COPULA JOINT MODELLING OF FOOD INSECURITY INDICATORS WITH APPLICATION TO FOOD INSECURITY PREVALENCE (FIP), HOUSEHOLD DIETARY DIVERSITY SCORE (HDDS) AND MONTHS OF INADEQUATE HOUSEHOLD FOOD PROVISIONING (MIHFP)

Laina Mbongo[1]*, Lillian Pazvakawambwa, Lawrence Kazembe
Department of Computing, Mathematics and Statistical Sciences,
University of Namibia, Windhoek, Namibia

Abstract

Food insecurity is expressed using various indicators to measure availability, access, utilization and stability. Some of the indicators used are household food insecurity prevalence (HFIP), household dietary diversity score (HDDS) and months of inadequate household food provisioning (MIHFP). These measures are often assumed to be independent, since they capture different spectrums of food insecurity. However, these are correlated to each other, and their dependence has rarely been analyzed. This study used generalized joint regression models through copulas to estimate the relationship between food security outcomes/indicators and exposure variables. Both Bernoulli and Poisson marginals were assumed to quantify both binary and count response variables. We further explored partial observability and sample selection in the outcomes. A national cross-sectional survey, NHIES, of 2015/2016 was used in this analysis. The results indicated that both the Frank copula and bivariate normal copula fitted the data better of establishing the relationship between HFIP and HDDS (AIC=2287.296), and between HFIP and MIHFP (AIC=2072.708) respectively. The partial observability and sample selection analysis to account for measurement errors indicated that there was no statistically significant relationship between the food insecurity indicators and the exposure variables. The chapter thus concluded that copula approaches provide an advantage

[1]* Corresponding Author Email: inambongo@gmail.com

of analyzing jointly two outcomes in order to test for significant relationships between high-level hierarchical effects (e.g., random effects). Specifically, the bivariate normal and the frank copula were found to fit the data best. One unique feature of the Gaussian Copula is that it does not allow for a different dependence structure between the outcomes while the frank copula does not have tail dependence and it can model both positive and negative dependencies as the normal copula.

Keywords: Copulas, Sample Selection, Partial observability, Household food insecurity prevalence (HFIP), Household dietary diversity score (HDDS), Months of inadequate household food provisioning (MIHFP)

1.1. Introduction

Food security (FS), according to FAO (2002) exists when all people, at all times, have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy lifestyle. Food security thus encompasses four dimensions namely: (1) food availability which addresses the "supply side" of food security and is determined by the level of food production, stock levels and net trade; (2) food accessibility (economic and physical), an adequate supply of food at national or international level does not itself guarantee household level food security; (3) utilization, which is commonly understood as the way the body makes the most of various nutrients in the food; (4) stability: Even if food intake is adequate today, one is still considered to be food insecure if there is inadequate access to food on a periodic basis, risking a deterioration of your nutritional status (FAO, 2010).

A variety of food security measures have been proposed to capture the four components above. These aim to capture the extent of food insecurity at individual and household level. Foremost is the household food insecurity prevalence (HFIP), a categorical measure that classifies each household into either food secure, mildly food insecure, moderately food insecure or severely food insecure. Households are categorized as increasingly food insecure as they respond affirmatively to more severe conditions and/or experience those conditions more frequently (Coates, Swindale, & Bilinsky, 2007). Measures of household dietary diversity (HDD) tend to be of two types: those based on whether an individual food is consumed or not and those that are based on whether any food from a particular group is consumed. According to Coates, Swindale, & Bilinsky (2007), the resource available to the household and the management and availability of these resources throughout the year defines food access, hence the need to estimate the proportion of households with an inadequate food supply in a month. This is considered as months of inadequate household food provisioning (MIHFP).

Although the definition of food security is clear, measurements of the different dimensions of food security are rare. Modelling of food insecurity, dietary diversity and months of inadequate food provisioning has often been applied independently at individual and household level. The main question of interest is, what are the chances that households or individuals that are food insecure are the same households that lack diversity in their diets and further experienced inadequate food provisioning throughout the year. The analysis of interdependence among two or more FS outcomes will help us to see the overall picture among outcomes and their correlations. Joint analysis has several advantages including avoiding multiple tests, increased power, better control of Type I error rates and efficiency handling of missing data (Leon & Wu, 2011).

According to Nieman (2015), the proper implementation of strategic probits and logits, however, is often made impossible by the outcome- rather than actor-specific structure of available data. While there are data on the aggregated outcomes of an interaction, there is no record of each player's actions at each of the interaction stages. During the analysis of observational data, it is often difficult to have data available for each actor at each information set of the game but instead the data is only available for the outcome of an interaction, with little to no data on the individual actions that led to the observed outcome. This translates that observational data such as food insecurity and dietary diversity are only partially observed. Traditional logistic and probit models often ignore the underlying partial observability problem, that might potentially lead to incorrect inferences.

The importance of dealing with these challenges motivated this study to employ alternative strategies that provide great flexibility in joint modeling of multimodal data. When there is an association between the two outcomes, a joint model will provide interesting and improved results than modelling the responses separately. The joint models significantly improve median log-loss and absolute residuals of cross-validation predictions (Broatch & Karl, 2017).

Additionally, the joint models provide the ability to test for significant relationships between high-level hierarchical effects (e.g., random team effects) since significant predictions for outcomes at individual level may not be important at the group level.

Survey data are sometimes affected by systematic non-participation (Marra & Radice, 2017). This can occur through various ways including directly declining to participate in the study. If individuals are selected into (or out of) the sample based on a combination of observed and unobserved characteristics then models that ignore such a mechanism will most likely yield estimates which are not representative of the population of interest. The bias arising from ignoring such systematic non-participation is known as non-random sample selection bias. Another bias arises through partial observability. Partial observability typically occurs when two decisions are made to jointly determine an outcome. By jointly determining the outcome, one might not be able to observe the specific responses of the two decisions but can only observe the joint outcome. The unobserved specific responses often lead to partial observability biasness. The bivariate Probit with partial observability acknowledges the biasness by assuming that the model which determines the observed outcome is a bivariate Probit in which only one of the four outcomes is observed (Marra & Radice, 2017).

The Copula approach is defined as a useful method for deriving joint distributions. The approach relates an arbitrary joint distribution to its corresponding univariate marginal distribution via copula (Skalar (1959) as cited in Kazembe (2016). Copulas have been applied in many applications of statistics such as in insurance, econometrics, medicine, marketing, spatial, time series and even sports (Perrone and Muller, 2016). Copula is a multivariate dependence structure for joint distribution of random variables that are parted from the marginal distribution of individual random variables (Zimmer & Trivedi, 2006). Copulas first link the marginal distribution together to form the joint distribution and then define the nonparametric measures of dependence of pairs of random variables.

In this chapter, we explored joint modelling of HFIAP, HDDS and MIHFP as joint of binary and count variables using copulas. To address shortcomings in traditional logistic and Probit models, we further conducted a bivariate Probit model with partial observability and sample selection to estimate HFIAP and HDDS, as well as HFIAP with MIHFP jointly.

1.2. Materials and Methods

1.2.1. Data

The study used cross-sectional survey data of the Namibian Household and Income Expenditure (NHIES) of 2015/2016. In order to be comparable with standards recommended for Africa by FAO, food groups in the NHIES 2015/2016 were re-grouped and re-arranged in order to make up the 12 food groups for the analysis of HFIP, HDDS and MIHFP. Statistical package R Version 3.6 was used to compute joint modelling of copulas. Three outcome variables were used in this chapter, namely Household food insecurity prevalence, household dietary diversity score, and months on inadequate household food provisioning.

1.2.2. Joint Modeling (JM)

Joint modelling has been defined according to the type of data used. This chapter adopted the definition of Marra & Radice (2017). Let us assume that there are two binary random variables (Y_{i1}, Y_{i2}) , , for $i = 1, \ldots, n,$ where n represents the sample size. The probability of event $(Y_{i1} = 1, Y_{i2} = 1)$ can be defined as:

$$p_{11i} = P(Y_{i1} = 1, Y_{i2} = 1) = C(P(Y_{i1} = 1), P(Y_{i2} = 1); \theta_i), \qquad (1)$$

Where $P\left(Y_{ij} = 1\right) = 1 - F_j\left(-\eta_{ji}\right)\, for\ j = 1,2, F_j$ $(\cdot)$ is the cumulative distribution function (cdf) of a standardized univariate distribution (in this case Gaussian, logistic or Gumbel), η_{ji} $\in$ $n_{ij} \epsilon\ \mathbb{R}$ is an additive predictor, C is a two-place copula function and θ_i is an association parameter measuring the dependence between the two random variables.

The marginal c.d.f.s in this model are conditioned on covariates through η_{1i} and η_{2i}, but for notational convenience they are suppressed when expressing them. The dependence parameter is provided as a function of an additive predictor because, for instance, the strength and direction of the relationship between the two marginals may differ between sets of observations. That is, $\theta_i = m(\eta_{ci})$, where m is a one-to-one transformation which ensures that θ_i lies in its range.

1.2.3. Parameter Estimation

The model specification allows for a high degree of flexibility in modeling covariate effects. If an unpenalized approach is employed to estimate the model's parameters, then over-fitting is the likely consequence. To prevent this, Marra & Radice (2017) maximized $\ell_p(\delta) = \ell(\delta) - \frac{1}{2}\delta^T S\delta$, where ℓ_p is the penalized model's log-likelihood, $\delta^T = (\beta_1^T, \beta_2^T, \beta_c^T)$ and $S = diag(D_1, D_2, D_c)$. The smoothing parameter vectors are collected in the overall vector $= (\lambda, \lambda_2^T, \lambda_c^T)$. Practically, it is advised that estimation of δ and λ should be obtained by using a stable and efficient trust region algorithm that is based on first and second order analytical derivative information, with integrated automatic multiple smoothing parameter selection (Marra & Radice, 2017).

1.2.4. Bivariate Binary Model with Non-random Sample Selection

According to Marra and Radice (2017), non-random sample selection occurs when individuals select themselves into (or out of) the sample based on a combination of observed and unobserved characteristics. Marra and Radice (2017) further noted that models that fail to take into account such a systematic selection could produce results that are unrepresentative of the population of interest. By adopting a two-equation structural latent variable framework where one equation defines the selection process (Y_{i1}) and the other describes the outcome Y_{i2}, a bivariate binary selection model may be used to address this problem and correct for non-random sample selection. (Y_{i1}) indicates whether an individual is selected into the sample whereas (Y_{i2}) is the outcome which is observed only if the individual is selected. In the same vein, to the endogenous model, the errors of the two equations are expected to follow a bivariate distribution with association parameter θ_i. In this case, the first additive looks like (Marra & Radice, 2017):

$$n_2 = \beta_{20}1_{ns} + Z_{21}\beta_{21} + \cdots + Z_{2k2}\beta_{2k2} = Z_2\beta_{2,} \quad (2)$$

$$n_c = \beta_{c0}1_{ns} + Z_{c1}\beta_{c1} + \cdots + Z_{ckc}\beta_{ckc} = Z_c\beta_{c,} \quad (3)$$

where 1_{ns} is an n_s-dimensional vector made up of ones corresponding to the selected observations, and Z_2 and Z_c have n_s rows. The log-likelihood function of the sample is:

$$\ell = \sum_{i=1}^{n}\{`I_{11i}\,log(p_{11i}) + I_{10i}\,log(P_{10i}) + (1 - y_{i1})\,log(p_{oi})\} \quad (4)$$

, where $p_{oi} = F1(-\eta_{1i})$.

1.2.5. Bivariate Probit Model with Partial Observability

The definition of partial observability in this section is derived from Marra & Radice (2017). The model tackles a problem in which an observed binary outcome reflects the joint realization of two unobserved binary outcomes. Therefore, the joint event $(Y_{i1} = 1, Y_{i2} = 1)$ has probability p_{11i} whereas all the other events have probability $1 - p_{11i}$.

The second predictor is defined as:

$$n_2 = \beta_{20}1_n + Z_{21}\beta_{21} + \cdots + Z_{2k2}\beta_{2k2} \quad (5)$$

The log-likelihood function can be written as:

$$\ell = \sum_{i=1}^{n}\{`I_{11i}\,log(p_{11i}) + (1 - I_{11i})\,log(1 - p_{11i})\} \quad (6)$$

Quantities of interest include estimates for p_{11i} and the impacts the covariates have on these probabilities. Note that this model is defined using Gaussian margins and a Gaussian copula.

1.2.6. The Copula Theory

The copula theory is used to determine the joint distribution of two variables and three variables in order to find the interdependence structure among the food security metrics. The copula theory was introduced by Sklar in 1959. It provides the opportunity to combine several single-variable distributions in various families of one, two, or multivariable distributions considering the interdependence of the variables (Mesbahzadeh, et al. 2019). According to Mesbahzadeh et al., (2019), one of the most important advantages of using copulas functions is that the structure of dependency between variables can be defined even if marginal distributions are different, meaning that in order to define a joint distribution function having equal marginal functions for each variable is not necessary.

1.2.7. Copula Functions

Copula functions include a variety of families such as Elliptical (t copula, Normal), Archimedean (Gumbel, Clayton, Frank, Ali-Mikhail-Haq), Extreme value (Husler-Reiss, Galambos, Tawn, and t-EV,Gumbel) and other families, namely, Plackett and Farlie-Gumbel-Morgensterm . Families of Archimedean and Elliptical are mostly considered (Mesbahzadeh, et al. 2019). In this chapter, we used the commonly used bivariate copulas. Table 23 shows a brief description of some copula functions:

Table 1: Copula families (Trivedi and Zimmer, 2005)

	Copula type	Joint CDF	$\boldsymbol{\theta}$	Kendall $\boldsymbol{\tau}$
Archimedean family	Frank	$C(\mu, v;\ \theta) = 1/\theta \ln[1 + \frac{(e^{-\theta\mu} - 1)(e^{-\theta v} - 1)}{e^{-\theta}} - 1]$	$R/\{0\}$	$1 - \left(\frac{4}{\theta}\right)(1 - D_1(\theta))$ $D_k(x) = k/x^k \int_0^x t^k /(\exp(t) - 1)dt$
	Rotated Joe	$1 - [1 - \prod_{i=1}^{m}(1 - (1 - \mu_i)^{\theta})]^{1/\theta}$	$(-\infty, -1)$	$-1 - 4\int x\,log(x)(1 - x)^{-\frac{2(1+\theta)}{\theta}}.dx$
	Rotated Gumbel	$C(\mu, v, \theta) = \mu + v - 1 + c(1 - \mu, 1 - v)$	$(-\infty, -1)$	$-1 - (1/\theta)$

	Rotated Clayton	$C(\mu, v, \theta) = \mu + v - 1 + c(1-\mu, 1 - v)$	$(-\infty, 0)$	$(\theta/(2-\theta)$
Elliptical Family	Gaussian	$C(\mu, v) = \int_0^{\mu} \Phi(\Phi^{-1}(v) - pxy\Phi^{-1}(t) / \sqrt{1-p^2xy})dt$	$(-1, +1)$	$(2/\pi)\ arcsin(\theta)$

1.2.8. Estimation of Parameters of Copula Function

Both parametric and nonparametric methods are used to estimate the parameters of copula function. In the parametric method, the relationship between generator function of each copula and Kendall coefficient Equation (87) is used (Mesbahzadeh, et al. 2019).

$$\tau = \frac{(c-d)}{\binom{n}{2}} \qquad (7)$$

In this equation, *c* and *d* are the number of pairs of concordant and discordant variables and *n* is number of observations. Two pairs of variables (X_i, Y_i) and (X_j, Y_j) are concurring if $X_j > X_i$ and $Y_j > Y_i$ or $X_i > Y_j$ and $Y_i > Y_j$. Alternatively, if $(X_i - X_j)$ $(Y_i - Y_j) > 0$, variables are concordant, and if $(X_i - X_i)$ $(Y_i - Y_j) < 0$, variables are discordant. In the parametric method, using the maximum log-likelihood function Equation (88), parameter of θ is estimated (Mesbahzadeh, et al. 2019).

$$L(\theta) = \sum_{k=1}^{n} log[c_\theta\{F_1(x_{1k}), \dots, F_p(X_{pk})\}] \qquad (8)$$

, where c_θ is the copula density function; F is the marginal distribution function; and $x_{1k}, x_{2k}, \dots., x_{pk}$ $k = 1, \dots, n$ are the dependent random variables.

Log-likelihood function estimates parameter of θ using density copula function. If dependent random variables are as $x_{1k}, x_{2k} \dots, x_{pk}$ $(k = 1, \dots, n)$ with copula function of $F_\theta\left(x_{1k,\dots,}x_{pk}\right) = C_\theta(F_1(X_{1k}), \dots, F_p(X_{pk}))$.

1.2.9. Goodness-of-Fit test for Copula Function

For selecting the best copula function, value of joint empirical probability of the variables were calculated through empirical copula in Equation (89) and then is compared with the values resulted from other copula functions (Archimedean and Elliptical families) (Mesbahzadeh, et al. 2019).

$$C_n(\mu, v) = \frac{1}{n}\sum_{t=1}^{n} 1(\mu_t < \mu, V_t < v) \tag{9}$$

Whereby μ and v are the empirical probabilities of the two variables.

To compare empirical copula with each copula functions, Normalized Root Mean Square Error (NRMSE) and Nash–Sutcliffe coefficient were selected equations (90) and (91)).

$$NRMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} \frac{(P_{ei}-P_i)^2}{(P_{ei},max-P_{ei},min}} \tag{10}$$

$$NSE = 1 - \frac{\sum_{n=1}^{N}(P_{ei}-P_i)^2}{(P_{ei}-\overline{P_i})^2} \tag{11}$$

Where P_{ei} is the value of empirical copula and P_i is the value of the copula theory.

Additionally, two criteria, namely, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Equation (92) and Equation 93) and are used. Furthermore, in equation 92 and equation 93, k is the model parameter, n is the number of observations and L is the value of the maximum log-likelihood function.

$$AIC = 2k - 2\,ln(L) \tag{12}$$

$$BIC = 2n\,log\,L + k\,log(n) \tag{13}$$

1.3. Results

1.3.1. Food Security, Dietary Diversity, and Months of Inadequate Food Provisioning

The following analysis shows the relationship between food security and socio-household characteristics (Table 24).

The variable sex (male, P=0.022), marital status (married (living with spouse), P =0.18), Education (primary and secondary, P<0.001), work status (working full time, P= 0.008, not working, P=0.020), access to water (no piped water, P= 0.037) showed a statistically significant relationship with food insecurity prevalence. Additionally, variables such as marital status (single, P=0.033), education (No education, P= 0.020, secondary, P=0.024), work status (working full time, P = 0.002 and not working-looking, P=0.008), Tenure status (owner/family, P= 0.046), water (no piped water, P= 0.012) and access to a flushing toilet (no toilet, P=0.012) had a statistically significant relationship with HDDS. In terms of MIHFP, variables such as marital status (not married but living with partner, P=0.004 and going steady (in a relationship), P=0.004, education (no education, P=<0.001 and secondary, P=0.001) and household structure (male centered, P= 0.031 and nuclear, P=0.011) were significant at 5% (Table 24).

Table 2: Association between HFIP, HDDS MIHFP and socio-household characteristics

	HFIP		**HDDS**		**MIHFP**	
	Pearson's Chi-Square					
Sex:	**Value**	**P-Value**	**Value**	**P-Value**	**Value**	**P-Value**
Male	5.239	**0.022**	0.066	0.797	1.345	0.246
Female	*Reference*					
Marital Status:						
Married (living with spouse)	5.552	**0.018**	3.732	0.053	1.959	0.162
Married (not living with spouse)	0.091	0.763	0.390	0.532	0.986	0.321
Not married (living with partner)	2.378	0.123	0.864	0.353	8.406	**0.004**
Going steady (in a relationship)	0.559	0.445	2.913	0.088	8.291	**0.004**
Single (not in a relationship)	0.330	0.566	4.536	**0.033**	0.791	0.374
Divorced separated	0.018	0.894	0.040	0.842	0.691	0.406
Widower/Widow	*Reference*					
Education:						
None	2.005	0.157	5.441	**0.020**	15.489	**<0.001**
Primary	35.588	**<0.001**	3.099	0.078	0.155	0.694
Secondary	33.010	**<0.001**	5.105	**0.024**	7.255	**0.007**
Tertiary	*Reference*					
Work status:						
Working full-time	6.958	**0.008**	9.454	**0.002**	3.723	0.054
Working part-time/casual work	0.249	0.618	1.000	0.317	0.054	0.815
Not working - looking	5.445	**0.020**	7.101	**0.008**	3.117	0.077
Not working - not looking	*Reference*					
Tenure Status:						
Owner/Family	0.428	0.513	3.141	**0.046**	0.830	0.362
Tenant/Lodger	0.047	0.828	2.348	0.125	0.613	0.434
Tied accommodation	*Reference*					
Household Structure:						
Female centered	0.200	0.655	0.207	0.649	0.056	0.814
Male centered	0.327	0.567	0.231	0.631	4.632	**0.031**
Nuclear	0.685	0.408	0.280	0.596	0.317	**0.011**
Extended	1.593	0.672	0.327	0.877	-1.942	0.573
Under 18 headed household	*Reference*					
No piped water - private	4.372	**0.037**	6.318	**0.012**	0.205	0.651
Piped Water - Private	*Reference*					
No electricity	1.927	0.165	0.037	0.847	1.323	0.250
Electricity available	*Reference*					
No toilet	0.250	0.617	6.332	**0.012**	0.762	0.383
Toilet available	*Reference*					

1.3.2. Logistic and Poisson Regression Models: HFIP, HDDS and MIHFP

Table 25 and Table 26 provides a summary of the logistic regression and Poisson regression models. Predictable variables such as education and accessibility to water influenced the food security level of a household. The household Dietary Diversity is affected by the educational level of the head of Household as well as accessibility to amenities such as electricity and toilet facilities. Months on Inadequate food Provisioning (MIHFP) is another indicator to measure the food security of a household. MIHFP was influenced by various factors including marital status (specifically by those that are not married but living with partners and those that are going steady in a relationship), Educational level, tenure status and the household structure. All these predictors were significant at 5% level (p-value<0.05).

Table 3: Modelling of FIP, HDDS and MIHFP

	Logistic Regression Model (HFIP)		**Poisson Regression Model (HDDS)**		**Poisson Regression Model (MIHFP)**	
Coefficients	**Std. Err**	**P-Value**	**Std. Err**	**P-Value**	**Std. Err**	**P-Value**
(Intercept)	623.74	0.978	0.617	0.004	1.942	0.011
Sex:					0.030	**0.002**
Male	0.342	0.087	0.064	0.812		
Female	*Reference*					
Marital Status:					-0.054	0.809
Married (living with spouse)	1.195	0.152	0.215	0.173		
Married (not living with spouse)	1.318	0.771	0.238	0.188	-0.490	0.090
Not married (living with partner)	1.200	0.348	0.216	0.353	0.462	**0.030**
Going steady (in a relationship)	1.194	0.579	0.214	0.269	-0.545	**0.013**
Single (not in a relationship)	1.180	0.411	0.211	0.517	0.012	0.952
Divorced / separated	1.755	0.677	0.349	0.924	0.248	0.466
Widower/Widow	*Reference*					
Education:						
None	0.905	**0.001**	0.141	**0.001**	0.798	**0.006**
Primary	0.805	**0.001**	0.110	**0.002**	0.388	0.172
Secondary	0.815	0.072	0.113	0.045	-0.002	0.928
Tertiary	*Reference*					

Table 4: Modelling of FIP, HDDS and MIHFP ….cont.

	Logistic Regression Model (HFIP)	**Poisson Regression Model**	**Poisson Regression Model**

			(HDDS)		**(MIHFP)**	
Employment Working full-time	0.496	0.271	0.092	0.269	-0.223	0.090
Working part-time/casual work	0.513	0.210	0.110	0.804	-0.154	0.237
Not working - looking	0.506	0.964	0.095	0.353	-0.006	0.624
Not working - not looking	*Reference*					
Tenure Status: Owner/Family	623.733	0.979	0.422	0.813	1.340	**0.049**
Tenant/Lodger	623.734	0.977	0.431	0.945	1.244	0.074
Tied accommodation	*Reference*					
Household Structure: Female centered	1.588	0.438	0.325	0.519	-1.942	**0.011**
Male centered	1.585	0.634	0.325	0.697	-1.942	**0.011**
Nuclear	1.595	0.777	0.326	0.680	-1.942	**0.011**
Extended	1.593	0.672	0.327	0.877	-1.942	**0.011**
Under 18 headed household	*Reference*					
No piped water - private	0.426	**0.035**	0.083	0.305	-0.159	0.157
Piped Water - private	*Reference*					
No electricity	0.620	0.119	0.141	**0.076**	-0.351	0.089
Electricity available	*Reference*					
No toilet	1.317	0.933	0.207	**0.019**	-0.141	0.807
Toilet available	*Reference*					
AIC	**468.81**		**1781.6**		**1752.019**	
BIC	**559.1909**		**1871.9515**		**1838.909**	

1.3.3. Joint Modelling of Household Food Insecurity Prevalence (HFIP) and Household Dietary Diversity Score (HDDS)

Generalized Joint Regression model was conducted, and copula estimates were performed using binary-binary margins (probit) to estimate several copulas with endogenous treatment, where the bivariate distributions are chosen so that the dependence is allowed. This is mainly because the models based on the Gaussian and Frank Copulas suggest that the dependence between the outcomes is positive, thus implying copulas which allow for negative association when the data do not support this will be misleading (Marra & Radice, 2017). The AICs were used to determine the best fitted model. According to Table 27, all the models are more or less equally good as their AICs did not differ much, however the Frank copula had the least AIC.

Table 5: AICs for copula models: FIP and HDDS

Family	**Df**	**AIC**
Bivariate Normal	65	2288.355
Frank	65	**2287.296**

Rotational Clayton	65	2288.158
Gumbel	65	2288.349

Table 28 shows that all the predictor variable estimates obtained for the Frank copula were not significant at 5%, thus indicating no existence of any positive association between the unstructured terms of the model equations.

Table 6: Estimates for Frank copula model (Margins: Bernoulli, Bernoulli)

Coefficients	**Estimate**		**Std. err**		**P-Value**	
	HFIP	**HDDS**	**HFIP**	**HDDS**	**HFIP**	**HDDS**
(Intercept)	7.587	2.408	7.144	7144.461	1.000	0.999
Sex:						
Male	-3.671	-4.211	5.435	5434.891	1.000	1.000
Female						
Marital Status:						
Married (Living with spouse)	0.7376	2.495	3.043	5434.891	1.000	1.000
Married not (living with	0.322	2.732	3.043	3042.879	1.000	1.000
spouse)	0.381	1.561	3.043	3042.879	1.000	1.000
Not married (living with	0.312	1.932	3.043	3042.879	1.000	1.000
partner)	0.120	9.466	3.043	3042.879	1.000	1.000

Coefficients	Estimate		Std. err		P-Value	
	HFIP	**HDDS**	**HFIP**	**HDDS**	**HFIP**	**HDDS**
Going steady (in a relationship)	3.043	-1.206	3.043	3042.879	1.000	1.000
Single (not in a relationship) Divorced / separated Widower/Widow						
Education:						
None	-0.328	-1.724	3.972	3971.539	1.000	1.000
Primary	-1.399	-1.466	3.972	3971.539	1.000	1.000
Secondary	-1.202	-1.350	3.972	3971.539	1.000	1.000
Tertiary	*Reference*					
Working full – time	0.421	-2.964	3.972	3971.539	1.000	1.000
Working part-time/Casual	0.211	-4.213	3.972	3971.539	1.000	1.000
Not working - looking	0.260	-4.865	3.972	3971.539	1.000	1.000
Not working- not looking	*Reference*					
Tenure Status:						
Owner/Family	-6.031	8.254	4.537	4543.257	1.000	0.999
Tenant/Lodger	-6.225	-4.681	4.537	4543.257	1.000	0.999
Tied accommodation	*Reference*					
Household Structure:						
Female centered	0.301	7.240	3.575	3574.794	1.000	1.000
Male centered	0.027	-9.714	3.575	3574.794	0.999	1.000
Nuclear	-0.137	-2.192	3.575	3574.794	0.999	1.000
Extended	-0.005	-8.334	3.575	3574.794	1.000	1.000
Under 18 headed household's	*Reference*					
No piped water- private	-0.249	-2.351	5.435	5434.892	1.000	1.000
Piped Water – Private	*Reference*					
No Electricity	-0.694	1.628	5.435	5434.892	1.000	1.000
Electricity available	*Reference*					
No Toilet	-0.050	1.919	5.435	5434.892	1.000	1.000
Toilet Available	*Reference*					
AIC: 2287.296; BIC; 2542.719, Theta= 0.419(-0.342, 1.05), Tau = 0.0464 (-0.0379, 0.116)						

5.3.4. Joint modelling of Household Food Insecurity Prevalence and Months of Inadequate Household Food Provision (MIHFP)

The joint modelling of food insecurity prevalence and months of inadequate food provisioning using different copula models shows that the Bivariate Normal copula is the preferred model (lowest AIC).

Table 7: AICs for copula models: HFIP and MIHFP (margins = Bernoulli, Poisson)

Family	Df	AIC
Bivariate Normal	65	**2072.708**
Frank	65	2074.352

Gumbel	65	2108.451

The estimates for the Bivariate Normal copula independent variables proved to have no positive association at 0.05 significant level (app *P*-values>0.005, and Theta (-0.32(-0.417, -0.219).

5.3.5. Sample Selection and Partial Observability: Food Insecurity Prevalence and Dietary Diversity Score

Sample selection and Partial observability were conducted to observe specific household responses. Table 30 shows that the determinants variables were not significant at 5%, suggesting that there is no statistically significant relationship between HFIP, HDDS and the independent variables. Sex of head of household was found to have a statistically significant relationship with household food insecurity prevalence (P-value<0.05) (Table 31).

Table 8: Sample selection: Food Insecurity Prevalence (HFIP) and Household Dietary Diversity Score (HDDS) (margins= Bernoulli, Bernoulli)

	Estimate		**Std. err**		**P Values**	
Coefficients	**HFIP**	**HDDS**	**FIP**	**HDDS**	**FIP**	**HDDSx**
(Intercept)	-7.808	-1.473	7082.429	8192.000	0.999	1.000
Sex:	-0.158	-14.423	0.211	8192.000	0.454	0.999
Male Female	*Reference*					
Marital Status: Married (Living with spouse)	-0.734	-14.723	3148.404	8192.000	1.000	0.999
Married not (living with	-0.790	-14.632	3148.404	8192.000	1.000	0.999
spouse)	-0.693	-14.641	3148.404	8192.000	1.000	0.999
Not married (living with	-1.170	34.629	3148.404	8192.000	1.000	0.999
partner)	-6.934	-14.638	3148.404	8192.000	1.000	0.999
Going steady (in a	-1.333	-14.692	3148.404	8192.000	1.000	0.997
relationship) Single (not in a relationship, Divorced / separated Widower/Widow)	*Reference*					

Education:						
None	-6.675	-14.574	2524.724	8192.000	0.998	0.999
Primary	-6.295	-14.541	2524.724	8192.000	0.998	0.999
Secondary	-5.995	-14.206	2524.724	8192.000	0.998	0.999
Tertiary	*Reference*					
Working full – time	-6.227	-14.289	2524.724	8192.000	0.998	0.999
Working part-time/Casual	-6.409	-14.405	2524.724	8192.000	0.998	0.999
Not working - looking	-5.520	-14.330	2524.724	8192.000	0.998	0.999
Not working- not looking	*Reference*					
Tenure Status:						
Owner/Family	-0.645	-14.553	5205.962	8192.000	1.000	0.999
Tenant/Lodger	-5.632	40.998	7740.02	8192.000	0.999	0.999
Tied accommodation	*Reference*					
Female centered	6.173	8.795	3918.062	8192.000	0.999	0.999
Male centered	5.990	8.860	3918.062	8192.000	0.999	0.999
Nuclear	5.945	8.793	3918.062	8192.000	0.999	0.999
Extended	-0.777	-15.056	6819.300	8192.000	1.000	0.999
Under 18 headed household's	*Reference*					
No piped water- private	-42.093	-18.209	5414.556	8192.000	0.994	0.999
Piped Water – Private	*Reference*					
No Electricity	68.147	8.714	5414.556	8192.000	0.990	0.999
Electricity available	*Reference*					
No Toilet	-3.959	69.296	6817.927	8192.000	1.000	0.999
Toilet Available	*Reference*					
AIC: 473.073; BIC; 27.626, Theta= 100(87, 100), Tau = 0.961 (0.955, 0.961)						

Table 9: Partial Observability: HFIP and HDDS (margins= Bernoulli, Bernoulli)

	Estimate		**Std. err**		**P-Values**	
Coefficients	**HFIP**	**HDDS**	**HFIP**	**HDDS**	**HFIP**	**HDDS**
(Intercept)	0.044	87.082	66.736	59.9021	0.999	0.146
Sex:						
Male	0.484	-24.148	0.210	20.396	**0.021**	0.236
Female	*Reference*					
Marital Status:						
Married (Living with spouse)	-0.899	-17.342	32.724	20.338	0.978	0.395
Married not (living with spouse)	-0.976	-18.269	32.724	20.397	0.976	0.369
Not married (living with partner)	-0.714	-27.820	32.724	20.344	0.982	0.171
Going steady (in a relationship)	-1.340	-27.326	32.724	20.333	0.967	0.179
Single (not in a relationship)	-5.858	-28.334	32.724	8192.000	0.962	0.997
Divorced / separated	-1.431	-19.456	32.724	20.346	0.965	0.339
Widower/Widow	*Reference*					
Education:						
None	-5.588	-11.029	24.434	20.304	0.819	0.587
Primary	-5.063	-12.938	24.434	20.304	0.836	0.523
Secondary	-4.557	-11.750	24.434	20.304	0.852	0.563
Tertiary	*Reference*					
Working full – time	-5.181	-6.048	24.455	20.343	0.832	0.766

Working part-time/Casual	-5.320	-4.986	24.454	20.342	0.828	0.806
Not working - looking	-4.725	-7.736	24.454	20.206	0.847	0.703
Not working- not looking	*Reference*					
Tenure Status:						
Owner/Family	-0.605	-2.319	55.287	24.218	0.991	0.924
Tenant/Lodger	-4.201	-4.795	81.594	8192.000	0.959	0.999
Tied accommodation	*Reference*					
Household Structure:						
Female centered	5.030	-23.047	53.353	58.850	0.925	0.695
Male centered	4.676	-20.748	53.354	58.881	0.930	0.724
Nuclear	4.971	-20.148	53.354	58.845	0.926	0.732
Extended	-0.498	-18.254	91.977	939.496	0.994	0.984
Under 18 headed household's	*Reference*					
No piped water- private	0.019	0.600	53.585	51.438	0.999	0.990
Piped Water – Private						
	Reference					
No Electricity	15.060	0.927	43.674	62.046	0.730	0.988
Electricity available	*Reference*					
No Toilet	-2.810	4.488	102.710	940.904	0.978	0.996
Toilet Available	*Reference*					
AIC: 3093.055; BIC: 3347.609, Theta= 0.272 (0.258, 0.289), Tau = 0.175 (0.166, 0.187)						

5.3.6. Sample Selection and Partial Observability: Food Insecurity Prevalence and Months of Inadequate Food Provision

Table 32 and Table 33 shows results from sample selection and partial observability. Apart from Sex, all other determinants variables were not significant at 5%, suggesting that there is no statistically significant relationship.

Table 10: Sample Selection: FIP and MIHFP (margins= Bernoulli, Poisson)

	Estimate		**Std. err**		**P Values**	
Coefficients	**HFIP**	**MIHFP**	**HFIP**	**MIHFP**	**HFIP**	**MIHFP**
(Intercept)	-5.728	9.692	6.639	7.213	0.993	0.999
Male	-1.014	-2.485	2.738	9.693	0.711	**0.010**
Female	*Reference*					
Married (Living with spouse)	-1.690	-1.157	1.940	3.580	0.999	0.999
Married not (living with spouse)	-1.735	-3.383	1.940	3.580	0.999	0.999
Not married (living with partner)	-1.552	4.724	1.940	3.580	0.999	0.999
Going steady (in a relationship)	-2.151	3.294	1.940	3.580	0.999	0.999
Single (not in a relationship)	-6.446	-1.786	1.940	8.192	0.999	1.000
Divorced / separated	-2.211	6.423	1.940		0.997	0.997
Widower/Widow	*Reference*					
Education:						
None	-1.970	2.266	1.793	3.979	0.999	0.999
Primary	-1.704	5.795	1.793	3.979	0.999	0.999

Secondary	-7.878	7.974	1.793	3.979	0.999	0.999
Tertiary	*Reference*					
Working full – time	-8.081	3.578	1.721	3.979	0.999	0.999
Working part-time/Casual	-1.029	-3.057	1.721	3.979	0.999	0.999
Not working - looking	-3.480	5.031	1.721	3.979	0.999	0.999
Not working- not looking	*Reference*					
Owner/Family	7.217	3.053	2.648	5.457	0.978	0.999
Tenant/Lodger	6.849	1.441	2.565	8.192	0.978	0.999
Tied accommodation	*Reference*					
Female centered	6.864	2.179	2.565	3.979	0.978	0.999
Male centered	6.845	5.120	2.565	3.979	0.978	0.999
Nuclear	6.817	-2.316	2.565	3.979	0.978	0.999
Extended	6.277	3.609	2.664	8.192	0.981	0.999
Under 18 headed household's	*Reference*					
No piped water- private	1.725	6.291	3.792	5.457	1.000	0.999
Piped Water – Private						
No Electricity	-7.443	2.641	3.932	8.192	0.984	0.999
Electricity available	*Reference*					
No Toilet	-6.591	1.611	2.121	5.457	0.999	0.999
Toilet Available	*Reference*					
AIC: 678.045, BIC: 985.021, Theta= 0.535 (0.382, 0.65), Tau = 0.359 (0.25, 0.45)						

Table 11: Partial Observability: FIP and MIHFP (margins= Bernoulli, Poisson)

	Estimate		**Std. err**		**P-Values**	
Coefficients	**FIP**	**MIHFP**	**FIP**	**MIHFP**	**FIP**	**MIHFP**
(Intercept)	0.056	87.082	77.736	49.9021	0.999	0.146
Sex:						
Male	0.556	-24.148	0.223	20.396	0.432	0.236
Female	*Reference*					
Marital Status:						
Married (Living with spouse)	-0.899	-17.342	32.724	20.338	0.978	0.395
Married not (living with spouse)	-0.976	-18.269	32.724	20.397	0.976	0.369
Not married (living with partner)	-0.714	-27.820	32.724	20.344	0.982	0.171
Going steady (in a relationship)	-1.340	-27.326	32.724	20.333	0.967	0.179
Single (not in a relationship)	-5.858	-28.334	32.724	8192.00	0.962	0.997
Divorced / separated	-1.431	-19.456	32.724	20.346	0.965	0.339
Widower/Widow	*Reference*					
Education:						
None	-5.588	-11.029	24.434	20.304	0.819	0.587
Primary	-5.063	-12.938	24.434	20.304	0.836	0.523
Secondary	-4.557	-11.750	24.434	20.304	0.852	0.563
Tertiary	*Reference*					
Working full – time	-5.181	-6.048	24.455	20.343	0.832	0.766
Working part-time/Casual	-5.320	-4.986	24.454	20.342	0.828	0.806
Not working - looking	-4.725	-7.736	24.454	20.206	0.847	0.703
Not working- not looking	*Reference*					
Tenure Status:						
Owner/Family	-0.605	-2.319	55.287	24.218	0.991	0.924
Tenant/Lodger	-4.201	-4.795	81.594	8192.000	0.959	0.999
Tied accommodation						

	Reference					
Female centered	5.030	-23.047	53.353	58.850	0.925	0.695
Male centered	4.676	-20.748	53.354	58.881	0.930	0.724
Nuclear	4.971	-20.148	53.354	58.845	0.926	0.732
Extended	-0.498	-18.254	91.977	939.496	0.994	0.984
Under 18 headed household's	*Reference*					
No piped water- private	0.019	0.600	53.585	51.438	0.999	0.990
Piped Water – Private	*Reference*					
No Electricity	15.060	0.927	43.674	62.046	0.730	0.988
Electricity available	*Reference*					
No Toilet	-2.810	4.488	102.710	940.904	0.978	0.996
Toilet Available	*Reference*					
AIC: 4031.044, BIC: 4712.055, Theta= -0.32(-0.417, -0.219), Tau = -0.207 (-0.274, -0.14)						

1.4. Discussion

Various food security measurements exist to measure the extent of food insecurity both at individual and household level. This chapter particularly applied bivariate joint regression models using copulas (Bivariate Normal, Frank, Rotational Clayton, Gumbel) to model food insecurity prevalence, Household Dietary Diversity and Months of Inadequate Food provisioning. The Bivariate Poisson models are appropriate for modeling paired count data exhibiting correlation and require joint estimation (Karlis & Ntzoufras, 2005). Sample selection and partial observability are errors that arise during the collection of data. For example, the implementation of strategic models is often made impossible by the outcome-rather than actor-specific structure of available data: while there are data on the aggregated outcomes of an interaction, there will be no record of each player's actions at each of the interaction stage.

Food insecurity is a major problem in the country. About 63% of the population are food insecure. This means, this proportion of the country does not have physical and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy lifestyle at all times. Food security puts an emphasis on all the 4 dimensions to be met: Food availability, Food accessibility, Food utilization and Food stability (FAO, The State of Food Security in the world, 2002). Dietary diversity is very critical in measuring food security. This means that most households consume a monotonous diet that lacks variety of diets. Additionally, food accessibility is defined by the availability of resources to the households throughout the month. According to Nickanor (2014), most households did not have enough resources for food in the months of January. January Precedes the month of December, that is mostly referred to as the Festival Month. Most households have utilized their savings and bonuses on these social gatherings. Apart from that, during the month of January, households further have to capitalize on other mandatory expenditures such as school uniforms and school fees and rural households investing in ploughing/ farming activities, as it is a rainy season. This leaves most households with little to spend on foods (Nickanor, 2014).

The results from the logistic and Poisson logistic regression models indicated that educational level of the head of household and accessibility to water influenced the food security level of a household. Other factors that influenced food security included marital status, tenure status and the household structures (female centered, male centered, Nuclear, Extended, under-18 headed households). Education improves food security more directly in two ways; firstly, by improving skills and income generating potentials, secondly, through greater employability opportunities and increased incomes from better employment (Ajieroh, 2009). The household structure also affects food insecurity of that house. Larger households tend to have a negative impact on individual caloric availability. The size of a household has a potential to directly affects is food insecurity level through its influence on consumption pattern (Nickanor, 2014).

This study utilized copula functions to jointly estimate the variables. A joint model provides improved results on modelling associations between two outcomes, rather than modelling them separately. It significantly improves the log-loss and absolute residuals of cross-validation predictions (Broatch & Karl, 2017). Frank copula fit the data better to estimate the relationship between food insecurity prevalence and household dietary diversity score. Bivariate Normal copula was the best to model an association between food insecurity prevalence and months of inadequate food provisioning. Sample selection and partial observability were conducted to determine the relationship between Food Insecurity Prevalence and Dietary diversity as well as between food insecurity prevalence and months of Inadequate household food provisioning. The models found that, apart from sex, all other social-demographic variables were not significant at 5% indicating a non-relationship between the exposure and outcome variables.

1.5. Conclusions

Generalized Joint Regression Models are used to model jointly binary outcomes. The aim of this study was to jointly model food insecurity indicators with application to FIP, HDDS and MIHFP. Copula approaches relate an arbitrary joint distribution to its corresponding univariate marginal distributions. Copulas were applied in this study to investigate the relationship between household food insecurity prevalence (HFIP) (1. Food Secure 2. Food Insecure); household dietary diversity score (1. Low diversity 2. High diversity) and Months of Inadequate Household Food Provisioning (MIHFP) (Seasonal, persistent). Food insecurity in Namibia is high and less varied with a monotonous diet. Households were further found to be more food insecure during the month of January. Measurement errors were accounted for by the modelling of sample selection and partial observability.

The Copula approach is defined as a useful method for deriving joint distributions. The approach relates an arbitrary joint distribution to its corresponding univariate marginal distributions via Copula. Specifically, five (5) Copula families namely the Bivariate normal, Frank, Survival, Clayton, Gumbel and the Survival Gumble were used in this analysis. The Frank Copula was identified to fit the data between FIP and HDDS better while the Bivariate normal better fitted the data between FIP and MIHFP. Sample selection and Partial Observability were conducted to observe specific responses between the three indicators. The socio-demographic variables were all not significant at 5% indicating a non-relationship between the exposure and outcome variables.

1.6. Acknowledgements

The authors had support from the Developing Excellence in Leadership, Training and Science (DELTAS) Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107754/Z/15/Z- DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme] and the UK government.