. C., GRAELER, B., NAGLER, T., ERHARDT, T., ALMEIDA, C., MIN, A., CZADO, C. & HOFMANN, M. (2015) Package â€˜VineCopulaâ€™. *R package version*, 2.

SCHIRMACHER, D. & SCHIRMACHER, E. (2008) Multivariate dependence modeling using pair-copulas. Technical report.

SHI, P. & YANG, L. (2016) Pair copula constructions for insurance experience rating. *Journal of the American Statistical Association*, 113, 122-133.

SHI, P. & ZHAO, Z. (2018) Predictive Modeling of Multivariate Longitudinal Insurance Claims Using Pair Copula Construction. *arXiv preprint arXiv:1805.07301*.

SHONE, M., DEMISSIE, T., YOHANNES, B. & YOHANNIS, M. (2017) Household food insecurity and associated factors in West Abaya district, Southern Ethiopia, 2015. *Agriculture & Food Security*, 6, 2.

SIRISRISAKULCHAI, J. & SRIBOONCHITTA, S. (2014) Modeling dependence of accident-related outcomes using pair copula constructions for discrete data. *Modeling Dependence in Econometrics. Springer International Publishing*, 215-228.

SKLAR, M. (1959) *Fonctions de répartition à n dimensions et leurs marges*, Université Paris 8.

SMITH, M., MIN, A., ALMEIDA, C. & CZADO, C. (2010) Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105, 1467-1479.

SMITH, M. S. (2015) Copula modelling of dependence in multivariate time series. *International Journal of Forecasting*, 31, 815-833.

STÖBER, J., HONG, H. G., CZADO, C. & GHOSH, P. (2015) Comorbidity of chronic diseases in the elderly: Patterns identified by a copula design for mixed responses. *Computational Statistics & Data Analysis*, 88, 28-39.

SUTKOFF, A. (2014) A regular vine-copula approach to endogenous regressors in brand value estimation. THE UNIVERSITY OF CHICAGO.

SWINDALE, A. & BILINSKY, P. (2007) Household food insecurity access scale (HFIAS) for measurement of household food access: indicator guide (v. 3). *Washington, DC: Food and Nutrition Technical Assistance Project, Academy for Educational Development*.

SYRING, N. A. (2013) Multivariate binary regression models with applications in health care utilisation. NORTHERN ILLINOIS UNIVERSITY.

TANTU, A. T., GAMEBO, T. D., SHENO, B. K. & KABALO, M. Y. (2017) Household food insecurity and associated factors among households in Wolaita Sodo town, 2015. *Agriculture & Food Security*, 6, 19.

TOULOUMIS, A., AGRESTI, A. & KATERI, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, 69, 633-640.

TUTZ, G. & HENNEVOGL, W. (1996) Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, 22, 537-557.

VANDENHENDE, F. O. & LAMBERT, P. (2000) Modeling repeated ordered categorical data using copulas. *Dis cussion Paper No 00-25: www. stat. ucl. ac. be/pub/papers/dp/dp00/dpO025. ps Institut de statistique, Uni versit*.

VARADHAN, R. & GROTHENDIECK, G. (2011) alabama: Constrained nonlinear optimization. *R package version*, 1, 2012.

VERBEKE, G., FIEUWS, S., MOLENBERGHS, G. & DAVIDIAN, M. (2014) The analysis of multivariate longitudinal data: A review. *Statistical methods in medical research,* 23, 42-59.

VERBEKE, G., FIEUWS, S., MOLENBERGHS, G. & DAVIDIAN, M. (2017) The analysis of multivariate longitudinal data: A review. *Statistical methods in medical research,* 26, 112-112.

WINEMAN, A. (2016) Multidimensional household food security measurement in rural Zambia. *Agrekon,* 55, 278-301.

YANG, L., FREES, E. W. & ZHANG, Z. (2020) Nonparametric estimation of copula regression models with discrete outcomes. *Journal of the American Statistical Association,* 115, 707-720.

ZERAY, D. D. N. (2017) Determinants of Rural Household Food Security in Wolaita Zone: The Case of Humbo Woreda. *Journal of Poverty, Investment and Development* 32, ISSN 2422-846X.

ZHANG, W., ZHANG, M. & CHEN, Y. (2019) A Copula-Based GLMM Model for Multivariate Longitudinal Data with Mixed-Types of Responses. *Sankhya B*, 1-27.

Appendices

Appendix A: Questionnaires

UNIVERSITY OF SOUTH AFRICA

COLLEGE OF SCIENCE, ENGINEERING AND TECHNOLOGY

Department of Statistics

Household Questionnaire for Modelling the Stability and Determinants of Household Food Insecurity, February, 2014-2015

Introduction and Consent

My name is____________ and I am attending post graduate class at University of South Africa (UNISA). We are conducting an assessment on modelling the determinant factors of household food insecurity using longitudinal multivariate ordinal logistic regression model. I would like to ask you some questions about you, your household, risks you face relating to food you are engaged in. The questionnaire usually takes between 20 -25 minutes to complete.

Whatever information you provide will be kept strictly confidential and will not be shown to other persons. Participation in this assessment is voluntary and you can choose not to answer any individual questions or all of the questions. However, we hope that you will participate fully in this assessment since your views are important.

Do you have any questions about the survey? May I begin the interview now?

☐

VERBAL CONSENT GIVEN TO INTERVIEW, CHECK BOX

Interview Information

Date of interview: (dd/mm/yyyy) ____/_____/______

Interviewer's name____________________________Signature___________________

Name of supervisor____________________________Signature___________________

Questionnaire ID: ______________________________

1. Area Identification		
No	**Question**	**Response**
1.1	Woreda	______________________
1.2	Kebele	______________________
1.3	Got	____________________
1.4	Household ID	_______________________

2. Demographic and Socioeconomic Characteristics			
No	Question	Response	Skip
2.1	Respondent	1. Household head 2. Housewife 3. Son/daughter 4. Other	
2.2	Age of the respondent in years	_____________________	
2.3	Age of the household head in years	_____________________	
2.4	Sex of respondent	1. Male 2. Female	
2.5	Sex of the household head	1. Male 2. Female	
2.6	Who is the household head?	1. Husband 2. Wife 3. Son/daughter 4. Other (specify)---------	
2.7	Family size	____________________	
2.8	Number of under 5 Children	____________________	
2.9	Whom do the household head live with?	1. Alone 2. Spouse/partner 3. Parents 4. Relatives 5. Others (specify)--------------	
2.10	What is the highest level of education of the	1. Unable to read and write	

	household head attained?	2. Can read and write 3. Regular Primary education (1-8) 4. General secondary education (9-10) 5. Preparatory education (11-12) 6. TVET 7. College /university education	
2.11	What is current marital status of the household's head?	1. Never married 2. Cohabiting 3. Married 4. Divorced 5. Widowed	
2.12	What is the household head current occupation? **(Select that all apply)**	1. Student 2. Unemployed 3. Professional employment 4. Self employed 5. Domestic worker 6. Casual worker 7. Housewife 8. Other (Specify)------------------	
3. Economic and Income Related Questions			
3.1	Main source of household income **(Select that all apply)**	1. Farming 2. Herding → 3. Merchant → 4. Daily labourer → 5. Other (specify)--------------	 3.9 3.10 3.12
3.2	How much is your total farmland size?	Land size in 1. Hectares ____________ 2. Timad ______________ 3. Gasha _______________ 4. Other ____________	

3.3	Slope of your land	1. Plain 2. Hilly 3. Steep	
3.4	How do you perceive the quality or fertility of your land?	1. Fertile 2. Medium fertile 3. Less fertile 4. Poor	
3.5	How many times do you cultivate within a year?	1. Yearly 2. Biannual 3. Three-times	
3.6	What type of the following cereals did you harvest during the last 12 months? **(Select that all apply)**	1. Barely 2. Millet 3. Wheat 4. Sorghum 5. Teff 6. Bean 7. Pea 8. Others specify ---------	
3.7	What are the main problems for farmers' incomes in your village? **(Select that all apply)**	1. Pests 2. Rainfall shortage 3. Disease 4. Lack of improved agricultural product input 5. Households head death 6. Excessive temperature 7. Excess rainfall 8. Increase in market price 9. Fall in market price 10. Property loss 11. Other specify ---------------------	
3.8	Have you used any of the following agricultural	1. Chemical fertilizer	

	technologies during the last 12 months production season? **(Select that all apply)**	2. Pesticides 3. Improved seeds 4. Farm credit 5. Access to irrigation water 6. Nothing 7. Compost 8. Others Specify ------------------	
3.9	What type of livestock do you have? **(Select that all apply)**	1. Ox 6. Donkey 2. Cow 7. Mule 3. Sheep 8. Camel 4. Goat 9. Chicken 5. Horse 10. Other (Specify)------	
3.10	The weathering condition of your village?	1. Hot (Kolla) 2. Medium (Wenadega) 3. Cold (Dega)	
3.11	How was the availability of rain on your village since last year?	1. Very high 2. High 3. Enough 4. Little 5. Very little 6. Too much 7. Too little 8. Other specify _______________	
3.12	Which season is the main production season in your village? **(Select that all apply)**	1. Winter (Dec-Feb) 2. Summer (Jun-Aug) 3. Autumn (March-May) 4. Spring (Sep-Nov)	
3.13	Member in the household contributing financially to incomes	________________	
3.14	Number of persons contributing financially to		

	incomes	________________	
3.15	Average monthly income of your family	1. Less than 500 2. 500-1500 3. 1501-2500 4. 2501-3000 5. 3001-4800 6. 4801-5000 7. Greater than 5001(Specify) ------	
4. Household Food Insecurity Availability, Access and Utilisation Scale tool			
A. Household Food Insecurity Utilisation Scale Tool			

4.1	In the past [24 hours], did you or any household member ate CEREAL CROPS (bread, noodles, biscuits, cookies or any other foods made from millet, sorghum, maize, rice, wheat other locally available grains)	1. Yes 2. No
4.2	In the past [24 hours], did you or any household member ate VITAMIN A RICH VEGETABLES AND FRUITS (carrots, squash, sweet potatoes, ripe mangoes, papayas or other locally available vitamin A-rich fruits or vegetables)	1. Yes 2. No
4.3	In the past [24 hours], did you or any household member ate MEAT (beef, pork, lamb, goat, rabbit, wild game, chicken, duck, or other birds, liver, kidney, heart or other organ meats or blood-based foods)	1. Yes 2. No
4.4	In the past [24 hours], did you or any household member ate EGGS	1. Yes 2. No
4.5	In the past [24 hours], did you or any household member ate FISH (fresh or dried fish or shellfish)	1. Yes 2. No
4.6	In the past [24 hours], did you or any household member ate LEGUMES, NUTS AND SEEDS (beans, peas, lentils, nuts, seeds or foods made from these)	1. Yes 2. No
4.7	In the past [24 hours], did you or any household member ate MILK AND MILK PRODUCTS (milk, cheese, yogurt or other milk products)	1. Yes 2. No
4.8	In the past [24 hours], did you or any household member ate OILS AND FATS (oil, fats or butter added to food or used for cooking)	1. Yes 2. No
4.9	In the past [24 hours], did you or any household member ate SWEETS (sugar, honey, sweetened soda or sugary foods such as chocolates, sweets or candies)	1. Yes 2. No

4.10	Did your households get enough and safe drinking water?	1. Yes 2. No
4.11	Did bone problem happen among your under five families?	1. Yes 2. No
4.12	Did diarrhoea disease mostly happen among your under five families?	1. Yes 2. No
4.13	Did anaemia disease happen among your under five families?	1. Yes 2. No
4.14	Did pregnant woman take balanced diet food than the other family members?	1. Yes 2. No
4.15	Did breast feeding woman take balanced diet food than the other family members?	1. Yes 2. No
4.16	Did you prepare appropriate place for dusts?	1. Yes 2. No
4.17	Does someone among your family members who didn't eat food that others ate it?	1. Ye 2. No
4.18	Has someone who can eat raw food (raw meat, milk and others) among your family members?	1. Yes 2. No
4.19	Do you toilet?	1. Yes 2. No

B. **Household Food Insecurity Availability Scale tool**

4.20	Have you plough land for cereal crops?	1. Yes 2. No	
4.21	If yes for 4.20, which of these statements best describes any cereal crops eaten in your household in the last 12 months through your own production?	1. Enough of the kinds of food we want to eat → 2. Enough but not always the kinds of food we want 3. Sometimes not enough to eat 4. Often not enough to eat	4.23
4.22	Which of these statements best describes that you usually able to buy all	1. Enough of the kinds of food we want to eat 2. Enough but not always the kinds of food we	

	of the cereal crops that you need for you and your family from the local market of your village or surrounding?	want 3. Sometimes not enough to eat 4. Often not enough to eat	
4.23	Have you plough land for fruits?	1. Yes 2. No	
4.24	If yes for 4.23, which of these statements best describes any fresh fruit eaten in your household in the last 12 months through your own production? Interviewer: Do not include juice or fruit that is frozen or canned.	1. Enough of the kinds of fruit we want to eat → 2. Enough but not always the kinds of fruit we want 3. Sometimes not enough to eat 4. Often not enough to eat	4.26
4.25	Which of these statements best describes that you usually able to buy all of the fruit that you need for you and your family from the local market of your village or surrounding?	1. Enough of the kinds of fruit we want to eat 2. Enough but not always the kinds of fruit we want 3. Sometimes not enough to eat 4. Often not enough to eat	
4.26	Have you plough land for vegetables?	1. Yes 2. No	
4.27	If yes for 4.26, which of these statements best describes any vegetables eaten in your household in the last 12 months through your own production?	1. Enough of the kinds of vegetables we want to eat → 2. Enough but not always the kinds of vegetables we want 3. Sometimes not enough to eat 4. Often not enough to eat	4.27
4.28	Which of these statements best describes that you usually able to buy all of the vegetables that you need for you and your family from the local market of your village or surrounding?	1. Enough of the kinds of vegetables we want to eat 2. Enough but not always the kinds of vegetables we want 3. Sometimes not enough to eat 4. Often not enough to eat	
4.29	Have you animals that produces milk?	1. Yes	

		2. No	
4.30	If yes for 4.29, which of these statements best describes any milk products eaten in your household in the last 12 months through your own production?	1. Enough of the kinds of food we want to eat → 2. Enough but not always the kinds of food we want 3. Sometimes not enough to eat 4. Often not enough to eat	4.32
4.31	Which of these statements best describes that you usually able to buy all of the milk products that you need for you and your family from the local market of your village or surrounding?	1. Enough of the kinds of food we want to eat 2. Enough but not always the kinds of food we want 3. Sometimes not enough to eat 4. Often not enough to eat	
4.32	Have you animals that produces milk?	1. Yes 2. No	
4.33	Which of these statements best describes any meat products eaten in your household in the last 12 months through your own production?	1. Enough of the kinds of food we want to eat → 2. Enough but not always the kinds of food we want 3. Sometimes not enough to eat 4. Often not enough to eat	4.35
4.34	Which of these statements best describes that you usually able to buy all of the meat products that you need for you and your family from the local market of your village or surrounding?	1. Enough of the kinds of food we want to eat 2. Enough but not always the kinds of food we want 3. Sometimes not enough to eat 4. Often not enough to eat	
4.35	Which of these statements best describes the kind of foods eaten in your household in the last 12 months through food aid in your village	1. Enough of the kinds of food we want to eat 2. Enough but not always the kinds of food we want 3. Sometimes not enough to eat 4. Often not enough to eat 5. Not aided	
4.36	Which of these statements best describes the kind of water used in your	1. Efficient water we want to use 2. Efficient but not always the kinds of water	

	household in the last 12 months in your village?	we want 3. Sometimes not efficient to use 4. Often not efficient to use	

C. Household Food Insecurity Access Scale tool

4.37	What best describes the food consumed in the household during the past 12 months.(due to lack of money to buy food)	1. Always enough of what wanted 2. Enough but not always what wanted 3. Sometimes not enough food 4. Often not enough food
4.38	In past 12 months were you and your household members worried that your food would run out before you had money to buy more?	1. No 2. Yes *No follow up question on frequency
4.39	In past 12 months did you have to eat the same food daily because you did not have money to buy other food?	1. No 2. Yes *No follow up question on frequency
4.40	In the past 12 months have you or any other adult in your household eaten less food than you wanted to because you did not have enough money to buy food?	1. No 2. Yes ⟶ How often? 1. More than half the time 2. Less than half the time but more than 30 days 3. Less than 30 days but more than 10 days 4. Less than 10 days
4.41	Did you or another adult in your household skip meals during the past 12 months because you did not have enough money to buy food?	1. No 2. Yes ⟶ How often? 1. More than half the time 2. Less than half the time but more than 30 days 3. Less than 30 days but more than 10 days 4. Less than 10 days
4.42	Did you or another adult in your household stop eating for an entire day (during the past 12 months) because you did not have enough money to buy food?	1. No 2. Yes ⟶ How often? 1. Less than half the time but more than 30 days 2. Less than 30 days but more than 10 days

		3. Less than 10 days
4.43	In the past 12 months, did you or anyone in the household borrow money for food from friends or relatives?	1. No 2. Yes ⟶ How often? 1. Less than half the time but more than 30 days 2. Less than 30 days but more than 10 days 3. Less than 10 days
4.44	In the past 12 months, did you or anyone in the household buy food on a credit account or credit card?	1. No 2. Yes ⟶ How often? 1. Less than half the time but more than 30 days 2. Less than 30 days but more than 10 days 3. Less than 10 days

5	**Household Food Insecurity Coping Mechanisms**	
5.1	Since crises, how did you overcome the food shortage? **(Select that all apply)**	1. Eating less/skipping meals 2. Eating food less preferred 3. Food or cash aid 4. Migrating household head to other villages 5. Migrating the younger household members to town 6. Selling assets 7. Eating wild food 8. Selling trees 9. Gardening (to grow food, mainly vegetables and green leaves) 10. Trade (commercial activities) 11. Little crafts 12. Small livestock raising 13. Other/specify -------------------------------
5.2	Did anyone in the family benefit from food aid rations in the last one year? **(Select that all apply)**	1. Yes, emergency food rations 2. Yes, safety net food rations 3. No 4. Other (specify)…………….
5.3	If you are safety net user have you graduate now?	1. Yes 2. No 3. No safety net user

Appendix B: Internal consistence analysis of the data collection tools.

Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Any household member ate CEREAL CROPS	61.00	74.572	.114	.737
Any household member ate VITAMIN A RICH VEGETABLES AND FRUITS	60.26	72.642	.243	.731
Any household member ate MEAT	60.12	74.049	.115	.735
Any household member ate EGGS	60.13	73.863	.116	.735
Any household member ate FISH	60.05	74.678	.112	.737
Any household member ate LEGUMES, NUTS AND SEEDS	60.85	73.268	.173	.733
Any household member ate MILK AND MILK PRODUCTS	60.31	72.401	.257	.730
Any household member ate OILS AND FATS	60.73	72.182	.278	.729
Any household member ate SWEETS	60.66	71.920	.294	.728
Did your households get enough and safe drinking water	60.70	73.010	.166	.733
Any cereal crops eaten in your household in the last 12 months through your own production	58.90	67.433	.360	.721
You are usually able to buy all of the cereal crops that you need for you and your family from the local market	60.46	71.192	.183	.732
Any fresh fruit eaten in your household in the last 12 months through your own production	57.53	67.432	.365	.720
You are usually able to buy all of the fruit that you need for you and your family from the local market	59.60	67.093	.298	.726
Any vegetables eaten in your household in the last 12 months through your own production	57.78	65.426	.417	.715

You are usually able to buy all of the vegetables that you need for you and your family from the local market	60.11	67.892	.303	.725
Any milk products eaten in your household in the last 12 months through your own production	58.24	66.221	.350	.721
You are usually able to buy all of the milk products that you need for you and your family from the local market	59.53	66.916	.297	.726
Any meat products eaten in your household in the last 12 months through your own production	58.02	68.625	.305	.725
You are usually able to buy all of the meat products that you need for you and your family from the local market	60.08	70.398	.147	.738
The kind of water used in your household in the last 12 months in your village	60.22	71.127	.145	.736
Food consumed in the household during the past 12 months. (due to lack of money to buy food)	59.11	66.207	.515	.711
You and your household members worried that your food would run out before you had money to buy more	60.18	71.795	.431	.726
You have to eat the same food daily because you did not have money to buy other food	60.51	72.332	.233	.730
You or any other adult in your household eaten less food than you wanted to because you did not have enough money to buy food	59.59	64.010	.372	.720
You or another adult in your household skip meals during the past 12 months because you did not have enough money to buy food	60.36	68.241	.307	.725

You or another adult in your household stop eating for an entire day (during the past 12 months) because you did not have enough money to buy food	60.94	73.694	.108	.735
You or anyone in the household borrow money for food from friends or relatives	60.37	67.808	.290	.726
You or anyone in the household buy food on a credit account or credit card	60.49	68.342	.262	.728
Cronbach's Alpha				.735

Appendix C: The joint probability distribution based on the D-vine pair copula was displayed as follows.

$$\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)$$

$$= \{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3} C_{13|2}(\frac{C_{12}(F_1(y_1-i_1),F_2(y_2))-C_{12}(F_1(y_1-i_1),F_2(y_2-1))}{F_2(y_2)-F_2(y_2-1)}, \frac{C_{23}(F_2(y_2),F_3(y_3-i_3))-C_{23}(F_2(y_2-1),F_3(y_3-i_3))}{F_2(y_2)-F_2(y_2-1)})\}[F_2(y_2)-F_2(y_2-1)].$$

Among the six proposed bivariate copula functions, AMH bivariate copula was selected for $F_1(y_1)$ and $F_2(y_2)$, Frank for $F_2(y_2)$ and $F_3(y_3)$, and Frank for $F_{1|2}(y_1 \mid y_2)$ and $F_{3|2}(y_3 \mid y_2)$. Hence the simplified joint probability distribution based on the D pair copula was displayed as follows.

when $i_1 = 0$ and $i_3 = 0$,

$$C_{12}^{AMH^{++}} = C_{12}\big(F_1(y_1), F_2(y_2)\big) = F_1(y_1) * F_2(y_2)/(1-\theta_{12}\big(1 - F_1(y_1)\big) * (1 - F_2(y_2)))$$

$$C_{12}^{AMH^{+-}} = C_{12}\big(F_1(y_1), F_2(y_2 - 1)\big) = F_1(y_1) * F_2(y_2 - 1)/(1-\theta_{12}\big(1 - F_1(y_1)\big) * (1 - F_2(y_2 - 1)))$$

$$C_{23}^{Fr^{++}} = C_{23}\big(F_2(y_2), F_3(y_3)\big) = {}^{-1}/_{\theta_{23}}\, log\left(1 + \frac{\big((\exp(-\theta_{23}F_2(y_2)) - 1\big)\big(\exp(-\theta_{23}F_3(y_3)) - 1)\big)}{\exp(-\theta_{23}) - 1}\right)$$

$$C_{23}^{Fr^{-+}} = C_{23}\big(F_2(y_2 - 1), F_3(y_3)\big) = {}^{-1}/_{\theta_{23}}\, log\left(1 + \frac{\big(\exp(-\theta_{23}F_2(y_2 - 1)) - 1\big)\big(\exp(-\theta_{23}F_3(y_3)) - 1\big)}{\exp(-\theta_{23}) - 1}\right)$$

$$C_{13|2}^{Fr^{00}} = {}^{-1}/_{\theta_{13|2}}\, log\left(1 + \frac{\left(\exp\left(-\theta_{13|2}\left(\frac{C_{12}^{AMH^{++}} - C_{12}^{AMH^{+-}}}{F_2(y_2) - F_2(y_2 - 1)}\right)\right) - 1\right)\left(\exp\left(-\theta_{13|2}\left(\frac{C_{23}^{Fr^{++}} - C_{23}^{Fr^{-+}}}{F_2(y_2) - F_2(y_2 - 1)}\right)\right) - 1\right)}{\exp\big(-\theta_{13|2}\big) - 1}\right)$$

when $i_1 = 1$ and $i_3 = 0$

$$C_{12}^{AMH^{-+}} = C_{12}\big(F_1(y_1 - 1), F_2(y_2)\big) = F_1(y_1 - 1) * F_2(y_2)/(1-\theta_{12}\big(1 - F_1(y_1 - 1)\big) * (1 - F_2(y_2)))$$

$$C_{12}^{AMH^{--}} = C_{12}\big(F_1(y_1-1), F_2(y_2-1)\big) = F_1(y_1-1) * F_2(y_2-1)/(1-\theta_{12}\big(1-F_1(y_1-1)\big) * (1-F_2(y_2-1)))$$

$$C_{23}^{Fr^{++}} = C_{23}\big(F_2(y_2), F_3(y_3)\big) = {}^{-1}/_{\theta_{23}}\, log\left(1 + \frac{\big((\exp(-\theta_{23}F_2(y_2)) - 1\big)\big(\exp(-\theta_{23}F_3(y_3)) - 1)\big)}{\exp(-\theta_{23}) - 1}\right)$$

$$C_{23}^{Fr^{-+}} = C_{23}\big(F_2(y_2-1), F_3(y_3)\big) = {}^{-1}/_{\theta_{23}}\, log\left(1 + \frac{\big(\exp(-\theta_{23}F_2(y_2-1)) - 1\big)\big(\exp(-\theta_{23}F_3(y_3)) - 1\big)}{\exp(-\theta_{23}) - 1}\right)$$

$$C_{13|2}^{Fr^{10}} = {}^{-1}/_{\theta_{13|2}}\, log\left(1 + \frac{\left(\exp\left(-\theta_{13|2}\left(\frac{C_{12}^{AMH^{-+}} - C_{12}^{AMH^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)\right) - 1\right)\left(\exp\left(-\theta_{13|2}\left(\frac{C_{23}^{Fr^{++}} - C_{23}^{Fr^{-+}}}{F_2(y_2) - F_2(y_2-1)}\right)\right) - 1\right)}{\exp\big(-\theta_{13|2}\big) - 1}\right)$$

when $i_1 = 0$ and $i_3 = 1$

$$C_{12}^{AMH^{++}} = C_{12}\big(F_1(y_1), F_2(y_2)\big) = F_1(y_1) * F_2(y_2)/(1-\theta_{12}\big(1-F_1(y_1)\big) * (1-F_2(y_2)))$$

$$C_{12}^{AMH^{+-}} = C_{12}\big(F_1(y_1), F_2(y_2-1)\big) = F_1(y_1) * F_2(y_2-1)/(1-\theta_{12}\big(1-F_1(y_1)\big) * (1-F_2(y_2-1)))$$

$$C_{23}^{Fr^{+-}} = C_{23}\big(F_2(y_2), F_3(y_3-1)\big) = {}^{-1}/_{\theta_{23}}\, log\left(1 + \frac{\big((\exp(-\theta_{23}F_2(y_2)) - 1\big)\big(\exp(-\theta_{23}F_3(y_3-1)) - 1)\big)}{\exp(-\theta_{23}) - 1}\right)$$

$$C_{23}^{Fr^{--}} = C_{23}\big(F_2(y_2-1), F_3(y_3-1)\big) = {}^{-1}/_{\theta_{23}}\, log\left(1 + \frac{\big(\exp(-\theta_{23}F_2(y_2-1)) - 1\big)\big(\exp(-\theta_{23}F_3(y_3-1)) - 1\big)}{\exp(-\theta_{23}) - 1}\right)$$

$$C_{13|2}^{Fr^{01}} = {}^{-1}/_{\theta_{13|2}}\, log\left(1 + \frac{\left(\exp\left(-\theta_{13|2}\left(\frac{C_{12}^{AMH^{++}} - C_{12}^{AMH^{+-}}}{F_2(y_2) - F_2(y_2-1)}\right)\right) - 1\right)\left(\exp\left(-\theta_{13|2}\left(\frac{C_{23}^{Fr^{+-}} - C_{23}^{Fr^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)\right) - 1\right)}{\exp\big(-\theta_{13|2}\big) - 1}\right)$$

when $i_1 = 1$ and $i_3 = 1$,

$$C_{12}^{AMH^{-+}} = C_{12}\big(F_1(y_1-1), F_2(y_2)\big) = F_1(y_1-1) * F_2(y_2)/(1-\theta_{12}\big(1-F_1(y_1-1)\big) * (1-F_2(y_2)))$$

$$C_{12}^{AMH^{--}} = C_{12}\big(F_1(y_1-1), F_2(y_2-1)\big) = F_1(y_1-1) * F_2(y_2-1)/(1-\theta_{12}\big(1-F_1(y_1-1)\big) * (1-F_2(y_2-1)))$$

$$C_{23}^{Fr^{+-}} = C_{23}\big(F_2(y_2), F_3(y_3-1)\big) = {}^{-1}\!/_{\theta_{23}}\, log\left(1+\frac{\big((\exp\left(-\theta_{23}F_2(y_2)\right))-1\big)\big(\exp\left(-\theta_{23}F_3(y_3-1)\right)-1\big)\big)}{\exp(-\theta_{23})-1}\right)$$

$$C_{23}^{Fr^{--}} = C_{23}\big(F_2(y_2-1), F_3(y_3-1)\big) = {}^{-1}\!/_{\theta_{23}}\, log\left(1+\frac{\big(\exp\left(-\theta_{23}F_2(y_2-1)\right)-1\big)\big(\exp\left(-\theta_{23}F_3(y_3-1)\right)-1\big)}{\exp(-\theta_{23})-1}\right)$$

$$C_{13|2}^{Fr^{11}} = {}^{-1}\!/_{\theta_{13|2}}\, log\left(1+\frac{\left(\exp\left(-\theta_{13|2}\left(\frac{C_{12}^{AMH^{-+}}-C_{12}^{AMH^{--}}}{F_2(y_2)-\ F_2(y_2-1)}\right)\right)-1\right)\left(\exp\left(-\theta_{13|2}\left(\frac{C_{23}^{Fr^{+-}}-C_{23}^{Fr^{--}}}{F_2(y_2)-\ F_2(y_2-1)}\right)\right)-1\right)}{\exp\big(-\theta_{13|2}\big)-1}\right)$$

Hence,

$$P(Y_1=y_1, Y_2=y_2, Y_3=y_3) = (C_{13|2}^{Fr^{00}} - C_{13|2}^{Fr^{01}} - C_{13|2}^{Fr^{10}} + C_{13|2}^{Fr^{11}})[\ F_2(y_2)-\ F_2(y_2-1)]$$

As a result, the likelihood function is

$$\mathrm{L}(c,\beta,\theta|X) = \prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{j=1}^{J-1}\big[\big(C_{13|2}^{Fr^{00}} - C_{13|2}^{Fr^{01}} - C_{13|2}^{Fr^{10}} + C_{13|2}^{Fr^{11}}\big)[\ F_2(y_2)-\ F_2(y_2-1)]\big]^{y_{1ij}y_{2ij}y_{3ij}}$$

The log-likelihood is

$$\mathrm{l}(c,\beta,\theta|X) = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J-1} y_{1ij}y_{2ij}y_{3ij} * log\Big(\big(C_{13|2}^{Fr^{00}} - C_{13|2}^{Fr^{01}} - C_{13|2}^{Fr^{10}} + C_{13|2}^{Fr^{11}}\big)[\ F_2(y_2)-\ F_2(y_2-1)]\Big)$$

$$\mathrm{l}(c,\beta,\theta|X) = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J-1} y_{1ij}y_{2ij}y_{3ij} * \big\{log\big(C_{13|2}^{Fr^{00}} - C_{13|2}^{Fr^{01}} - C_{13|2}^{Fr^{10}} + C_{13|2}^{Fr^{11}}\big) + log\big(\ F_2(y_2)-\ F_2(y_2-1)\big)\big\}$$

I. The steps possessed to simplify and obtain the log-likelihood maximum likelihood for the D- vine longitudinal discrete random variables application to household food security.

The joint probability distribution of longitudinal ordinal data based on the D pair copula was displayed as follows.

$$\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)$$
$$= \{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3} C_{13|2}(\frac{C_{12}(F_1(y_1-i_1),F_2(y_2))-C_{12}(F_1(y_1-i_1),F_2(y_2-1))}{F_2(y_2)-F_2(y_2-1)}, \frac{C_{23}(F_2(y_2),F_3(y_3-i_3))-C_{23}(F_2(y_2-1),F_3(y_3-i_3))}{F_2(y_2)-F_2(y_2-1)})\}[F_2(y_2)-F_2(y_2-1)].$$

Among the six proposed bivariate copula functions, Gumbel copula was selected for $F_1(y_1)$ and $F_2(y_2)$, Gumbel for $F_2(y_2)$ and $F_3(y_3)$, and Gumbel for $F_{1|2}(y_1 \mid y_2)$ and $F_{3|2}(y_3 \mid y_2)$. Hence the simplified joint probability distribution given by

when $i_1 = 0$ and $i_3 = 0$,

$$C_{12}^{Gu^{++}} = C_{12}(F_1(y_1), F_2(y_2)) = exp\left[-\left(\left(-log(F_1(y_1))\right)^{\theta_{12}} + \left(-log(F_2(y_2))\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{12}^{Gu^{+-}} = C_{12}(F_1(y_1), F_2(y_2-1)) = exp\left[-\left(\left(-log(F_1(y_1))\right)^{\theta_{12}} + \left(-log(F_2(y_2-1))\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{23}^{Gu^{++}} = C_{23}(F_2(y_2), F_3(y_3)) = exp\left[-\left(\left(-log(F_2(y_2))\right)^{\theta_{23}} + \left(-log(F_3(y_3))\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{23}^{Gu^{-+}} = C_{23}(F_2(y_2-1), F_3(y_3)) = exp\left[-\left(\left(-log(F_2(y_2-1))\right)^{\theta_{23}} + \left(-log(F_3(y_3))\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{13|2}^{Gu^{00}} = exp\left[-\left(\left(-log\left(\frac{C_{12}^{Gu^{++}} - C_{12}^{Gu^{+-}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}} + \left(-log\left(\frac{C_{23}^{Gu^{++}} - C_{23}^{Gu^{-+}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}}\right)^{1/\theta_{13|2}}\right]$$

when $i_1 = 1$ and $i_3 = 0$

$$C_{12}^{Gu^{-+}} = C_{12}\big(F_1(y_1-1), F_2(y_2)\big) = exp\left[-\left(\left(-log\big(F_1(y_1-1)\big)\right)^{\theta_{12}} + \left(-log\big(F_2(y_2)\big)\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{12}^{Gu^{--}} = C_{12}\big(F_1(y_1-1), F_2(y_2-1)\big) = exp\left[-\left(\left(-log\big(F_1(y_1-1)\big)\right)^{\theta_{12}} + \left(-log\big(F_2(y_2-1)\big)\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{23}^{Gu^{++}} = C_{23}\big(F_2(y_2), F_3(y_3)\big) = exp\left[-\left(\left(-log\big(F_2(y_2)\big)\right)^{\theta_{23}} + \left(-log\big(F_3(y_3)\big)\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{23}^{Gu^{-+}} = C_{23}\big(F_2(y_2-1), F_3(y_3)\big) = exp\left[-\left(\left(-log\big(F_2(y_2-1)\big)\right)^{\theta_{23}} + \left(-log\big(F_3(y_3)\big)\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{13|2}^{Gu^{10}} = exp\left[-\left(\left(-log\left(\frac{C_{12}^{Gu^{-+}} - C_{12}^{Gu^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}} + \left(-log\left(\frac{C_{23}^{Gu^{++}} - C_{23}^{Gu^{-+}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}}\right)^{1/\theta_{13|2}}\right]$$

when $i_1 = 0$ and $i_3 = 1$

$$C_{12}^{Gu^{++}} = C_{12}\big(F_1(y_1), F_2(y_2)\big) = exp\left[-\left(\left(-log\big(F_1(y_1)\big)\right)^{\theta_{12}} + \left(-log\big(F_2(y_2)\big)\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{12}^{Gu^{+-}} = C_{12}\big(F_1(y_1), F_2(y_2-1)\big) = exp\left[-\left(\left(-log\big(F_1(y_1)\big)\right)^{\theta_{12}} + \left(-log\big(F_2(y_2-1)\big)\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{23}^{Gu^{+-}} = C_{23}\big(F_2(y_2), F_3(y_3-1)\big) = exp\left[-\left(\left(-log\big(F_2(y_2)\big)\right)^{\theta_{23}} + \left(-log\big(F_3(y_3-1)\big)\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{23}^{Gu^{--}} = C_{23}\big(F_2(y_2-1), F_3(y_3-1)\big) = exp\left[-\left(\left(-log\big(F_2(y_2-1)\big)\right)^{\theta_{23}} + \left(-log\big(F_3(y_3-1)\big)\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{13|2}^{Gu^{01}} = exp\left[-\left(\left(-log\left(\frac{C_{12}^{Gu^{++}} - C_{12}^{Gu^{+-}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}} + \left(-log\left(\frac{C_{23}^{Gu^{+-}} - C_{23}^{Gu^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}}\right)^{1/\theta_{13|2}}\right]$$

when $i_1 = 1$ and $i_3 = 1$,

$$C_{12}^{Gu^{-+}} = C_{12}\big(F_1(y_1-1), F_2(y_2)\big) = exp\left[-\left(\left(-log\big(F_1(y_1-1)\big)\right)^{\theta_{12}} + \left(-log\big(F_2(y_2)\big)\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{12}^{Gu^{--}} = C_{12}\big(F_1(y_1-1), F_2(y_2-1)\big) = exp\left[-\left(\left(-log\big(F_1(y_1-1)\big)\right)^{\theta_{12}} + \left(-log\big(F_2(y_2-1)\big)\right)^{\theta_{12}}\right)^{1/\theta_{12}}\right]$$

$$C_{23}^{Gu^{+-}} = C_{23}\big(F_2(y_2), F_3(y_3-1)\big) = exp\left[-\left(\left(-log\big(F_2(y_2)\big)\right)^{\theta_{23}} + \left(-log\big(F_3(y_3-1)\big)\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{23}^{Gu^{--}} = C_{23}\big(F_2(y_2-1), F_3(y_3-1)\big) = exp\left[-\left(\left(-log\big(F_2(y_2-1)\big)\right)^{\theta_{23}} + \left(-log\big(F_3(y_3-1)\big)\right)^{\theta_{23}}\right)^{1/\theta_{23}}\right]$$

$$C_{13|2}^{Gu^{11}} = exp\left[-\left(\left(-log\left(\frac{C_{12}^{Gu^{-+}} - C_{12}^{Gu^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}} + \left(-log\left(\frac{C_{23}^{Gu^{+-}} - C_{23}^{Gu^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)\right)^{\theta_{13|2}}\right)^{1/\theta_{13|2}}\right]$$

Hence,

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = (C_{13|2}^{Gu^{00}} - C_{13|2}^{Gu^{01}} - C_{13|2}^{Gu^{10}} + C_{13|2}^{Gu^{11}})[\, F_2(y_2) - F_2(y_2-1)]$$

As a result, the likelihood function is

$$L(c,\beta,\theta|X) = \prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{j=1}^{J-1}\left[\left(C_{13|2}^{Gu^{00}} - C_{13|2}^{Gu^{01}} - C_{13|2}^{Gu^{10}} + C_{13|2}^{Gu^{11}}\right)[\,F_2(y_2) - F_2(y_2-1)]\right]^{y_{1ij}y_{2ij}y_{3ij}}$$

The log-likelihood is

$$l(c,\beta,\theta|X) = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J-1} y_{1ij}y_{2ij}y_{3ij} * log\left(\left(C_{13|2}^{Gu^{00}} - C_{13|2}^{Gu^{01}} - C_{13|2}^{Gu^{10}} + C_{13|2}^{Gu^{11}}\right)[\,F_2(y_2) - F_2(y_2-1)]\right)$$

$$l(c,\beta,\theta|X) = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J-1} y_{1ij}y_{2ij}y_{3ij} * \left\{log\left(C_{13|2}^{Gu^{00}} - C_{13|2}^{Gu^{01}} - C_{13|2}^{Gu^{10}} + C_{13|2}^{Gu^{11}}\right) + log\left(F_2(y_2) - F_2(y_2-1)\right)\right\}$$

II. The steps possessed to simplify and obtain the log-likelihood maximum likelihood for the D- vine multivariate longitudinal discrete random variables application to household food security.

M-dimensional multivariate discrete random variables were observed repeatedly for T time points and re-ordered the observations of the multivariate series into the univariate outcomes of dimensions $N = T * n$ given by $Y = (Y_1, Y_2, \dots, Y_M)$, where $Y_1 = (y_{11}, y_{21}, \dots, y_{N1})'$, $Y_2 = (y_{12}, y_{22}, \dots, y_{N2})'and\ Y_M = (y_{1m}, y_{2m}, .., y_{Nm})'$. The joint probability distribution of multivariate longitudinal ordinal data based on the D pair copula was displayed as follows.

$$\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)$$
$$= \{\sum_{i_1=0,1}\sum_{i_3=0,1}(-1)^{i_1+i_3} C_{13|2}\left(\frac{C_{12}(F_1(y_1-i_1),F_2(y_2)) - C_{12}(F_1(y_1-i_1),F_2(y_2-1))}{F_2(y_2)-F_2(y_2-1)}, \frac{C_{23}(F_2(y_2),F_3(y_3-i_3)) - C_{23}(F_2(y_2-1),F_3(y_3-i_3))}{F_2(y_2)-F_2(y_2-1)}\right)\}[F_2(y_2)-F_2(y_2-1)].$$

Among the six proposed bivariate copula functions, Clayton (Cl) copula was selected for $F_1(y_1)$ and $F_2(y_2)$, AMH for $F_2(y_2)$ and $F_3(y_3)$, and Independent for $F_{1|2}(y_1 \mid y_2)$ and $F_{3|2}(y_3 \mid y_2)$. Hence the simplified joint probability distribution given by

when $i_1 = 0$ and $i_3 = 0$,

$$C_{12}^{Cl^{++}} = C_{12}\left(F_1(y_1), F_2(y_2)\right) = \left(\left(F_1(y_1)\right)^{-\theta_{12}} + \left(F_2(y_2)\right)^{-\theta_{12}} - 1\right)^{-1/\theta_{12}}$$

$$C_{12}^{Cl^{+-}} = C_{12}\big(F_1(y_1), F_2(y_2-1)\big) = \Big(\big(F_1(y_1)\big)^{-\theta_{12}} + \big(F_2(y_2-1)\big)^{-\theta_{12}} - 1\Big)^{-1/\theta_{12}}$$

$$C_{23}^{AMH^{++}} = C_{23}\big(F_2(y_2), F_3(y_3)\big) = F_2(y_2) * F_2(y_3)/(1-\theta_{23}\big(1-F_2(y_2)\big) * (1-F_3(y_3)))$$

$$C_{23}^{AMH^{-+}} = C_{23}\big(F_2(y_2-1), F_3(y_3)\big) = F_2(y_2-1) * F_2(y_3)/(1-\theta_{23}\big(1-F_2(y_2-1)\big) * (1-F_3(y_3)))$$

$$C_{13|2}^{Ind^{00}} = \left(\frac{C_{12}^{Cl^{++}} - C_{12}^{Cl^{+-}}}{F_2(y_2) - F_2(y_2-1)}\right)\left(\frac{C_{23}^{AMH^{++}} - C_{23}^{AMH^{-+}}}{F_2(y_2) - F_2(y_2-1)}\right)$$

when $i_1 = 1$ and $i_3 = 0$

$$C_{12}^{Cl^{-+}} = C_{12}\big(F_1(y_1-1), F_2(y_2)\big) = \Big(\big(F_1(y_1-1)\big)^{-\theta_{12}} + \big(F_2(y_2)\big)^{-\theta_{12}} - 1\Big)^{-1/\theta_{12}}$$

$$C_{12}^{Cl^{--}} = C_{12}\big(F_1(y_1-1), F_2(y_2-1)\big) = \Big(\big(F_1(y_1-1)\big)^{-\theta_{12}} + \big(F_2(y_2-1)\big)^{-\theta_{12}} - 1\Big)^{-1/\theta_{12}}$$

$$C_{23}^{AMH^{++}} = C_{23}\big(F_2(y_2), F_3(y_3)\big) = F_2(y_2) * F_2(y_3)/(1-\theta_{23}\big(1-F_2(y_2)\big) * (1-F_3(y_3)))$$

$$C_{23}^{AMH^{-+}} = C_{23}\big(F_2(y_2-1), F_3(y_3)\big) = F_2(y_2-1) * F_2(y_3)/(1-\theta_{23}\big(1-F_2(y_2-1)\big) * (1-F_3(y_3)))$$

$$C_{13|2}^{Ind^{10}} = \left(\frac{C_{12}^{Cl^{-+}} - C_{12}^{Cl^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)\left(\frac{C_{23}^{AMH^{++}} - C_{23}^{AMH^{-+}}}{F_2(y_2) - F_2(y_2-1)}\right)$$

when $i_1 = 0$ and $i_3 = 1$

$$C_{12}^{Cl^{++}} = C_{12}\big(F_1(y_1), F_2(y_2)\big) = \Big(\big(F_1(y_1)\big)^{-\theta_{12}} + \big(F_2(y_2)\big)^{-\theta_{12}} - 1\Big)^{-1/\theta_{12}}$$

$$C_{12}^{Cl^{+-}} = C_{12}\big(F_1(y_1), F_2(y_2-1)\big) = \Big(\big(F_1(y_1)\big)^{-\theta_{12}} + \big(F_2(y_2-1)\big)^{-\theta_{12}} - 1\Big)^{-1/\theta_{12}}$$

$$C_{23}^{AMH^{+-}} = C_{23}\big(F_2(y_2), F_3(y_3-1)\big) = F_2(y_2) * F_2(y_3-1)/(1-\theta_{23}\big(1-F_2(y_2)\big) * (1-F_3(y_3-1)))$$

$$C_{23}^{AMH^{--}} = C_{23}\big(F_2(y_2-1), F_3(y_3-1)\big) = F_2(y_2-1) * F_2(y_3-1)/(1-\theta_{23}\big(1-F_2(y_2-1)\big) * (1-F_3(y_3-1)))$$

$$C_{13|2}^{Ind^{01}} = \left(\frac{C_{12}^{Cl^{++}} - C_{12}^{Cl^{+-}}}{F_2(y_2) - F_2(y_2-1)}\right)\left(\frac{C_{23}^{AMH^{+-}} - C_{23}^{AMH^{--}}}{F_2(y_2) - F_2(y_2-1)}\right)$$

when $i_1 = 1$ and $i_3 = 1$,

$$C_{12}^{Cl^{-+}} = C_{12}\big(F_1(y_1 - 1), F_2(y_2)\big) = \Big(\big(F_1(y_1 - 1)\big)^{-\theta_{12}} + \big(F_2(y_2)\big)^{-\theta_{12}} - 1\Big)^{-1/\theta_{12}}$$

$$C_{12}^{Cl^{--}} = C_{12}\big(F_1(y_1 - 1), F_2(y_2 - 1)\big) = \Big(\big(F_1(y_1 - 1)\big)^{-\theta_{12}} + \big(F_2(y_2 - 1)\big)^{-\theta_{12}} - 1\Big)^{-1/\theta_{12}}$$

$$C_{23}^{AMH^{+-}} = C_{23}\big(F_2(y_2), F_3(y_3 - 1)\big) = F_2(y_2) * F_2(y_3 - 1)/(1 - \theta_{23}\big(1 - F_2(y_2)\big) * (1 - F_3(y_3 - 1)))$$

$$C_{23}^{AMH^{--}} = C_{23}\big(F_2(y_2 - 1), F_3(y_3 - 1)\big) = F_2(y_2 - 1) * F_2(y_3 - 1)/(1 - \theta_{23}\big(1 - F_2(y_2 - 1)\big) * (1 - F_3(y_3 - 1)))$$

$$C_{13|2}^{Ind^{11}} = \left(\frac{C_{12}^{Cl^{-+}} - C_{12}^{Cl^{--}}}{F_2(y_2) - F_2(y_2 - 1)}\right)\left(\frac{C_{23}^{AMH^{+-}} - C_{23}^{AMH^{--}}}{F_2(y_2) - F_2(y_2 - 1)}\right)$$

Hence,

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = \left(C_{13|2}^{Ind^{00}} - C_{13|2}^{Ind^{01}} - C_{13|2}^{Ind^{10}} + C_{13|2}^{Ind^{11}}\right)[\,F_2(y_2) - F_2(y_2 - 1)]$$

As a result, the likelihood function is

$$\mathrm{L}(c, \beta, \theta | X) = \prod_{i=1}^{n}\prod_{t=1}^{T}\prod_{j=1}^{J-1}\left[\left(C_{13|2}^{Ind^{00}} - C_{13|2}^{Ind^{01}} - C_{13|2}^{Ind^{10}} + C_{13|2}^{Ind^{11}}\right)[\,F_2(y_2) - F_2(y_2 - 1)]\right]^{y_{1ij}y_{2ij}y_{3ij}}$$

The log-likelihood is

$$\mathrm{l}(c, \beta, \theta | X) = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J-1} y_{1ij}y_{2ij}y_{3ij} * log\left(\left(C_{13|2}^{Ind^{00}} - C_{13|2}^{Ind^{01}} - C_{13|2}^{Ind^{10}} + C_{13|2}^{Ind^{11}}\right)[\,F_2(y_2) - F_2(y_2 - 1)]\right)$$

$$\mathrm{l}(c, \beta, \theta | X) = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J-1} y_{1ij}y_{2ij}y_{3ij} * \left\{log\left(C_{13|2}^{Ind^{00}} - C_{13|2}^{Ind^{01}} - C_{13|2}^{Ind^{10}} + C_{13|2}^{Ind^{11}}\right) + log\big(\,F_2(y_2) - F_2(y_2 - 1)\big)\right\}$$

Appendix D: R codes for the log-likelihood D-vines

I. The R code to estimate the parameters for the simplified log-likelihood function in Appendix B I.

```
# Import the Food security data and Pre-processing ##
### X are Covariates for All Dimensions ##
### Dep is Dependent variables ##

library(Alabama)

X<-read.table(file.choose(), header=TRUE, sep=",")
Dep<-read.table(file.choose(), header=TRUE, sep=",")
Y<-Dep$Y1
Z<-Dep$Y2
W<-Dep$Y3
y1<-ifelse(Y==1,1,0)
y2<-ifelse(Y==2,1,0)
y3<-ifelse(Y==3,1,0)
y4<-ifelse(Y==4,1,0)
z1<-ifelse(Z==1,1,0)
z2<-ifelse(Z==2,1,0)
z3<-ifelse(Z==3,1,0)
z4<-ifelse(Z==4,1,0)
w1<-ifelse(W==1,1,0)
w2<-ifelse(W==2,1,0)
w3<-ifelse(W==3,1,0)
w4<-ifelse(W==4,1,0)
X=as.vector(X)
G=function(z)
{
G=exp(z)/(1+exp(z))
```

```
return(G)
}

g=function(z)
{
g=exp(z)/(1+exp(z))^2
return(g)
}
### Cumulative logit multivariate Ordinal longitudinal Model ###
alpha1<-vector(length=3,mode="numeric")
alpha1[1]<-0
alpha2<-vector(length=3,mode="numeric")
alpha2[1]<-0
alpha3<-vector(length=3,mode="numeric")
alpha3[1]<-0
beta<-vector(length=16,mode="numeric")  ### coefficients for Availability covariates ###
gamma<-vector(length=16,mode="numeric")  ### coefficients for Accessibility covariates ###
zeta<-vector(length=6,mode="numeric")   ### coefficients for Utilisation covariates ###
r1<-vector(length=1,mode="numeric")     ### PCC parameter for Availability and Accessibility
###
r2<-vector(length=1,mode="numeric")     ### PCC parameter for Accessibility and Utilisation
###
r3<-vector(length=1,mode="numeric")     ### PCC parameter for Availability|Accessibility and
Utilisation|Accessibility ###
par<-vector(length=60,mode="numeric")

logL.cum <-function(par)
{
comp1<-comp2<-comp3<-comp4<-vector(length=dim(X)[1],mode="numeric")
z10<-z11<-z12<-z13<-z14<-z20<-z21<-z22<-z23<-z24<-z30<-z31<-z32<-z33<-z34<-
vector(length=dim(X)[1],mode="numeric")
```

```
AMH_11<-AMH_21<-AMH_31<-AMH_41<-AMH_12<-AMH_22<-AMH_32<-AMH_42<-
AMH_13<-AMH_23<-AMH_33<-AMH_43<-AMH_14<-AMH_24<-AMH_34<-AMH_44<-
vector(length=dim(X)[1],mode="numeric")
Fr_001<-Fr_101<-Fr_011<-Fr_111<- Fr_002<-Fr_102<-Fr_012<-Fr_112<-Fr_003<-Fr_103<-
Fr_013<-Fr_113<-Fr_004<-Fr_104<-Fr_014<-Fr_114<-
vector(length=dim(X)[1],mode="numeric")
Fr_11<-Fr_21<-Fr_31<-Fr_41<-Fr_12<-Fr_22<-Fr_32<-Fr_42<-Fr_13<-Fr_23<-Fr_33<-
Fr_43<-Fr_14<-Fr_24<-Fr_34<-Fr_44<-vector(length=dim(X)[1],mode="numeric")
f1<-f2<-f3<-f4<-vector(length=dim(X)[1],mode="numeric")
alpha1<-par[1:3]; alpha2<-par[4:6]; alpha3<-par[7:9]; beta<-par[10:25]; gamma<-par[26:41];
zeta<-par[42:57]; r1<-par[58]; r2<-par[59]; r3<-par[60]
for (ii in 1: dim(X)[1])
{
z10[ii]<--Inf+sum(beta*X[ii,])
z11[ii]<-alpha1[1]+sum(beta*X[ii,])
z12[ii]<-alpha1[2]+sum(beta*X[ii,])
z13[ii]<-alpha1[3]+sum(beta*X[ii,])
z14[ii]<-100+sum(beta*X[ii,])
z20[ii]<--Inf+sum(gamma*X[ii,])
z21[ii]<-alpha2[1]+sum(gamma*X[ii,])
z22[ii]<-alpha2[2]+sum(gamma*X[ii,])
z23[ii]<-alpha2[3]+sum(gamma*X[ii,])
z24[ii]<-100+sum(gamma*X[ii,])
z30[ii]<--Inf+sum(zeta*X[ii,])
z31[ii]<-alpha3[1]+sum(zeta*X[ii,])
z32[ii]<-alpha3[2]+sum(zeta*X[ii,])
z33[ii]<-alpha3[3]+sum(zeta*X[ii,])
z34[ii]<-100+sum(zeta*X[ii,])
}

d<-data.frame(z10, z11, z12, z13, z14)
```

```
z<-data.frame(z20, z21, z22, z23, z24)
w<-data.frame(z30, z31, z32, z33, z34)

AHM_11<-(G(d[,2])*G(z[,2]))/(1-r1*(1-G(d[,2]))*(1-G(z[,2])))
AHM_21<-(G(d[,2])*G(z[,1]))/(1-r1*(1-G(d[,2]))*(1-G(z[,1])))
AHM_31<-(G(d[,1])*G(z[,2]))/(1-r1*(1-G(d[,1]))*(1-G(z[,2])))
AHM_41<-(G(d[,1])*G(z[,1]))/(1-r1*(1-G(d[,1]))*(1-G(z[,1])))

Fr_11<-(-1/r2)*log(1+((exp(-r2*G(z[,2]))-1)*(exp(-r2*G(w[,2]))-1)/(exp(-r2)-1)))
Fr_21<-(-1/r2)*log(1+((exp(-r2*G(z[,1]))-1)*(exp(-r2*G(w[,2]))-1)/(exp(-r2)-1)))
Fr_31<-(-1/r2)*log(1+((exp(-r2*G(z[,2]))-1)*(exp(-r2*G(w[,1]))-1)/(exp(-r2)-1)))
Fr_41<-(-1/r2)*log(1+((exp(-r2*G(z[,1]))-1)*(exp(-r2*G(w[,1]))-1)/(exp(-r2)-1)))

f2<-G(z[,2])-G(z[,1])

Fr_001<-(-1/r3)*log(1+((exp(-r3*(AHM_11 - AHM_21)/f2)-1)*(exp(-r3*(Fr_11 - Fr_21)/f2)-
1)/(exp(-r3)-1)))
Fr_101<-(-1/r3)*log(1+((exp(-r3*(AHM_31 - AHM_41)/f2)-1)*(exp(-r3*(Fr_11 - Fr_21)/f2)-
1)/(exp(-r3)-1)))
Fr_011<-(-1/r3)*log(1+((exp(-r3*(AHM_11 - AHM_21)/f2)-1)*(exp(-r3*(Fr_31 - Fr_41)/f2)-
1)/(exp(-r3)-1)))
Fr_111<-(-1/r3)*log(1+((exp(-r3*(AHM_31 - AHM_41)/f2)-1)*(exp(-r3*(Fr_31 - Fr_41)/f2)-
1)/(exp(-r3)-1)))

comp1<-y1*z1*w1*log((Fr_001 - Fr_101 - Fr_011 + Fr_111)*f2)

AHM_12<-(G(d[,3])*G(z[,3]))/(1-r1*(1-G(d[,3]))*(1-G(z[,3])))
AHM_22<-(G(d[,3])*G(z[,2]))/(1-r1*(1-G(d[,3]))*(1-G(z[,2])))
AHM_32<-(G(d[,2])*G(z[,3]))/(1-r1*(1-G(d[,2]))*(1-G(z[,3])))
AHM_42<-(G(d[,2])*G(z[,2]))/(1-r1*(1-G(d[,2]))*(1-G(z[,2])))
```

```
Fr_12<-(-1/r2)*log(1+((exp(-r2*G(z[,3]))-1)*(exp(-r2*G(w[,3]))-1)/(exp(-r2)-1)))
Fr_22<-(-1/r2)*log(1+((exp(-r2*G(z[,2]))-1)*(exp(-r2*G(w[,3]))-1)/(exp(-r2)-1)))
Fr_32<-(-1/r2)*log(1+((exp(-r2*G(z[,3]))-1)*(exp(-r2*G(w[,2]))-1)/(exp(-r2)-1)))
Fr_42<-(-1/r2)*log(1+((exp(-r2*G(z[,2]))-1)*(exp(-r2*G(w[,2]))-1)/(exp(-r2)-1)))

f3<-G(z[,3])-G(z[,2])

Fr_002<-(-1/r3)*log(1+((exp(-r3*(AHM_12 - AHM_22)/f3)-1)*(exp(-r3*(Fr_12 - Fr_22)/f3)-
1)/(exp(-r3)-1)))
Fr_102<-(-1/r3)*log(1+((exp(-r3*(AHM_32 - AHM_42)/f3)-1)*(exp(-r3*(Fr_12 - Fr_22)/f3)-
1)/(exp(-r3)-1)))
Fr_012<-(-1/r3)*log(1+((exp(-r3*(AHM_12 - AHM_22)/f3)-1)*(exp(-r3*(Fr_32 - Fr_42)/f3)-
1)/(exp(-r3)-1)))
Fr_112<-(-1/r3)*log(1+((exp(-r3*(AHM_32 - AHM_42)/f3)-1)*(exp(-r3*(Fr_32 - Fr_42)/f3)-
1)/(exp(-r3)-1)))

comp2<-y2*z2*w2*log((Fr_002 - Fr_102 - Fr_012 + Fr_112)*f3)

AHM_13<-(G(d[,4])*G(z[,4]))/(1-r1*(1-G(d[,4]))*(1-G(z[,4])))
AHM_23<-(G(d[,4])*G(z[,3]))/(1-r1*(1-G(d[,4]))*(1-G(z[,3])))
AHM_33<-(G(d[,3])*G(z[,4]))/(1-r1*(1-G(d[,3]))*(1-G(z[,4])))
AHM_43<-(G(d[,3])*G(z[,3]))/(1-r1*(1-G(d[,3]))*(1-G(z[,3])))

Fr_13<-(-1/r2)*log(1+((exp(-r2*G(z[,4]))-1)*(exp(-r2*G(w[,4]))-1)/(exp(-r2)-1)))
Fr_23<-(-1/r2)*log(1+((exp(-r2*G(z[,3]))-1)*(exp(-r2*G(w[,4]))-1)/(exp(-r2)-1)))
Fr_33<-(-1/r2)*log(1+((exp(-r2*G(z[,4]))-1)*(exp(-r2*G(w[,3]))-1)/(exp(-r2)-1)))
Fr_43<-(-1/r2)*log(1+((exp(-r2*G(z[,3]))-1)*(exp(-r2*G(w[,3]))-1)/(exp(-r2)-1)))

f4<-G(z[,4])-G(z[,3])
```

```
Fr_003<-(-1/r3)*log(1+((exp(-r3*(AHM_13 - AHM_23)/f4)-1)*(exp(-r3*(Fr_13 - Fr_23)/f4)-1)/(exp(-r3)-1)))
Fr_103<-(-1/r3)*log(1+((exp(-r3*(AHM_33 - AHM_43)/f4)-1)*(exp(-r3*(Fr_13 - Fr_23)/f4)-1)/(exp(-r3)-1)))
Fr_013<-(-1/r3)*log(1+((exp(-r3*(AHM_13 - AHM_23)/f4)-1)*(exp(-r3*(Fr_33 - Fr_43)/f4)-1)/(exp(-r3)-1)))
Fr_113<-(-1/r3)*log(1+((exp(-r3*(AHM_33 - AHM_43)/f4)-1)*(exp(-r3*(Fr_33 - Fr_43)/f4)-1)/(exp(-r3)-1)))

comp3<-y3*z3*w3*log((Fr_003 - Fr_103 - Fr_013 + Fr_113)*f4)

AHM_14<-(G(d[,5])*G(z[,5]))/(1-r1*(1-G(d[,5]))*(1-G(z[,5])))
AHM_24<-(G(d[,5])*G(z[,4]))/(1-r1*(1-G(d[,5]))*(1-G(z[,4])))
AHM_34<-(G(d[,4])*G(z[,5]))/(1-r1*(1-G(d[,4]))*(1-G(z[,5])))
AHM_44<-(G(d[,4])*G(z[,4]))/(1-r1*(1-G(d[,4]))*(1-G(z[,4])))

Fr_14<-(-1/r2)*log(1+((exp(-r2*G(z[,5]))-1)*(exp(-r2*G(w[,5]))-1)/(exp(-r2)-1)))
Fr_24<-(-1/r2)*log(1+((exp(-r2*G(z[,4]))-1)*(exp(-r2*G(w[,5]))-1)/(exp(-r2)-1)))
Fr_34<-(-1/r2)*log(1+((exp(-r2*G(z[,5]))-1)*(exp(-r2*G(w[,4]))-1)/(exp(-r2)-1)))
Fr_44<-(-1/r2)*log(1+((exp(-r2*G(z[,4]))-1)*(exp(-r2*G(w[,4]))-1)/(exp(-r2)-1)))

f5<-G(z[,5])-G(z[,4])

Fr_004<-(-1/r3)*log(1+((exp(-r3*(AHM_14 - AHM_24)/f5)-1)*(exp(-r3*(Fr_14 - Fr_24)/f5)-1)/(exp(-r3)-1)))
Fr_104<-(-1/r3)*log(1+((exp(-r3*(AHM_34 - AHM_44)/f5)-1)*(exp(-r3*(Fr_14 - Fr_24)/f5)-1)/(exp(-r3)-1)))
Fr_014<-(-1/r3)*log(1+((exp(-r3*(AHM_14 - AHM_24)/f5)-1)*(exp(-r3*(Fr_34 - Fr_44)/f5)-1)/(exp(-r3)-1)))
Fr_114<-(-1/r3)*log(1+((exp(-r3*(AHM_34 - AHM_44)/f5)-1)*(exp(-r3*(Fr_34 - Fr_44)/f5)-1)/(exp(-r3)-1)))
```

```
comp4<-y4*z4*w4*log((Fr_004 - Fr_104 - Fr_014 + Fr_114)*f5)

-sum(comp1+comp2+comp3+comp4)
}
hin<-function(par)
{
alpha<-par[1:9]
h<-rep(NA,1)
h[1]<-alpha1[2]-alpha[1]
h[2]<-alpha1[3]-alpha1[2]
h[3]<-alpha2[2]-alpha2[1]
h[4]<-alpha2[3]-alpha2[2]
h[5]<-alpha3[2]-alpha3[1]
h[6]<-alpha3[3]-alpha3[2]
h
}
hin.jac<-function(par){
alpha<-par[1:9]
j<-matrix(NA,6, length(par))
j[1,]<-
c(1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0)
j[2,]<-
c(0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0)
j[3,]<-
c(0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0)
```

```
j[4,]<-
c(0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0)
j[5,]<-
c(0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0)
j[6,]<-
c(0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0)
j
}
par<-c(-4.331,-1.526,0.834,-4.685,-2.885,-.806,-1.526,2.243,4.140,-0.514,-0.698,-0.192,-1.09,-
0.231,-1.258,-0.445,0.432,-0.492,-0.503,0.85,0.586,1.65,-0.578,-1.622,-1.191,-.865,-.751,-
.242,.647,.185,-.223,.058,-.050,-.086,-.267,-1.586,.388,.159,.132,-.920,-.589,-1.065,-.908,-.111,-
.544,-.184,1.117,.910,.349,-.582,.066,1.542,.492,.290,.010,.012,.169, 0.5, 1.5, 1.2)
fit.cum<- auglag(par, logL.cum , hin = hin, hin.jac = hin.jac)
s.e<- sqrt(diag(solve(fit.cum$hessian)))
```

II. The R code to estimate the parameters for the simplified log-likelihood function in Appendix B II.

```
library(alabama)

X<-read.table(file.choose(), header=TRUE, sep=",")
Dep<-read.table(file.choose(), header=TRUE, sep=",")
Y<-Dep$Y1
Z<-Dep$Y2
W<-Dep$Y3

y1<-ifelse(Y==1,1,0)
y2<-ifelse(Y==2,1,0)
y3<-ifelse(Y==3,1,0)
y4<-ifelse(Y==4,1,0)
```

```
z1<-ifelse(Z==1,1,0)
z2<-ifelse(Z==2,1,0)
z3<-ifelse(Z==3,1,0)
z4<-ifelse(Z==4,1,0)
w1<-ifelse(W==1,1,0)
w2<-ifelse(W==2,1,0)
w3<-ifelse(W==3,1,0)
w4<-ifelse(W==4,1,0)

X=as.vector(X)

G=function(z)
{
G=exp(z)/(1+exp(z))
return(G)
}

g=function(z)
{
g=exp(z)/(1+exp(z))^2
return(g)
}

### Cumulative logit Ordinal Model ###
alpha1<-vector(length=3,mode="numeric")
alpha1[1]<-0
alpha2<-vector(length=3,mode="numeric")
alpha2[1]<-0
alpha3<-vector(length=3,mode="numeric")
alpha3[1]<-0
```

```
beta<-vector(length=7,mode="numeric")
gamma<-vector(length=7,mode="numeric")
zeta<-vector(length=7,mode="numeric")
r1<-vector(length=1,mode="numeric")
r2<-vector(length=1,mode="numeric")
r3<-vector(length=1,mode="numeric")

par<-vector(length=33,mode="numeric")

logL.cum <-function(par)
{
comp1<-comp2<-comp3<-comp4<-vector(length=dim(X)[1],mode="numeric")
z10<-z11<-z12<-z13<-z14<-z20<-z21<-z22<-z23<-z24<-z30<-z31<-z32<-z33<-z34<-
vector(length=dim(X)[1],mode="numeric")

Gu_11<-Ga_21<-Ga_31<-Ga_41<-Ga_12<-Ga_22<-Ga_32<-Ga_42<-Ga_13<-Ga_23<-
Ga_33<-Ga_43<-Ga_14<-Ga_24<-Ga_34<-Ga_44<-vector(length=dim(X)[1],mode="numeric")
Gu_11<-Gu_21<-Gu_31<-Gu_41<-Gu_12<-Gu_22<-Gu_32<-Gu_42<-Gu_13<-Gu_23<-
Gu_33<-Gu_43<-Gu_14<-Gu_24<-Gu_34<-Gu_44<-
vector(length=dim(X)[1],mode="numeric")
Ga_001<-Ga_011<-Ga_101<-Ga_111<-Ga_002<-Ga_012<-Ga_102<-Ga_112<-Ga_003<-
Ga_013<-Ga_103<-Ga_113<-Ga_004<-Ga_014<-Ga_104<-Ga_114<-
vector(length=dim(X)[1],mode="numeric")
Gu_001<-Gu_011<-Gu_101<-Gu_111<-Gu_002<-Gu_012<-Gu_102<-Gu_112<-Gu_003<-
Gu_013<-Gu_103<-Gu_113<-Gu_004<-Gu_014<-Gu_104<-Gu_114<-
vector(length=dim(X)[1],mode="numeric")

f1<-f2<-f3<-f4<-vector(length=dim(X)[1],mode="numeric")
alpha1<-par[1:3]; alpha2<-par[4:6]; alpha3<-par[7:9]; beta<-par[10:16]; gamma<-par[17:23];
zeta<-par[24:30]; r1<-par[31]; r2<-par[32]; r3<-par[33]
```

```
for (ii in 1: dim(X)[1])
{
z10[ii]<--Inf+sum(beta*X[ii,])
z11[ii]<-alpha1[1]+sum(beta*X[ii,])
z12[ii]<-alpha1[2]+sum(beta*X[ii,])
z13[ii]<-alpha1[3]+sum(beta*X[ii,])
z14[ii]<-100+sum(beta*X[ii,])

z20[ii]<--Inf+sum(gamma*X[ii,])
z21[ii]<-alpha2[1]+sum(gamma*X[ii,])
z22[ii]<-alpha2[2]+sum(gamma*X[ii,])
z23[ii]<-alpha2[3]+sum(gamma*X[ii,])
z24[ii]<-100+sum(gamma*X[ii,])

z30[ii]<--Inf+sum(zeta*X[ii,])
z31[ii]<-alpha3[1]+sum(zeta*X[ii,])
z32[ii]<-alpha3[2]+sum(zeta*X[ii,])
z33[ii]<-alpha3[3]+sum(zeta*X[ii,])
z34[ii]<-100+sum(zeta*X[ii,])

}
d<-data.frame(z10, z11, z12, z13, z14)
z<-data.frame(z20, z21, z22, z23, z24)
w<-data.frame(z30, z31, z32, z33, z34)

Ga_11<-exp(-((-log(G(d[,2])))^r1 + (-log(G(z[,2])))^r1)^1/r1)
Ga_21<-exp(-((-log(G(d[,2])))^r1 + (-log(G(z[,1])))^r1)^1/r1)
Ga_31<-exp(-((-log(G(d[,1])))^r1 + (-log(G(z[,2])))^r1)^1/r1)
Ga_41<-exp(-((-log(G(d[,1])))^r1 + (-log(G(z[,1])))^r1)^1/r1)

Gu_11<-exp(-((-log(G(z[,2])))^r2 + (-log(G(w[,2])))^r2)^1/r2)
```

```
Gu_21<-exp(-((-log(G(z[,1])))^r2 + (-log(G(w[,2])))^r2)^1/r2)
Gu_31<-exp(-((-log(G(z[,2])))^r2 + (-log(G(w[,1])))^r2)^1/r2)
Gu_41<-exp(-((-log(G(z[,1])))^r2 + (-log(G(w[,1])))^r2)^1/r2)
f2<-G(z[,2])-G(z[,1])

Ga_001<-exp(-((-log((Ga_11 - Ga_21)/f2))^r3 + (-log((Gu_11 - Gu_21)/f2))^r3)^1/r3)
Ga_101<-exp(-((-log((Ga_11 - Ga_21)/f2))^r3 + (-log((Gu_31 - Gu_41)/f2))^r3)^1/r3)
Ga_011<-exp(-((-log((Ga_31 - Ga_41)/f2))^r3 + (-log((Gu_11 - Gu_21)/f2))^r3)^1/r3)
Ga_111<-exp(-((-log((Ga_31 - Ga_41)/f2))^r3 + (-log((Gu_31 - Gu_41)/f2))^r3)^1/r3)

comp1<-y1*z1*w1*log((Ga_001 - Ga_101 - Ga_011 + Ga_111)*f2)


Ga_12<-exp(-((-log(G(d[,3])))^r1 + (-log(G(z[,3])))^r1)^1/r1)
Ga_22<-exp(-((-log(G(d[,3])))^r1 + (-log(G(z[,2])))^r1)^1/r1)
Ga_32<-exp(-((-log(G(d[,2])))^r1 + (-log(G(z[,3])))^r1)^1/r1)
Ga_42<-exp(-((-log(G(d[,2])))^r1 + (-log(G(z[,2])))^r1)^1/r1)

Gu_12<-exp(-((-log(G(z[,3])))^r2 + (-log(G(w[,3])))^r2)^1/r2)
Gu_22<-exp(-((-log(G(z[,2])))^r2 + (-log(G(w[,3])))^r2)^1/r2)
Gu_32<-exp(-((-log(G(z[,3])))^r2 + (-log(G(w[,2])))^r2)^1/r2)
Gu_42<-exp(-((-log(G(z[,2])))^r2 + (-log(G(w[,2])))^r2)^1/r2)

f3<-G(z[,3])-G(z[,2])

Ga_002<-exp(-((-log((Ga_12 - Ga_22)/f3))^r3 + (-log((Gu_12 - Gu_22)/f3))^r3)^1/r3)
Ga_102<-exp(-((-log((Ga_12 - Ga_22)/f3))^r3 + (-log((Gu_32 - Gu_42)/f3))^r3)^1/r3)
Ga_012<-exp(-((-log((Ga_32 - Ga_42)/f3))^r3 + (-log((Gu_12 - Gu_22)/f3))^r3)^1/r3)
Ga_112<-exp(-((-log((Ga_32 - Ga_42)/f3))^r3 + (-log((Gu_32 - Gu_42)/f3))^r3)^1/r3)

comp2<-y2*z2*w2*log((Ga_002 - Ga_102 - Ga_012 + Ga_112)*f3)
```

```
Ga_13<-exp(-((-log(G(d[,4])))^r1 + (-log(G(z[,4])))^r1)^1/r1)
Ga_23<-exp(-((-log(G(d[,4])))^r1 + (-log(G(z[,3])))^r1)^1/r1)
Ga_33<-exp(-((-log(G(d[,3])))^r1 + (-log(G(z[,4])))^r1)^1/r1)
Ga_43<-exp(-((-log(G(d[,3])))^r1 + (-log(G(z[,3])))^r1)^1/r1)

Gu_13<-exp(-((-log(G(z[,4])))^r2 + (-log(G(w[,4])))^r2)^1/r2)
Gu_23<-exp(-((-log(G(z[,3])))^r2 + (-log(G(w[,4])))^r2)^1/r2)
Gu_33<-exp(-((-log(G(z[,4])))^r2 + (-log(G(w[,3])))^r2)^1/r2)
Gu_43<-exp(-((-log(G(z[,3])))^r2 + (-log(G(w[,3])))^r2)^1/r2)

f4<-G(z[,4])-G(z[,3])

Ga_003<-exp(-((-log((Ga_13 - Ga_23)/f4))^r3 + (-log((Gu_13 - Gu_23)/f4))^r3)^1/r3)
Ga_103<-exp(-((-log((Ga_13 - Ga_23)/f4))^r3 + (-log((Gu_33 - Gu_43)/f4))^r3)^1/r3)
Ga_013<-exp(-((-log((Ga_33 - Ga_43)/f4))^r3 + (-log((Gu_13 - Gu_23)/f4))^r3)^1/r3)
Ga_113<-exp(-((-log((Ga_33 - Ga_43)/f4))^r3 + (-log((Gu_33 - Gu_43)/f4))^r3)^1/r3)

comp3<-y3*z3*w3*log((Ga_003 - Ga_103 - Ga_013 + Ga_113)*f4)

Ga_14<-exp(-((-log(G(d[,5])))^r1 + (-log(G(z[,5])))^r1)^1/r1)
Ga_24<-exp(-((-log(G(d[,5])))^r1 + (-log(G(z[,4])))^r1)^1/r1)
Ga_34<-exp(-((-log(G(d[,4])))^r1 + (-log(G(z[,5])))^r1)^1/r1)
Ga_44<-exp(-((-log(G(d[,4])))^r1 + (-log(G(z[,4])))^r1)^1/r1)

Gu_14<-exp(-((-log(G(z[,5])))^r2 + (-log(G(w[,5])))^r2)^1/r2)
Gu_24<-exp(-((-log(G(z[,4])))^r2 + (-log(G(w[,5])))^r2)^1/r2)
Gu_34<-exp(-((-log(G(z[,5])))^r2 + (-log(G(w[,4])))^r2)^1/r2)
Gu_44<-exp(-((-log(G(z[,4])))^r2 + (-log(G(w[,4])))^r2)^1/r2)

f5<-G(z[,5])-G(z[,4])
```

```
Ga_004<-exp(-((-log((Ga_14 - Ga_24)/f5))^r3 + (-log((Gu_14 - Gu_24)/f5))^r3)^1/r3)
Ga_104<-exp(-((-log((Ga_14 - Ga_24)/f5))^r3 + (-log((Gu_34 - Gu_44)/f5))^r3)^1/r3)
Ga_014<-exp(-((-log((Ga_34 - Ga_44)/f5))^r3 + (-log((Gu_14 - Gu_24)/f5))^r3)^1/r3)
Ga_114<-exp(-((-log((Ga_34 - Ga_44)/f5))^r3 + (-log((Gu_34 - Gu_44)/f5))^r3)^1/r3)

comp4<-y4*z4*w4*log((Ga_004 - Ga_104 - Ga_014 + Ga_114)*f5)

-sum(comp1+comp2+comp3+comp4)
}
hin<-function(par)
{
alpha<-par[1:9]
h<-rep(NA,1)
h[1]<-alpha1[2]-alpha[1]
h[2]<-alpha1[3]-alpha1[2]
h[3]<-alpha2[2]-alpha2[1]
h[4]<-alpha2[3]-alpha2[2]
h[5]<-alpha3[2]-alpha3[1]
h[6]<-alpha3[3]-alpha3[2]
h}
hin.jac<-function(par){
alpha<-par[1:9]
j<-matrix(NA,6, length(par))
j[1,]<-c(1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[2,]<-c(0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[3,]<-c(0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[4,]<-c(0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[5,]<-c(0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[6,]<-c(0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j
}
```

```
par<-c(-2.515, .104, 2.926, -2.157, .444, 3.298, -3.366, -.244, 2.113, .259, .990, -.222, .299, .905,
-.360, -.363, .258, 1.103, -.215, .405, .927, .003, -.186, .022, .896, -.577, -.456, -.717, .412, .984,
1.1, 1.1, 1.1)
fit.cum<- auglag(par, logL.cum , hin = hin, hin.jac = hin.jac)
s.e<- sqrt(diag(solve(fit.cum$hessian)))
```

III. Code to estimate the parameters for the simplified log-likelihood function in Appendix B III.

```
# Import the Food security data and Preprocessing ##
### X1 are Covariates for Availability Dimension ##
### X2 are Covariates for Accessibility Dimension ##
### X3 are Covariates for Utilisation Dimension ##
### Dep are Dependent variables ##

library(alabama)

X1<-read.table(file.choose(), header=TRUE, sep=",")
X2<-read.table(file.choose(), header=TRUE, sep=",")
X3<-read.table(file.choose(), header=TRUE, sep=",")
Dep<-read.table(file.choose(), header=TRUE, sep=",")
Y<-Dep$Y1
Z<-Dep$Y2
W<-Dep$Y3
y1<-ifelse(Y==1,1,0)
y2<-ifelse(Y==2,1,0)
y3<-ifelse(Y==3,1,0)
y4<-ifelse(Y==4,1,0)
z1<-ifelse(Z==1,1,0)
z2<-ifelse(Z==2,1,0)
z3<-ifelse(Z==3,1,0)
```

```
z4<-ifelse(Z==4,1,0)
w1<-ifelse(W==1,1,0)
w2<-ifelse(W==2,1,0)
w3<-ifelse(W==3,1,0)
w4<-ifelse(W==4,1,0)
X1=as.vector(X1)
X2=as.vector(X2)
X3=as.vector(X3)
G=function(z)
{
G=exp(z)/(1+exp(z))
return(G)
}

g=function(z)
{
g=exp(z)/(1+exp(z))^2
return(g)
}
### Cumulative logit multivariate Ordinal longitudinal Model ###
alpha1<-vector(length=3,mode="numeric")
alpha1[1]<-0
alpha2<-vector(length=3,mode="numeric")
alpha2[1]<-0
alpha3<-vector(length=3,mode="numeric")
alpha3[1]<-0
beta<-vector(length=13,mode="numeric")  ### coefficients for Availability covariates ###
gamma<-vector(length=11,mode="numeric")  ### coefficients for Accessibility covariates ###
zeta<-vector(length=9,mode="numeric")   ### coefficients for Utilisation covariates ###
r1<-vector(length=1,mode="numeric")     ### PCC parameter for Availability and Accessibility
###
```

```
r2<-vector(length=1,mode="numeric")     ### PCC parameter for Accessibility and Utilisation
###
par<-vector(length=44,mode="numeric")

logL.cum <-function(par)
{
comp1<-comp2<-comp3<-comp4<-vector(length=dim(X1)[1],mode="numeric")
z10<-z11<-z12<-z13<-z14<-z20<-z21<-z22<-z23<-z24<-z30<-z31<-z32<-z33<-z34<-
vector(length=dim(X1)[1],mode="numeric")
Cl_11<- Cl_21<- Cl_31<- Cl_41<-Cl_11<- Cl_21<- Cl_31<- Cl_41<-Cl_11<- Cl_21<- Cl_31<-
Cl_41<-Cl_11<- Cl_21<- Cl_31<- Cl_41<-vector(length=dim(X1)[1],mode="numeric")
AMH_11<-AMH_21<-AMH_31<-AMH_41<-AMH_12<-AMH_22<-AMH_32<-AMH_42<-
AMH_13<-AMH_23<-AMH_33<-AMH_43<-AMH_14<-AMH_24<-AMH_34<-AMH_44<-
vector(length=dim(X1)[1],mode="numeric")
IND_001<-IND_101<-IND_011<-IND_111<-IND_002<-IND_102<-IND_012<-IND_112<-
IND_003<-IND_103<-IND_013<-IND_113<-IND_004<-IND_104<-IND_014<-IND_114<-
vector(length=dim(X1)[1],mode="numeric")
f1<-f2<-f3<-f4<-vector(length=dim(X1)[1],mode="numeric")
alpha1<-par[1:3]; alpha2<-par[4:6]; alpha3<-par[7:9]; beta<-par[10:22]; gamma<-par[23:33];
zeta<-par[34:42]; r1<-par[44]; r2<-par[44]
for (ii in 1: dim(X1)[1])
{
z10[ii]<--Inf+sum(beta*X1[ii,])
z11[ii]<-alpha1[1]+sum(beta*X1[ii,])
z12[ii]<-alpha1[2]+sum(beta*X1[ii,])
z13[ii]<-alpha1[3]+sum(beta*X1[ii,])
z14[ii]<-100+sum(beta*X1[ii,])
z20[ii]<--Inf+sum(gamma*X2[ii,])
z21[ii]<-alpha2[1]+sum(gamma*X2[ii,])
z22[ii]<-alpha2[2]+sum(gamma*X2[ii,])
z23[ii]<-alpha2[3]+sum(gamma*X2[ii,])
```

```
z24[ii]<-100+sum(gamma*X2[ii,])
z30[ii]<--Inf+sum(zeta*X3[ii,])
z31[ii]<-alpha3[1]+sum(zeta*X3[ii,])
z32[ii]<-alpha3[2]+sum(zeta*X3[ii,])
z33[ii]<-alpha3[3]+sum(zeta*X3[ii,])
z34[ii]<-100+sum(zeta*X3[ii,])
}

d<-data.frame(z10, z11, z12, z13, z14)
z<-data.frame(z20, z21, z22, z23, z24)
w<-data.frame(z30, z31, z32, z33, z34)

Cl_11<-((G(d[,2]))^(-r1) + (G(z[,2]))^(-r1) - 1)^(-1/r1)
Cl_21<-((G(d[,2]))^(-r1) + (G(z[,1]))^(-r1) - 1)^(-1/r1)
Cl_31<-((G(d[,1]))^(-r1) + (G(z[,2]))^(-r1) - 1)^(-1/r1)
Cl_41<-((G(d[,1]))^(-r1) + (G(z[,1]))^(-r1) - 1)^(-1/r1)

AMH_11<-(G(z[,2])*G(w[,2]))/((1-r2*(1-G(z[,2]))*(1-G(w[,2]))))
AMH_21<-(G(z[,1])*G(w[,2]))/((1-r2*(1-G(z[,1]))*(1-G(w[,2]))))
AMH_31<-(G(z[,2])*G(w[,1]))/((1-r2*(1-G(z[,2]))*(1-G(w[,1]))))
AMH_41<-(G(z[,1])*G(w[,1]))/((1-r2*(1-G(z[,1]))*(1-G(w[,1]))))

f2<-G(z[,2])-G(z[,1])
IND_001<-((AMH_11-AMH_21)/f2)*((Cl_11-Cl_21)/f2)
IND_101<-((AMH_31-AMH_41)/f2)*((Cl_11-Cl_21)/f2)
IND_011<-((AMH_11-AMH_21)/f2)*((Cl_31-Cl_41)/f2)
IND_111<-((AMH_31-AMH_41)/f2)*((Cl_31-Cl_41)/f2)

comp1<-y1*z1*w1*log((IND_001 - IND_101 - IND_011 + IND_111)*f2)

Cl_12<-((G(d[,3]))^(-r1) + (G(z[,3]))^(-r1) - 1)^(-1/r1)
```

```
Cl_22<-((G(d[,3]))^(-r1) + (G(z[,2]))^(-r1) - 1)^(-1/r1)
Cl_32<-((G(d[,2]))^(-r1) + (G(z[,3]))^(-r1) - 1)^(-1/r1)
Cl_42<-((G(d[,2]))^(-r1) + (G(z[,2]))^(-r1) - 1)^(-1/r1)

AMH_12<-(G(z[,3])*G(w[,3]))/((1-r2*(1-G(z[,3]))*(1-G(w[,3]))))
AMH_22<-(G(z[,2])*G(w[,3]))/((1-r2*(1-G(z[,2]))*(1-G(w[,3]))))
AMH_32<-(G(z[,3])*G(w[,2]))/((1-r2*(1-G(z[,3]))*(1-G(w[,2]))))
AMH_42<-(G(z[,2])*G(w[,2]))/((1-r2*(1-G(z[,2]))*(1-G(w[,2]))))

f3<-G(z[,3])-G(z[,2])
IND_002<-((AMH_12-AMH_22)/f3)*((Cl_12-Cl_22)/f3)
IND_102<-((AMH_32-AMH_42)/f3)*((Cl_12-Cl_22)/f3)
IND_012<-((AMH_12-AMH_22)/f3)*((Cl_32-Cl_42)/f3)
IND_112<-((AMH_32-AMH_42)/f3)*((Cl_32-Cl_42)/f3)

comp2<-y2*z2*w2*log((IND_002 - IND_102 - IND_012 + IND_112)*f3)

Cl_13<-((G(d[,4]))^(-r1) + (G(z[,4]))^(-r1) - 1)^(-1/r1)
Cl_23<-((G(d[,4]))^(-r1) + (G(z[,3]))^(-r1) - 1)^(-1/r1)
Cl_33<-((G(d[,3]))^(-r1) + (G(z[,4]))^(-r1) - 1)^(-1/r1)
Cl_43<-((G(d[,3]))^(-r1) + (G(z[,3]))^(-r1) - 1)^(-1/r1)

AMH_13<-(G(z[,4])*G(w[,4]))/((1-r2*(1-G(z[,4]))*(1-G(w[,4]))))
AMH_23<-(G(z[,3])*G(w[,4]))/((1-r2*(1-G(z[,3]))*(1-G(w[,4]))))
AMH_33<-(G(z[,4])*G(w[,3]))/((1-r2*(1-G(z[,4]))*(1-G(w[,3]))))
AMH_43<-(G(z[,3])*G(w[,3]))/((1-r2*(1-G(z[,3]))*(1-G(w[,3]))))

f4<-G(z[,4])-G(z[,3])
IND_003<-((AMH_13-AMH_23)/f4)*((Cl_13-Cl_23)/f4)
IND_103<-((AMH_33-AMH_43)/f4)*((Cl_13-Cl_23)/f4)
IND_013<-((AMH_13-AMH_23)/f4)*((Cl_33-Cl_43)/f4)
```

```
IND_113<-((AMH_33-AMH_43)/f4)*((Cl_33-Cl_43)/f4)

comp3<-y3*z3*w3*log((IND_003 - IND_103 - IND_013 + IND_113)*f4)


Cl_14<-((G(d[,5]))^(-r1) + (G(z[,5]))^(-r1) - 1)^(-1/r1)
Cl_24<-((G(d[,5]))^(-r1) + (G(z[,4]))^(-r1) - 1)^(-1/r1)
Cl_34<-((G(d[,4]))^(-r1) + (G(z[,5]))^(-r1) - 1)^(-1/r1)
Cl_44<-((G(d[,4]))^(-r1) + (G(z[,4]))^(-r1) - 1)^(-1/r1)

AMH_14<-(G(z[,5])*G(w[,5]))/((1-r2*(1-G(z[,5]))*(1-G(w[,5]))))
AMH_24<-(G(z[,4])*G(w[,5]))/((1-r2*(1-G(z[,4]))*(1-G(w[,5]))))
AMH_34<-(G(z[,5])*G(w[,4]))/((1-r2*(1-G(z[,5]))*(1-G(w[,4]))))
AMH_44<-(G(z[,4])*G(w[,4]))/((1-r2*(1-G(z[,4]))*(1-G(w[,4]))))

f5<-G(z[,5])-G(z[,4])
IND_004<-((AMH_14-AMH_24)/f5)*((Cl_14-Cl_24)/f5)
IND_104<-((AMH_34-AMH_44)/f5)*((Cl_14-Cl_24)/f5)
IND_014<-((AMH_14-AMH_24)/f5)*((Cl_34-Cl_44)/f5)
IND_114<-((AMH_34-AMH_44)/f5)*((Cl_34-Cl_44)/f5)

comp4<-y4*z4*w4*log((IND_004 - IND_104 - IND_014 + IND_114)*f5)

-sum(comp1+comp2+comp3+comp4)
}

hin<-function(par)
{
alpha<-par[1:9]
h<-rep(NA,1)
h[1]<-alpha1[2]-alpha[1]
```

```
h[2]<-alpha1[3]-alpha1[2]
h[3]<-alpha2[2]-alpha2[1]
h[4]<-alpha2[3]-alpha2[2]
h[5]<-alpha3[2]-alpha3[1]
h[6]<-alpha3[3]-alpha3[2]
h
}
hin.jac<-function(par)
{
alpha<-par[1:9]
j<-matrix(NA,9, length(par))
j[1,]<-c(1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[2,]<-c(0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[3,]<-c(0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[4,]<-c(0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[5,]<-c(0,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j[6,]<-c(0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
j
}
par<-c(-4.066, -1.203, 1.072, -2.577, -.679, 1.256, -5.060, -1.239, 1.185, -.060, -.409, -.385, -
1.137, -1.443, .380, -.730, .279, .458, .265, 1.064, -.680, .834, .073, -1.095, -1.207, .268, .442,
.250, .274, .294, .091, -.766, -.634, .039, -.539, -.056, -1.278, -1.429, .332, -.341, .362, -.363, 1.5,
0.5)
fit.cum<- auglag(par, logL.cum , hin = hin, hin.jac = hin.jac)
s.e<- sqrt(diag(solve(fit.cum$hessian)))
```

Appendix E: Plagiarism Report

Yimam Final Thesis

by Yimam J

Submission date: 05-Nov-2020 07:47AM (UTC+0200)

Submission ID: 1438690888

File name: Yimam_JA_50879235-edited1305_Final_20_10_20.docx (1.55M)

Word count: 46959

Character count: 285422

Yimam Final Thesis

ORIGINALITY REPORT

3%	2%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Anastasios Panagiotelis, Claudia Czado, Harry Joe. "Pair Copula Constructions for Multivariate Discrete Data", Journal of the American Statistical Association, 2012 Publication	1%
2	link.springer.com Internet Source	1%
3	Submitted to wbhs Student Paper	1%
4	ageconsearch.umn.edu Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		

Appendix F: Ethical Clearance Approval

UNISA university of south africa

UNISA SCHOOL OF SCIENCE ETHICS REVIEW COMMITTEE

26 June 2019

ERC Reference # : 2019/SSR-ERC/031
Name : Mr Jemal Ayalew Yimam
Student #: 50879235

Dear Mr JA Yimam

Decision: Ethics Approval from 26 June 2019 to 25 June 2024.

Researcher(s):	Name:	Mr Jemal Ayalew Yimam
	E-mail address:	50879235@mylife.unisa.ac.za
	Cell #:	+251920488641
Supervisor (s):	Name:	Prof John O Olaomi
	E-mail address:	olaomjo@unisa.ac.za
	Telephone #:	+27116709056
	Cell #:	+27723240066

RECEIVED
2019 -06- 26
OFFICE OF THE EXECUTIVE DEAN
College of Science, Engineering and Technology

Working title of research:

Modeling the Determinant Factors of Household Food Insecurity: A Multivariate Longitudinal Ordinal Logistic Regression Approach

Qualification: PHD

Thank you for the application for research ethics clearance by the Unisa School of Science Ethics Review Committee for the above mentioned research. Ethics approval is granted until **25 June 2024**.

*The **low risk application** was **reviewed** by the School of Science Ethics Review Committee on 26 June 2019 in compliance with the Unisa Policy on Research Ethics and the Standard Operating Procedure on Research Ethics Risk Assessment.*

The proposed research may now commence with the provisions that:

1. The researcher(s) will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

2. Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study should be communicated in writing to the *School of Science Ethics Review Committee.*
3. The researcher(s) will conduct the study according to the methods and procedures set out in the approved application.
4. Any changes that can affect the study-related risks for the research participants, particularly in terms of assurances made with regards to the protection of participants' privacy and the confidentiality of the data, should be reported to the Committee in writing, accompanied by a progress report.
5. The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study. Adherence to the following South African legislation is important, if applicable: Protection of Personal Information Act, no 4 of 2013; Children's act no 38 of 2005 and the National Health Act, no 61 of 2003.
6. Only de-identified research data may be used for secondary research purposes in future on condition that the research objectives are similar to those of the original research. Secondary use of identifiable human research data require additional ethics clearance.
7. No field work activities may continue after the expiry date (**25 June 2024**). Submission of a completed research ethics progress report will constitute an application for renewal of Ethics Research Committee approval.
8. Field work activities may only commence from the date on this ethics certificate.

Note:

*The reference number **2019/SSR-ERC/031** should be clearly indicated on all forms of communication with the intended research participants, as well as with the Committee.*

Yours sincerely,

Ms S Muchengetwa
Chair : School of Science ERC
Tel: 011 670 9253
E-mail: muchess@unisa.ac.za

Prof Mantile Lekala
Director : Science
Tel: 011 670 9091
E-mail:lekalml@unisa.ac.za

Prof BB Mamba
Executive Dean : CSET
Tel: 011 670 9231
Email:mambabb@unisa.ac.za

University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

Appendix F: Language editing certificate

7542 Galangal Street
Lotus Gardens
Pretoria
0008

07 January 2021

TO WHOM IT MAY CONCERN

This certificate serves to confirm that I have edited JA Yimam's thesis entitled, **Modelling the Stability and Determinants of Household Food Insecurity: A Multivariate Longitudinal Ordinal Logistic Regression Approach.**

I found the work easy and intriguing to read. Much of my editing basically dealt with obstructionist technical aspects of language, which could have otherwise compromised smooth reading as well as the sense of the information being conveyed. I hope that the work will be found to be of an acceptable standard. I am a member of Professional Editors' Guild.

Hereunder are my particulars:

Jack Chokwe (Mr)
Contact numbers: 072 214 5489
jackchokwe@gmail.com