MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Bayesian Inference for a 1-Factor Copula Model

Ban Kheng Tan, Anastasios Panagiotelis, George Athanasopoulos

March 2017

Working Paper 06/17

Bayesian Inference for a 1-Factor Copula Model

Ban Kheng Tan
Department of Econometrics and Business Statistics
Monash University, 900 Dandenong Road
Caulfield East, Victoria 3145, Australia.
Email: ban.tan@monash.edu

Anastasios Panagiotelis
Department of Econometrics and Business Statistics
Monash University, 900 Dandenong Road
Caulfield East, Victoria 3145, Australia.
Email: Anastasios.Panagiotelis@monash.edu

George Athanasopoulos
Department of Econometrics and Business Statistics
Monash University, 900 Dandenong Road
Caulfield East, Victoria 3145, Australia.
Email: George.Athanasopoulos@monash.edu

24 March 2017

JEL classification: C11, C15, C32, C58, C63

Bayesian Inference for a 1-Factor Copula Model

Abstract

We develop efficient Bayesian inference for the 1-factor copula model with two significant contributions over classical inference. First, our approach leads to straightforward inference on the latent factor since iterates of the latent factor are generated as a by-product in the proposed Markov chain Monte Carlo algorithm. In contrast, there is no known classical approach for inference on the latents. Second, by developing a reversible jump Markov chain Monte Carlo scheme, we are able to select or average over factor copula specifications that are constructed from a large set of candidate parametric bivariate copula building blocks. Our approach can accommodate margins that are discrete, continuous or a combination of both. Through extensive simulations multiple schemes are compared on the basis of computational and Monte Carlo efficiency. The preferred schemes provide reliable inference on all parameters including the latent factor and model space. The potential of the proposed methodology is highlighted in an empirical study of ten binary variables measuring the multidimensional nature of poverty collected for 11463 East Timorese households. We construct a poverty index using estimates of the latent factor. Compared to a classical analysis, our method yields better out-of-sample fit and uncovers a variety of flexible relationships between the latent measure and observed variables by averaging over a diverse set of copulas.

Keywords: Model Averaging; Reversible jump MCMC; Vine Copulas; Dimension Reduction; Multidimensional poverty index.

1 INTRODUCTION

In recent years, high dimensional datasets have been increasingly used and analyzed in statistics particularly in business and economics. In some cases these involve multiple measurements collected over time, for example the returns of all stocks in the S&P 500 and the large panel of disaggregate US macroeconomic series first introduced by Stock and Watson (2002). In other instances, multivariate data may be collected across individuals or households, for example surveys with a large number of questions, or the data considered in our application which are multiple measures of deprivation collected across households in East Timor. In factor analysis the dependence structure between variables is based on a small number of latent unobservable factors, thus providing an elegant solution to dealing with the high dimensionality of such datasets. In applications to economics, finance, marketing and psychometrics, the latent factors often have meaningful interpretations and are an object of inference themselves. For example, a latent factor obtained from macroeconomic data may provide a coincident index for the business cycle (Geweke, 1977; Sargent, Sims et al., 1977; Stock & Watson, 1989). As shall be seen in our application, a latent factor can also be used as an overall measure of poverty that takes the multidimensional nature of well being into account.

The classical factor model is defined in Anderson and Rubin (1956) as

$$y_i = \Lambda f_i + \xi_i, \quad \xi_i \sim N(0, \Sigma), \quad \forall i = 1, \dots, n \tag{1.1}$$

where y_i is an $m \times 1$ response vector, f_i is an $r \times 1$ vector of common factors, Λ is a matrix of factor loadings, and Σ is assumed to be diagonal together with normality which implies that the components of each ξ_i are independent. We index observations by the subscript i rather than subscript t, since our focus and indeed our application involve the case where multiple measures are made across individuals, rather than time.

The assumptions of classical factor analysis can be relaxed in a number of ways, for example principal components provide consistent estimates of the factors even if the data are not normally distributed. Also, the literature on approximate factor models (Chamberlain & Rothschild, 1982) and dynamic factor models (Forni, Hallin, Lippi & Reichlin, 2000; Stock & Watson, 2002) show that principal components still consistently estimate factors even when idiosyncratic terms are correlated as long as this correlation is bounded in the limit as $n, m \to \infty$.

However, these approaches all assume a linear relationship between the response variables and the factors.

A recent extension to classical factor analysis is the factor copula model proposed by Krupskii and Joe (2013). This approach follows classical factor analysis by assuming that the responses are independent when conditioned on r latent variables. However, rather than assuming Gaussianity, dependence between each response and the latent factor is modelled using a copula. The r-factor copula is a special case of a C-vine copula, such that the central nodes of the first r trees are latent variables and all pair copulas on trees $r+1$ onward are independence copulas (see Joe, 2014). This leads to a simple form of the density for $\boldsymbol{y}$ in terms of bivariate copulas. For example, the density $\boldsymbol{y}$ in the 1-factor copula model is given by,

$$p(\boldsymbol{y}) = \left\{ \int_0^1 \left[\prod_{j \in \mathcal{F}_\mathcal{C}} c_j(P_j(y_j), f) \times \prod_{j \in \mathcal{F}_\mathcal{D}} C_j(P_j(y_j)|f) - C_j(P_j(y_j - 1)|f) \right] df \right\} \times \prod_{j \in \mathcal{F}_\mathcal{C}} p_j(y_j), \qquad (1.2)$$

where $c_j(.,.)$ is the copula density of the j^{th} continuous response and latent factor, $C_j(.|.)$ is the conditional copula distribution of the j^{th} discrete response and latent factor, $\mathcal{F}_\mathcal{C}$ is the set of continuous responses, $\mathcal{F}_\mathcal{D}$ is the set of discrete responses, $P_j(.)$ and $p_j(.)$ are respectively the distribution and density (mass) function of the j^{th} continuous (discrete) response. Note here that $\mathcal{F}_\mathcal{C}$ or $\mathcal{F}_\mathcal{D}$ can be an empty set.

Another class of factor copula model which appears in the literature is the model proposed by Oh and Patton (2012) which is an extension of Hull and White (2004). Oh and Patton (2012) create a copula using a similar structure to (1.1) where $\boldsymbol{f}_i$ and ξ_i follow non-Gaussian distributions. In their approach the $\boldsymbol{y}_i$ in (1.1) is not the response variable but an additional vector of m latent variables, and the factor copula is defined as the copula of these latent variables.

The factor copula model we consider builds on Krupskii and Joe (2013) which has a number of advantages relative to classical analysis. First, the margins need not be Gaussian or even continuous, different models can be used for each margin, or they can be modelled non-parametrically. Second, the use of non-Gaussian copulas allows for non-linear dependence features such as tail dependence, which is common in many types of data including financial returns data. Third, the model is parsimonious, the number of parameters is of $O(m)$ making it well suited to high dimensional problems. Fourth, and crucially, the factor copula allows for a non-linear relationship between each response and the factor.

This paper addresses two issues yet to be resolved in the factor copula framework: the first is inference on the latent factor and the second is model selection. Both problems are addressed by developing Bayesian Markov chain Monte Carlo (MCMC) inference for the 1-factor copula model. Under both frequentist and Bayesian approaches the latent factor in (1.2) must be integrated out numerically. However, in contrast to frequentist approach, the Bayesian framework achieves this by simulating from the joint posterior of the latent variables and parameters. The draws of the latent variables can then be used to conduct inference on the factor in the same way as the other parameters. The specific proposal distributions used in the MCMC scheme were carefully chosen on the basis of extensive simulation studies and our scheme provides efficient inference on the latent factors.

For model selection, each copula in (1.2) can be any element of a set of b bivariate parametric copula families, leading to m^b models. Since m^b can be quite large, it is infeasible to enumerate the BIC, DIC or marginal likelihood for every model. Therefore we develop a reversible jump MCMC scheme (Green, 1995), which is shown in simulation studies to efficiently traverse this large model space. Overall, our approach is formal, automatic (and therefore feasible in high dimensions) while the Bayesian paradigm also produces natural weights for model averaging. As such, it compares favourably to the model selection heuristics recommended by Krupskii and Joe (2013) and Nikoloulopoulos and Joe (2015) that are based on diagnostic methods such as the bivariate normal scores plot.

The rest of the paper is organized as follows. Section 2 outlines the formulation of Bayesian parameter estimation for a 1-factor copula model where all the choices of bivariate copula families are prespecified/known. The most efficient MCMC sampling scheme between three different sampling schemes proposed is identified based on a simulation study. Formulation of Bayesian model estimation along with the parameter estimates using RJMCMC will be addressed in Section 4. Section 5 demonstrates the performance of our proposed sampling algorithm for model estimation while Section 6 contains an example of the application of 1-factor copula model to East Timor poverty data. Section 7 concludes with a discussion of possible future research.

2 BAYESIAN MODEL

In this section, we formulate the prior, augmented likelihood and posterior for a 1-factor copula model. We assume that all bivariate copulas come from parametric families that are known, although we will explicitly allow for model uncertainty in Section 4.

2.1 The Augmented Likelihood

Let $Y = \{y_{ij}\}$, where y_{ij} is the response for the i^{th} observation of the j^{th} variable, $\boldsymbol{f} = (f_1, \ldots, f_n)'$ be the common factor, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_m')'$ be a vector of parameters associated with marginal distributions and $\boldsymbol{\tau} = (\boldsymbol{\tau}_1', \ldots, \boldsymbol{\tau}_m')'$ be a vector of parameters associated with the bivariate copulas. The augmented likelihood is given by

$$p(Y, \boldsymbol{f}|\boldsymbol{\theta}, \boldsymbol{\tau}) = \prod_{i=1}^{n} \Bigg\{ \prod_{j \in \mathcal{F}_C} c_j(P_{\boldsymbol{\theta}_j}(y_{ij}), P(f_i); \tau_j) \times \prod_{j \in \mathcal{F}_D} \Big\{ C_j(P_{\boldsymbol{\theta}_j}(y_{ij}) | P(f_i); \tau_j) - C_j(P_{\boldsymbol{\theta}_j}(y_{ij} - 1) | P(f_i); \tau_j) \Big\} \times \prod_{j \in \mathcal{F}_C} p_{\boldsymbol{\theta}_j}(y_{ij}) \times p(f_i) \Bigg\}, \tag{2.1}$$

where P and p are used as generic notation for a distribution and density respectively, $P_{\boldsymbol{\theta}_j}$ ($p_{\boldsymbol{\theta}_j}$) denotes that a distribution (density) depends on marginal parameters $\boldsymbol{\theta}_j$ and $c_j(a, b; d)$ denotes a copula density corresponding to the j^{th} margin with arguments a and b and parameter d and $C_j(a|b; d) = \int_0^a c_j(u, b; d)du$. Note that by the invariance property of likelihood, we are able to reparameterize one parameter copula families in terms of the Kendall's τ by exploiting the bijective relationship $\tau_j = G_j^{-1}(\rho_j)$ between τ and the usual parameterization ρ. For example, for a Gaussian copula $\tau = \frac{2}{\pi} arcsin(\rho)$ where ρ is the correlation in a Gaussian copula. Expressing non-nested copulas in terms of this common parameterization simplifies our specification of priors in Section 2.3 and facilitates the development of our algorithm in Section 4 which jumps between non-nested models. Two parameter copula families can also be used and are discussed explicitly in Section 4.

2.2 Model Identification

For an r-factor model, r^2 restrictions are imposed on factor loadings, variances and/or covariances to identify the model. There are large number of possible ways to place the restrictions (see Asparouhov & Muthén, 2009, and references therein). Even in the classical 1-factor model, the mean and variance of the latent factor are unidentified and are usually normalized to zero and one respectively. In the case of a 1-factor copula model a similar restriction must be applied, specifically the marginal distribution of the factors must be fixed. We follow Krupskii and Joe (2013) in assuming $f_i \sim U(0,1)$ since this facilitates the interpretation of the latent factor in terms of quantiles, for example if an individual has a factor score of 0.9 then their factor score is higher than 90% of the observations. We note that a $f_i \sim N(0,1)$ could also be used to identify the factor. Identification is carried out naturally in the Bayesian framework; the restriction is equivalent to the prior on the latent factor.

Another well understood identification issue in the classical factor model, is that the latent factors and loadings are only identified up to a sign. A similar issue can arise for the factor copula model when the bivariate copula families are reflection symmetric around 0.5, i.e $c_j(P(x_j), f_i; \tau_j) = c_j(P(x_j), 1-f_i; -\tau_j)$. In this case a model with parameters $(\boldsymbol{\tau}, \boldsymbol{f})$ will be observationally equivalent to a model with $\left(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{\tau}}\right)$ when $\tilde{f}_i = 1 - f_i$ for $i = 1, \ldots n$ and $\tilde{\tau}_j = -\tau_j$ for $j = 1, \ldots m$. In order to identify the latent factor, we impose identifying restriction(s) on the sign of τ_j based on prior knowledge. For example in our application in Section 6, we constrain τ_j corresponding to one variable to be positive so that increasing levels of the latent factor are associated with higher levels poverty rather than higher levels of welfare. This restriction also resolves issues associated with sampling from a multimodal posterior and in the context of model selection reduces the size of the model subspace. If the identifying restriction is not strong, the Markov chains could still jump between multiple modes, particularly when we consider the more complicated issue of model selection. This can reduce the efficiency of MCMC schemes and at times lead to a failure to converge. In these cases we recommend imposing at least one strong identifying restriction (for example $\tau_j > 0.5$) for some j. In cases where imposing such a single strong restriction cannot be justified, imposing several weaker restrictions can be used instead.

2.3 The Priors

In this study, we have adopted flat priors across all the parameters where possible. In the original model, the unknown parameters of interest are $\boldsymbol{\rho}$, $\boldsymbol{\theta}$ and $\boldsymbol{f}$. For identification purposes discussed previously, $f_i \sim U(0,1)$ for $i = 1, 2, ..., n$, which is analogous to the assumption on the distribution of latent variable adopted in Krupskii and Joe (2013) as seen in (1.2).

In this paper, we use uniform, $U(-1,1)$ priors on the $\boldsymbol{\tau}$. The alternative of a uniform prior on $\boldsymbol{\rho}$ would imply different priors on the common scale of $\boldsymbol{\tau}$ for different copula families. Under a reasonable assumption of a priori independence, uniform priors on $\boldsymbol{f}$, $\boldsymbol{\tau}$ give the joint probability density function of the prior, $p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau}) = p(\boldsymbol{f}) \times p(\boldsymbol{\theta}) \times p(\boldsymbol{\tau}) = \prod_{i=1}^{n} \mathbb{1}(0 < f_i < 1) \times p(\boldsymbol{\theta}) \times \prod_{j=1}^{m} \mathbb{1}(-1 < \tau_j < 1)$. The choice of priors for the marginal parameters $p(\boldsymbol{\theta})$ is context specific, in our paper we use standard uninformative Jeffreys' priors unless otherwise stated.

2.4 The Augmented Posterior

The posterior of interest is $p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau}|Y)$ given by

$$\begin{aligned} p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau}|Y) \propto & p(Y|\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau}) \times p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau}) \\ & \propto \prod_{i=1}^{n} \Bigg\{ \prod_{j\in\mathcal{F}_C} c_j(P_{\boldsymbol{\theta}_j}(y_{ij}), P_f(f_i); \tau_j) \times \prod_{j\in\mathcal{F}_C} p_{\boldsymbol{\theta}_j}(y_{ij}) \times p_f(f_i) \\ & \times \prod_{j\in\mathcal{F}_D} \Big\{ C_j(P_{\boldsymbol{\theta}_j}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\boldsymbol{\theta}_j}(y_{ij}-1)|P(f_i);\tau_j) \Big\} \Bigg\} \times p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau}) \end{aligned} \tag{2.2}$$

where $p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau})$ is the joint prior density. Updating (2.2) with the chosen priors simplifies the joint posterior density to

$$\begin{aligned} p(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\tau}|Y) \propto \prod_{i=1}^{n} \Bigg[& \prod_{j\in\mathcal{F}_C} c_j(P_{\boldsymbol{\theta}_j}(y_{ij}), f_i;\tau_j) p_{\boldsymbol{\theta}_j}(y_{ij}) \times \prod_{j\in\mathcal{F}_D} \Big\{ C_j(P_{\boldsymbol{\theta}_j}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\boldsymbol{\theta}_j}(y_{ij}-1)|P(f_i);\tau_j) \Big\} \\ & \times \mathbb{1}(0 < f_i < 1) \times \prod_{j=1}^{m} p(\boldsymbol{\theta}_j) \times \prod_{j=1}^{m} \mathbb{1}(-1 < \tau_j < 1) \Bigg]. \end{aligned} \tag{2.3}$$

3 BAYESIAN ESTIMATION UNDER KNOWN COPULA FAMILIES

In this section we develop Bayesian inference scheme for the case where the parametric copula families are known, before relaxing this assumption in Section 4. The structure of the factor copula model supports a Gibbs sampling algorithm which simplifies simulation from the full joint posterior compared to a single using Metropolis Hastings step for a high dimensional parameter space. The parameter vector is broken into blocks and random draws from individual conditional posterior densities are considered. In order to obtain random draws from the parameter of interest in the respective blocks, the Metropolis-Hastings algorithm is adopted. The Metropolis-Hastings within-Gibbs algorithm for the 1-factor copula model is as follows:

1. Set initial values for parameters, $f_i^{[0]}$, $\tau_j^{[0]}$ and $\boldsymbol{\theta}_j^{[0]}$.
2. Generate $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{f},\boldsymbol{\tau},Y)$ by generating $p(\boldsymbol{\theta}_j|\boldsymbol{f},\boldsymbol{\tau},Y)$ for $j = 1,...,m$.
3. Generate $\boldsymbol{\tau}$ from $p(\boldsymbol{\tau}|\boldsymbol{f},\boldsymbol{\theta},Y)$ by generating $p(\tau_j|\boldsymbol{f},\boldsymbol{\theta},Y)$ for $j = 1,...,m$.
4. Generate $\boldsymbol{f}$ from $p(\boldsymbol{f}|\boldsymbol{\theta},\boldsymbol{\tau},Y)$ by generating $p(f_i|\boldsymbol{\theta},\boldsymbol{\tau},Y)$ for $i = 1,...,n$.
5. Repeat steps 2, 3 and 4 for $k = 2,...,K$ where K is the Monte Carlo sample size.
6. Finally, discard an initial set of burn-in draws to remove the effect of initial choice of parameter values for convergence of chain.

We emphasise that although parameters are generated one-at-a-time in steps 2-4, this is equivalent to sampling from just three blocks due to the conditional independence structure that arises from the factor copula model. From the joint posterior density in (2.3), the conditional posterior densities for each of the parameters are given by

$$p(\boldsymbol{\theta}|\boldsymbol{f},\boldsymbol{\tau},Y) \propto \prod_{j=1}^{m} p(\boldsymbol{\theta}_j|\boldsymbol{f},\boldsymbol{\tau},Y) \propto \prod_{j\in\mathcal{F}_C}\left[\prod_{i=1}^{n} c_j(P_{\boldsymbol{\theta}_j}(y_{ij}),f_i;\tau_j)p_{\boldsymbol{\theta}_j}(y_{ij})\right]\times\prod_{j=1}^{m} p(\boldsymbol{\theta}_j)\times \prod_{j\in\mathcal{F}_D}\prod_{i=1}^{n}\left\{C_j(P_{\boldsymbol{\theta}_j}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\boldsymbol{\theta}_j}(y_{ij}-1)|P(f_i);\tau_j)\right\} \quad (3.1)$$

$$p(\boldsymbol{\tau}|\boldsymbol{f},\boldsymbol{\theta},Y) \propto \prod_{j=1}^{m} p(\tau_j|\boldsymbol{f},\boldsymbol{\theta},Y) \propto \prod_{j\in\mathcal{F}_C}\left[\prod_{i=1}^{n} c_j(P_{\boldsymbol{\theta}_j}(y_{ij}),f_i;\tau_j)\right]\times\prod_{j=1}^{m} \mathbb{1}(-1<\tau_j<1)\times \prod_{j\in\mathcal{F}_D}\prod_{i=1}^{n}\left\{C_j(P_{\boldsymbol{\theta}_j}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\boldsymbol{\theta}_j}(y_{ij}-1)|P(f_i);\tau_j)\right\} \quad (3.2)$$

$$p(\boldsymbol{f}|\boldsymbol{\theta},\boldsymbol{\tau},Y) \propto \prod_{i=1}^{n} p(f_i|\boldsymbol{\theta},\boldsymbol{\tau},Y) \propto \prod_{i=1}^{n}\left[\prod_{j\in\mathcal{F}_C} c_j(P_{\boldsymbol{\theta}_j}(y_{ij}),f_i;\tau_j)\times\prod_{j\in\mathcal{F}_D}\left\{C_j(P_{\boldsymbol{\theta}_j}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\boldsymbol{\theta}_j}(y_{ij}-1)|P(f_i);\tau_j)\right\}\times\mathbb{1}(0<f_i<1)\right] \quad (3.3)$$

Note that the elements of $\boldsymbol{f}$ are independent conditional on $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ and can easily be generated as a block. Similarly, the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ from equations (3.1) and (3.2) can also be generated as a block leading to a very efficient sampling scheme. Also, once $\boldsymbol{\tau}$ are found, $\boldsymbol{\rho}$ can be simply computed using the corresponding bijective relationships of the copulas.

3.1 Initial Values

The initial values of the model parameters are obtained as follow:

1. $\boldsymbol{\theta}^{[0]} = (\boldsymbol{\theta}_1^{[0]}, \boldsymbol{\theta}_2^{[0]}, ..., \boldsymbol{\theta}_m^{[0]})$ are estimated using maximum likelihood estimates (MLE) on each margin one at a time. Other alternatives to MLE such as a Bayesian mode are possible here.

2. $\tau_j^{[0]}$ is estimated based on the bijective relationship of the jth bivariate Gaussian copula for $j = 1, 2, ..., m$ with the respective loadings obtained from linear factor analysis. In particular, we start off by obtaining the correlation coefficient (loadings l) from running classical factor analysis on data transformed to normality and map the respective loadings through the bijective relationship $\tau = \frac{2}{\pi} \arcsin l$ to obtain the corresponding τ. Instead of usual Pearson correlation matrix, the polychoric correlation matrix (assuming underlying multivariate normal latent response variables) is used when conducting factor analysis on the discrete responses.

3. Initial values for $\boldsymbol{f}$ are set by first simulating $\tilde{f}_i^j \sim p(\boldsymbol{f}|\boldsymbol{\theta}_j, \boldsymbol{\tau}_j, \mathbf{y}_j)$ by setting $\tilde{f}_i^j = \tilde{C}_j^{-1}(w_i^j | F_{\boldsymbol{\theta}_j^{[0]}}(y_{ij}); \tau_j^{[0]})$ for each $i = 1, ..., n$, $j = 1, ..., m$, where $\tilde{C}_j^{-1}(.|.;\tau_j)$ is an inverse conditional distribution function of the jth response and $w_i^j \sim U(0,1)$ (see Joe, 2014). We then compute $f_i^{[0]} = m^{-1} \sum_{j=1}^{m} \tilde{f}_i^j$ for $i = 1, 2, ..., n$.

3.2 Proposal Densities

In many cases, the conditional posterior of $\boldsymbol{\theta}$ will be recognizable under independence. Therefore, a good choice of proposal density for $\boldsymbol{\theta}$ is the recognized distribution from (3.1) which simplifies the computation of the acceptance probability α_{θ_j} to

$$\alpha_{\theta_j} = \begin{cases} min\left[1, \dfrac{\prod_{i=1}^{n} c_j(P_{\theta_j^{new}}(y_{ij}), f_i; \tau_j) p_{\theta_j^{new}}(y_{ij}) q(\theta_j^{old}|\theta_j^{new}) \times p(\theta_j^{new})}{\prod_{i=1}^{n} c_j(P_{\theta_j^{old}}(y_{ij}), f_i; \tau_j) p_{\theta_j^{old}}(y_{ij}) q(\theta_j^{new}|\theta_j^{old}) \times p(\theta_j^{old})}\right] & \text{if } j \in \mathcal{F}_{\mathcal{C}}, \\ min\left[1, \dfrac{\prod_{i=1}^{n}\left\{C_j(P_{\theta_j^{new}}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\theta_j^{new}}(y_{ij}-1)|P(f_i);\tau_j)\right\} q(\theta_j^{old}|\theta_j^{new}) \times p(\theta_j^{new})}{\prod_{i=1}^{n}\left\{C_j(P_{\theta_j^{old}}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\theta_j^{old}}(y_{ij}-1)|P(f_i);\tau_j)\right\} q(\theta_j^{new}|\theta_j^{old}) \times p(\theta_j^{old})}\right] & \text{if } j \in \mathcal{F}_{\mathcal{D}}, \end{cases}$$

$$= \begin{cases} min\left[1, \dfrac{\prod_{i=1}^{n} c_j(P_{\theta_j^{new}}(y_{ij}), f_i; \tau_j) \times p(\theta_j^{new})}{\prod_{i=1}^{n} c_j(P_{\theta_j^{old}}(y_{ij}), f_i; \tau_j) \times p(\theta_j^{old})}\right] & \text{if } j \in \mathcal{F}_{\mathcal{C}}, \\ min\left[1, \dfrac{\prod_{i=1}^{n}\left\{C_j(P_{\theta_j^{new}}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\theta_j^{new}}(y_{ij}-1)|P(f_i);\tau_j)\right\} \prod_{i=1}^{n} dP_{\theta_j^{old}}(y_{ij}) \times p(\theta_j^{new})}{\prod_{i=1}^{n}\left\{C_j(P_{\theta_j^{old}}(y_{ij})|P(f_i);\tau_j) - C_j(P_{\theta_j^{old}}(y_{ij}-1)|P(f_i);\tau_j)\right\} \prod_{i=1}^{n} dP_{\theta_j^{new}}(y_{ij}) \times p(\theta_j^{old})}\right] & \text{if } j \in \mathcal{F}_{\mathcal{D}}, \end{cases} \tag{3.4}$$

since we chose $q(\theta_j^{'}|\theta_j) = p_{\theta_j^{'}}(y_{ij})$ if $j \in \mathcal{F}_{\mathcal{C}}$ while $q(\theta_j^{'}|\theta_j) = \prod_{i=1}^{n}\left\{P_{\theta_j^{'}}(y_{ij}) - P_{\theta_j^{'}}(y_{ij}-1)\right\} = \prod_{i=1}^{n} dP_{\theta_j^{'}}(y_{ij})$ if $j \in \mathcal{F}_{\mathcal{D}}$.

At the outset, there are no clear choices of proposal densities for $\boldsymbol{\tau}$ (as well as any corresponding ν if Student's t-copula is chosen) and $\boldsymbol{f}$. Here, we implement two Random Walk Metropolis-Hastings schemes: (i) adaptive method, scaling using Robbins-Monro process (AMRM) (see Garthwaite, Fan & Sisson, 2010) and (ii) a simple adaptive method which halves or doubles the tuning parameter until the acceptance rate falls in the desired range (AMDH). Essentially, the AMRM aims to determine the tuning parameter ϕ that yields a specified/desired overall sampler acceptance rate by relying on the use of a stochastic search algorithm known as the Robbins-Monro process (Monro, 1951). Some restarting conditions are imposed to ensure the search process is well-behaved and the selected ϕ is close to optimality. For comparison, we consider another proposal based on a Gaussian[1] Laplace approximation to the conditional posterior. The mean and variance of this Gaussian distribution are given by mode and Hessian of the log of the target distribution, both of which are found numerically. Note that when estimating the copula parameter τ, some bounds have to be set in order to avoid numerical instabilities at the extreme values. In particular we restrict the range of τ such that $-0.99 < \tau < 0.99$ which is decided based on extensive stability tests.

[1] Student's-t with degrees of freedom of 3 may be used to deal with target density with fat tails.

3.3 Simulation Results

The aim here is to find the most efficient scheme among the three proposed sampling schemes. To compare the efficiency of the three schemes, we carried out a large simulation study. For brevity, we only present results where all margins are continuous and all bivariate copula pairs are Clayton. These results are representative of our simulation study and a full set of results are available from the authors upon request.

One way to measure the efficiency of the sampling algorithm is to calculate the effective sample size (ESS) for an MCMC sequence as discussed in Kass, Carlin, Gelman and Neal (1998), following an earlier suggestion by Neal (1993). The ESS gives an approximation of the equivalent number of independent samples that the correlated samples represent. To take into account of both computational and Monte Carlo efficiency, we compare the random walk schemes with a Laplace approximation algorithm based on effective sample size per second (ESS/s). Therefore, a higher ESS/s indicates that there are more independent samples per second derived from the correlated MCMC iterates.

Our simulation results in Table 1 showed that the two Random Walk Metropolis-Hastings schemes are more efficient (by a factor of at least 30 times) compared to the Laplace approximation scheme for the different combination of n and m we consider. Between the two Random Walk Metropolis-Hastings schemes, the one estimated using AMRM performs more efficiently as indicated by the total average ESS per second. In addition, whether the parameter of interest is θ, τ or f, the Random Walk AMRM scheme still dominates.

No. of Margins, m	No. of Observations, n	Average ESS/t by case		
		AMRM	AMDH	Laplace
4	100	67.62488	37.93871	1.00906
	500	32.07149	17.90954	0.29595
12	100	12.69896	10.68965	0.18920
	500	9.01946	5.22198	0.09089
36	100	2.6080	1.9370	0.0272
	500	2.0767	1.1301	0.0203
Total Average ESS/t		126.0995	74.8270	1.6326

Table 1: *Average ESS per second over all parameters across different cases and sampling schemes.*

In this section, we have compared three different schemes to estimate a one factor copula model with known bivariate copula pairs. Following this section, we will focus on simultaneous copula pair selection and parameter estimation utilizing the most efficient scheme (AMRM of the random walk scheme which is the random walk scheme using the Robbins-Monro Process).

4 BAYESIAN MODEL SELECTION

4.1 Bayesian Model with Unknown Model Specification

In the Bayesian framework, we can allow for model uncertainty simply by implementing Bayesian model averaging. Extending (2.2) to allow for simultaneous estimation of model and the corresponding parameters (see Carlin & Chib, 1995) results in the joint posterior expressed as

$$\begin{aligned} p(M, \boldsymbol{f}^{(M)}, \boldsymbol{\theta}^{(f)}, \boldsymbol{\tau}^{(M)}|Y) &\propto p(Y|M, \boldsymbol{f}^{(M)}, \boldsymbol{\theta}^{(M)}, \boldsymbol{\tau}^{(M)}) \times p(M, \boldsymbol{f}^{(M)}, \boldsymbol{\theta}^{(M)}, \boldsymbol{\tau}^{(M)}) \\ &\propto p(Y|M, \boldsymbol{f}^{(M)}, \boldsymbol{\theta}^{(M)}, \boldsymbol{\tau}^{(M)}) \times p(\boldsymbol{f}^{(M)}, \boldsymbol{\theta}^{(M)}, \boldsymbol{\tau}^{(M)}|M) \times P(M) \end{aligned} \tag{4.1}$$

which is the product of likelihood, priors on parameters and a prior on the model. To define the model set $\mathcal{M}$ we first define a candidate set of parametric copula families $\mathcal{C}$, for example $\mathcal{C}$ = {Gaussian, Clayton, Gumbel}. The model M is parameterized through indicators $\eta_j \in \mathcal{C}$ that assign a bivariate copula in $\mathcal{C}$ to the dependence between the latent factor and j^{th} variable. The complete model is therefore characterised by M={η_1,η_2,...,η_m}. Here, we adopt a uniform prior over the model set and as in Section 2.3, we adopt flat priors on all the parameters.

Updating (4.1) with the chosen priors simplifies the joint posterior density to

$$\begin{aligned} p(M, \boldsymbol{f}^{(M)}, \boldsymbol{\theta}^{(M)}, \boldsymbol{\tau}^{(M)}|Y) \propto \prod_{i=1}^{n} \Bigg[&\prod_{j \in \mathcal{F}_\mathcal{C}} c_{\eta_j}(P_{\boldsymbol{\theta}_j}(y_{ij}), f_i; \tau_j) p_{\boldsymbol{\theta}_j}(y_{ij}) \times \mathbb{1}(0 < f_i < 1) \times p(\boldsymbol{\theta}) \times \mathbb{1}(-1 < \tau_j < 1) \\ &\times \prod_{j \in \mathcal{F}_\mathcal{D}} \left\{ C_{\eta_j}(P_{\boldsymbol{\theta}_j}(y_{ij})|P(f_i); \tau_j) - C_{\eta_j}(P_{\boldsymbol{\theta}_j}(y_{ij}-1)|P(f_i); \tau_j) \right\} \Bigg] \end{aligned} \tag{4.2}$$

noting that it differs to (2.3) because of the additional subscript η_j on the copula density and conditional distribution $c_{\eta_j}(.,.;\tau_j)$ and $C_{\eta_j}(.|.;\tau_j)$ respectively which determines the copula family between the jth marginal and the latent, which in turn determines the overall model. For simplicity of notation, we have expressed the copula density and conditional distribution as a 1-parameter copula family in the preceding sections $c_{\eta_j}(.,.;\tau_j)$ and $C_{\eta_j}(.|.;\tau_j)$ respectively throughout the paper. However, $\mathcal{C}$ can include parametric families with two or more parameters,

Table 2: *Candidate pair copula families and their parameter transformation.*

Copula	Notation	Parameters	Kendall's τ	Tail Dependence
Gaussian	Ga	$\rho \in (1,-1)$	$\tau = \frac{2}{\pi} arcsin(\rho)$	$\lambda_L = \lambda_U = 0$
Student's t	T	$\rho \in (1,-1), \nu > 2$	$\tau = \frac{2}{\pi} arcsin(\rho)$	$\lambda_L = \lambda_U = 2T_{\nu+1}\left(-\sqrt{(\nu+1)\frac{1-\rho}{1+\rho}}\right)$
Clayton I	C_1	$\kappa \in \mathbb{R}\backslash\{0\}$	$\tau = \frac{\kappa}{2+\lvert\kappa\rvert}$	$\lambda_L = 2^{-\frac{1}{\lvert\kappa\rvert}}, \lambda_U = 0$
Gumbel I	G_1	$\kappa \in \mathbb{R}\backslash(-1,1)$	$\tau = \frac{\lvert\kappa\rvert-1}{\kappa}$	$\lambda_L = 0, \lambda_U = 2 - 2^{\frac{1}{\lvert\kappa\rvert}}$
Clayton II	C_2	$\kappa \in \mathbb{R}\backslash\{0\}$	$\tau = \frac{\kappa}{2+\lvert\kappa\rvert}$	$\lambda_L = 0, \lambda_U = 2^{-\frac{1}{\lvert\kappa\rvert}}$
Gumbel II	G_2	$\kappa \in \mathbb{R}\backslash(-1,1)$	$\tau = \frac{\lvert\kappa\rvert-1}{\kappa}$	$\lambda_L = 2 - 2^{\frac{1}{\lvert\kappa\rvert}}, \lambda_U = 0$

κ denotes the corresponding Archimedean copula's natural parameter.

for example we will include the t-copula in $\mathcal{C}$. In this case the density of the Student's t-copula is written as $c_{\mathrm{T}}(.,.;\tau,\nu)$ where ν is the degrees of freedom parameter.

Throughout the remainder of the paper we use the following copula families in $\mathcal{C}$. In order to allow for negative dependence for Gumbel and Clayton copulas, we redefine the Clayton and Gumbel copulas by combining them with their corresponding 2-reflected counterparts:

$$c_{\text{Clayton I}}(u,v;\tau) = \begin{cases} c_{\text{Clayton}}(u,v;\tau) & \text{if} \quad \tau \geq 0 \\ c_{\text{Clayton}}(u,1-v;-\tau) & \text{if} \quad \tau < 0 \end{cases} \tag{4.3}$$

$$c_{\text{Gumbel I}}(u,v;\tau) = \begin{cases} c_{\text{Gumbel}}(u,v;\tau) & \text{if} \quad \tau \geq 0 \\ c_{\text{Gumbel}}(u,1-v;-\tau) & \text{if} \quad \tau < 0. \end{cases} \tag{4.4}$$

To allow for richer dependence structure, survival copulas for Clayton I and Gumbel I are introduced as candidate copula families. The survival copula is also combined with the 2-reflected version to allow for tail dependence.:

$$c_{\text{Clayton II}}(u,v;\tau) = c_{\text{Clayton I}}(1-u,1-v;\tau) \tag{4.5}$$

$$c_{\text{Gumbel II}}(u,v;\tau) = c_{\text{Gumbel I}}(1-u,1-v;\tau). \tag{4.6}$$

Overall, candidate pair copula families considered in this study are $\mathcal{C}$ ={Gaussian, Student's t, Clayton I, Gumbel I, Clayton II, Gumbel II} which result in m^6 models in the model space. Some essential details and parameter transformation of these copula families are given in Table 2 while Figure 1 shows the pairs plots with normalized margins of these copulas at $\tau = 0.7$ (Top row) and $\tau = -0.7$ (Bottom row). Note that given the definitions of Clayton I and Clayton II, all 4 possible reflections of the bivariate Clayton are included in $\mathcal{C}$ (and likewise for the Gumbel).

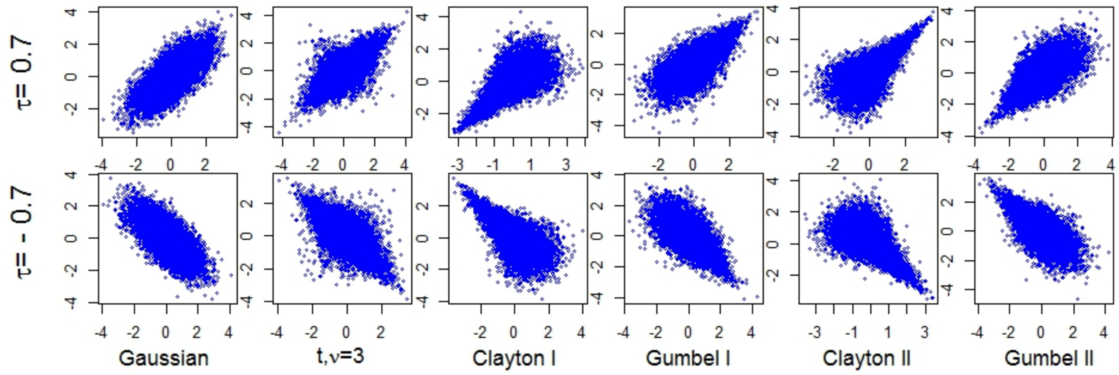

Figure 1: *Pairs plots of several different bivariate copulas at* $\tau = 0.7$ *and* $\tau = -0.7$.

4.2 Reversible Jump MCMC Algorithm

The reversible jump methodology introduced by Green (1995) is a more general MCMC technique that allows for comparison between models of different dimensions. The formulation is based on creating irreducible and aperiodic Markov chains which are able to jump between models with differing parameter spaces, while retaining the detailed balance condition which is needed to ensure a correct limiting distribution. For extensive reviews and advances of the methodology, see Sisson (2005) and Hastie and Green (2012).

Assuming we have a set of competing models $M \in \mathcal{M}$ where M is defined by the corresponding copula families, i.e. $M = \{\eta_1, \eta_2, ..., \eta_m\}$ for m marginals and $\boldsymbol{\Psi}_M$ denotes the corresponding parameter vector for the model. Given the current state of the Markov chain is $(M, \boldsymbol{\Psi}_M)$ where $\boldsymbol{\Psi}_M$ has dimension d_M, the general algorithm is as follows:

1. Set initial values for parameters, $f_i^{[0]}$, $\tau_j^{[0]}$, $\boldsymbol{\theta}_j^{[0]}$ and $c_{\eta_j}^{[0]}$ for $i = 1, ..., n$ and $j = 1, ..., m$.
2. Perform within-model move (as in Section 3) to update $\boldsymbol{\Psi}_M = \{\boldsymbol{\theta}^{(M)}, \boldsymbol{\tau}^{(M)}, \boldsymbol{f}^{(M)}\}$.
3. Propose a between-model move, from $M \to M'$.
4. Repeat steps 2 and 3 for $k = 2, ..., K$ where K is the MCMC sample size.
5. Finally, discard an initial set of burn-in draws to remove the effect of initial choice of parameter/indicator values for convergence of chain.

The initial values for the model parameters in this section are generated in the same manner as Section 3.1. As for the initial model, we have found that the convergence of our scheme is robust to the initial combination of pair copula families. In the event where we start at a Student's t-copula, the degrees of freedom parameter ν is initialized at the value 5.

Since the between model move in step 3 is based on making a jump from the current model state M to a neighbouring model M' it is important to define a neighbourhood of models. For the factor copula model this is reasonably straightforward. Two models $M = \{\eta_1, \eta_2, \ldots, \eta_m\}$ and $M' = \{\eta_1', \eta_2', \ldots, \eta_m'\}$ are neighbours if $\eta_j \neq \eta_j'$ for only one value of j and all $\eta_{i \neq j} = \eta_{i \neq j}'$. Proposing a jump $M \to M'$ involves two steps, the first is to uniformly select a random pair copula j, the second is to select $\eta_j' \in \mathcal{C} \backslash \eta_j$. This construction reduces the complex problem of dimension matching for the overall model to a simpler dimension matching problem for single copula families with either the same or a differing number of parameters. We now discuss each of these cases in turn. In this section, for conciseness, we will assume the margin j is continuous. In the case where the margin j is discrete, we replace $\prod_{i=1}^{n} c_{\eta_j}(P_{\theta_j}(y_{ij}), f_i; \tau_j)$ with $\prod_{i=1}^{n} \left\{ C_{\eta_j}(P_{\theta_j}(y_{ij})|P(f_i); \tau_j) - C_{\eta_j}(P_{\theta_j}(y_{ij}-1)|P(f_i); \tau_j) \right\}$.

Case 1: Fixed Dimension Transitions Consider a set of candidate copula families with the same number of parameters, $\mathcal{C}_1$ ={Gaussian, Clayton I, Gumbel I, Clayton II, Gumbel II}. The transition between these one-parameter copula families is less complicated since the dependence structure is captured by a single measure which has a bijective relationship with Kendall's Tau,τ. Effectively, this bijective relationship allows us to match the equivalence of the dependence measure in a general subspace where they are directly comparable. Thus, the c_{η_j} for $j = 1, ..., m$ can be sampled easily with the form of the acceptance probability for each proposed copula pair, $\alpha_{c_{\eta_j}}$ is given by

$$\alpha_{c_{\eta_j}} = min\left[1, \frac{\prod_{i=1}^{n} c_{\eta_j^{new}}(P_{\theta_j}(y_{ij}), f_i; \tau_j) r(c_{\eta_j^{new}}, c_{\eta_j^{old}})}{\prod_{i=1}^{n} c_{\eta_j^{old}}(P_{\theta_j}(y_{ij}), f_i; \tau_j) r(c_{\eta_j^{old}}, c_{\eta_j^{new}})}\right] \tag{4.7}$$

for $j = 1, 2, ..., m$ where $r(b', b)$ is the probability of making the move from b' to b. If η_j^{new} is selected uniformly from $\mathcal{C} \backslash \eta_j^{old}$ then terms with $r(b', b)$ will cancel out. However, although j is selected uniformly, η_j^{new} need not be and in Section 4.2.1 we will discuss a better proposal for $r(b', b)$.

Case 2: Varying Dimension Transitions If the proposed jump is to a copula family with a different number of parameters, for example a jump from a t-copula to a Gumbel the issue of dimension matching must be addressed. An invertible function g is introduced and auxiliary variables $\boldsymbol{a}$ are generated to map the parameter spaces to an equivalent/comparable subspace.

Whenever there is a jump or move from a 2-parameter copula to 1-parameter copula or vice versa[2], we have to perform dimension matching to ensure the detailed balance condition holds. Specifically, the dimension of the parameter space of the new copula family, c' is the same as the current copula family, c by introducing auxiliary variables, such that $d_{c'} + d_{\boldsymbol{a}'} = d_c + d_{\boldsymbol{a}}$. In this case, the form of the acceptance probability for each proposed copula pair to ensure the detailed balance condition holds is given by

$$\alpha_{c_{\eta_j^{old}}, c_{\eta_j^{new}}} = min\left[1, \frac{\prod_{i=1}^{n} c_{\eta_j^{new}}(P_{\boldsymbol{\theta}_j}(y_{ij}), f_i; \tau_j^{new}) r(c_{\eta_j^{new}}, c_{\eta_j^{old}}) q_a(\boldsymbol{a}'|c_{\eta_j^{new}}, c_{\eta_j^{old}})}{\prod_{i=1}^{n} c_{\eta_j^{old}}(P_{\boldsymbol{\theta}_j}(y_{ij}), f_i; \tau_j^{old}) r(c_{\eta_j^{old}}, c_{\eta_j^{new}}) q_a(\boldsymbol{a}|c_{\eta_j^{old}}, c_{\eta_j^{new}})} \left|\frac{\partial g(\tau_j^{new}, \boldsymbol{a}')}{\partial(\tau_j^{old}, \boldsymbol{a})}\right|\right] \tag{4.8}$$

for $j = 1, 2, ..., m$ where q_a is the proposal distribution for auxiliary variable $\boldsymbol{a}$ for matching the dimension of the copula pairs and $\frac{\partial g(\tau^{new}, \boldsymbol{a}')}{\partial(\tau^{old}, \boldsymbol{a})} = \begin{bmatrix} \frac{\partial \tau^{new}}{\partial \tau^{old}} & \frac{\partial \tau^{new}}{\partial \boldsymbol{a}} \\ \frac{\partial \boldsymbol{a}'}{\partial \tau^{old}} & \frac{\partial \boldsymbol{a}'}{\partial \boldsymbol{a}} \end{bmatrix}$ is the Jacobian of transformation. Note that we have $\tau^{new} = \tau^{old} = \tau$ since τ has the same definition across both of the copula families' parameter space. The efficiency and practicality of the implementation of the algorithm depends highly on the choice of the proposal distribution of $\boldsymbol{a}$, as well as the g function which will be covered in Section 4.2.1.

4.2.1 Proposal Density and Mapping Function

Proposal for new copula families r To ensure better mixing of the chain, it is better to propose jumps to neighbouring models that are more likely to be accepted. As such, the proposal utilizes the updated information (updated parameter estimates) available at the latest MCMC iteration as well as the data. At the kth iteration, for each $j = 1, ..., m$, we perform the following steps:

1. First, we compute the likelihood $L(\boldsymbol{y_j}|\boldsymbol{\eta_j^*}, \boldsymbol{f}^{[k]}, \boldsymbol{\theta}_j^{[k]}, \tau_j^{[k]})$ for $\boldsymbol{\eta_j^*} \in \mathcal{C} \setminus \eta_j^{[k]}$ where $\boldsymbol{\eta_j^*}$ are all the candidate copula families except the current copula family $\eta_j^{[k]}$ for the given marginal, j.

[2]For example, a Student's t-copula has two parameters, τ and ν categorizing the dependence structure while the other copula families considered here only have one parameter,τ.

2. Using the computed likelihood as weights, a set of proposal/transition probabilities r is created.

3. Finally, we sample a new copula family η_j^{new} based on the transition probabilities r.

Mapping function, g and proposal for auxiliary variables q_a Instead of proposing random auxiliary variables and using ad-hoc choice of g functions, we could make use of an additional piece of information we know about the copula families, that is the strength of the tail dependence, λ. Therefore, the term, $\frac{q_a(\boldsymbol{a}'|c_{\eta_j^{new}}, c_{\eta_j^{old}})}{q_a(\boldsymbol{a}|c_{\eta_j^{old}}, c_{\eta_j^{new}})}$ from (4.8) disappears, simplifying the computation of the acceptance probability. However, the tail dependence measure λ is generally not always directly comparable across different copula families, for example λ_C and λ_G is a measure of the one sided tail dependence while λ_T is a measure of 2-sided tail dependence[3]. To deal with this, we introduce an appropriate mapping function for each of the transitions to project them to a new comparable space. In particular we choose the auxiliary variable, $\lambda^* = \frac{\lambda_G}{2} = \frac{\lambda_C}{2} = \lambda_T$ which is just a measure of the average tail dependence over both tails. Another issue to be addressed is moving from a Gaussian copula to a Student's t-copula (and vice versa) since the Gaussian copula has no tail dependence. We will address this issue and the various possible varying dimension transitions in turn.

Consider the current copula family is c and the new copula family proposed is c'. Now, we will present the approaches to obtain the remaining terms needed to compute acceptance probability for:

- Transitioning from $c = c_{Ga}$ (Gaussian) to $c' = c_T$ (Student's t):
  As the degrees of freedom, $\nu \to \infty$, a Student's t-copula reduces to a Gaussian copula, the Gaussian copula is a limiting case of Student's t-copula. This means that the dimension is already implicitly matched and all we need now is to propose a value of ν for the Student's t-copula. Here, we choose to adopt an independent proposal for new ν since it is often difficult to obtain sufficient information from the past iterations to identify good proposal values for ν. Moreover, adopting an independent proposal reduces the Jacobian determinant to one. In this case, the ν value is proposed from a normal distribution truncated from 2 to 30 centered around 5 with standard deviation of 3, $TN \sim (5,9)[2,30]$. The choice of the mean and standard deviation here is reasonable since we are better off proposing a Gaussian copula if we accept large ν values. Putting everything together, the

[3] Subscript C denotes both Clayton I and Clayton II copulas and G denotes both Gumbel I and Gumbel II copulas.

acceptance probability is:

$$\alpha_{c_{Ga},c_T} = min\left[1, \frac{\prod_{i=1}^{n} c_T(F_{\boldsymbol{\theta}}(y_i), L_i; \tau) r(c_T, c_{Ga})}{\prod_{i=1}^{n} c_{Ga}(F_{\boldsymbol{\theta}}(y_i), L_i; \tau) r(c_{Ga}, c_T) q(\nu|c_{Ga}, c_T)}\right] \tag{4.9}$$

where $q(\nu|c_{Ga}, c_T)$ is the proposal density for ν.

- The remaining transitions:

  As discussed previously, the acceptance probabilities for these transitions can be computed using

$$\alpha_{c,c'} = min\left[1, \frac{\prod_{i=1}^{n} c'(P_{\boldsymbol{\theta}}(y_i), f_i; \tau) r(c', c)}{\prod_{i=1}^{n} c(P_{\boldsymbol{\theta}}(y_i), f_i; \tau) r(c, c')} \left|\frac{\partial g(\tau, \lambda^*)}{\partial(\tau, \lambda_c)}\right|\right] \tag{4.10}$$

  where only the Jacobian terms are not known up to this point. Here, we make use of the proposed auxiliary variable, λ^*. Denote g_1 as the mapping function from λ_c to λ^*. The Jacobian terms are tabulated in Table 3 for different transitions.

Table 3: *Computation of Jacobian determinants for various transitions.*

c	c'	$g_1(z)$	Jacobian determinant
Student's t	Gaussian	-	1
Clayton or Gumbel	Student's t	$2z$	$\frac{1}{2}$
Student's t	Clayton or Gumbel	$\frac{z}{2}$	2

Moving from Student's t to Gaussian means we have implicitly introduced $\nu = \infty$. There is no need for proposing any auxiliary variable resulting in acceptance probability which is the same as the fixed dimension transition case.

The transitions between different copula pairs can be easily extended to accommodate other parametric copula families, not limited to just the 1 or 2-parameter copula families. This can be achieved by proposing sufficient number of auxiliary variables for dimension matching when transitioning between different model size. Several other copula properties such as upper tail dependence, lower tail dependence, quantiles and various concordance measures are viable options as auxiliary variables for dimension matching across the different copula families. By adopting such approach instead of proposing random auxiliary variables, sampling efficiency is improved.

5 SIMULATION STUDY

In this section, we will first study the performance of the scheme that simultaneously carries out parameter estimation and model selection. We then compare the performance of our algorithm for different m and n. For these simulations, we only impose a weak identifying restriction $\tau_j > 0$ for the marginals with the strongest dependence parameter based on the loadings/initial values generated from linear factor analysis.

5.1 Simulation Setup

Initially, we consider 100 replications from the data-generating process with $M_1 = \{c_{Ga}, c_T, c_{C_1}, c_{G_1}, c_{C_2}, c_{G_2}\}$ with τ parameter (-0.5,0.5,-0.5,0.5,-0.5,0.5) respectively and $\nu = 3$ for a Student's t-copula. We set the first three margins to be discrete margins and the remaining three to be continuous margins. The latent values are fixed across all 100 replications in order to assess the performance of the sampling scheme in latent variable estimation. This is a challenging scenario since we attempt to estimate the marginal parameters, latent factors and true model from a large model space with a small sample size of only 250 observations.

We then consider four cases with different dimensionality and sample sizes, specifically all combinations of $m = 6, 12$ and $n = 250, 500$. The true data-generating process chosen for this part is M_1 whenever $m = 6$ and $M_2 = \{M_1, M_1\}$ whenever $m = 12$. As before, we let the first three margins to be discrete margins and the remaining three to be continuous margins when $m = 6$. We set $j = 1, 2, 3, 7, 8, 9$ to be discrete margins and the remaining six to be continuous margins when $m = 12$. We assume normal margins with $\boldsymbol{\theta}_j = (\mu_j, \sigma_j^2)$ where $\mu_j = 0$ and $\sigma_j^2 = 1$ if $j \in \mathcal{F}_\mathcal{C}$ and multinomial margins with $\boldsymbol{\theta}_j = (\boldsymbol{\pi_j})$ where $\boldsymbol{\pi_j} = \{\pi_{j,1}, ...\pi_{j,L_j}\}$ and number of categories for the margins, $L_j = 5$ if $j \in \mathcal{F}_\mathcal{D}$ for $j = 1, 2, ..., m$. However, one could easily extend the model to accommodate different marginal distributions if needed. In this simulation study, we jointly estimate the marginal parameters along with other parameters. All data are generated using an algorithm given in Joe (2014).

Due to difficulty of conducting model selection, we adopt a large Monte Carlo size, $K = 20000$. Our burn-in period is 10% of the Monte Carlo sample size. We check that the sampling scheme has converged to the limiting distribution by inspecting respective trace plots for all parameters and the Robbins-Monro tuning process is well-behaved. The sampling scheme seems to be efficient considering the quick estimation process which took only approximately 20 minutes to complete a single replication. All computations were carried out on a Red Hat Linux cluster using the Sun Grid Engine with a typical CPU clockspeed of 2.2GHz.

5.1.1 Analysis of RJMCMC output

Table 4 displays the 10 models with the highest posterior model probability for a single, typical replication. In this case the 'true' DGP does not have the highest posterior model probability rather it is the model with the seventh highest posterior probability (1.38%). The 'true' DGP cannot be easily identified here since the model space is large and furthermore some candidate copula families share similar dependence characteristics. For instance, the Clayton I (Gumbel I) and Gumbel II (Clayton II) copulas both capture lower (upper) tail dependence and the t-copula is similar to the Gaussian copula when the degrees of freedom parameter is large. It is difficult to distinguish between parametric families that only differ in the tails especially when n is small and data in the tails of the distribution is sparse. The effect of this sampling error is reduced when we consider results from all 100 replications. Table 5 summarises the number of times each model is equal to the posterior mode across 100 replications. The 'true' DGP is the posterior mode 18 times out of 100 replications, more than any other model. Although this is still low, it highlights the difficulty of distinguishing between neighbouring models in large model spaces when the sample size is small.

Table 4: *Posterior Model Probabilities of the top 10 models in the first replication.*

Replication 1							
Rank	η_1	η_2	η_3	η_4	η_5	η_6	Posterior Model Probability
1	Ga	T	G_2	G_1	C_2	G_2	0.1495
2	T	T	G_2	G_1	C_2	G_2	0.1247
3	Ga	T	G_2	G_1	C_2	G_2	0.0773
4	T	T	G_2	G_1	C_2	C_1	0.1141
5	G_1	T	G_2	G_1	C_2	G_2	0.0773
6	G_1	T	G_2	G_1	C_2	C_1	0.0669
7	$\mathbf{Ga}$	$\mathbf{T}$	$\mathbf{C_1}$	$\mathbf{G_1}$	$\mathbf{C_2}$	$\mathbf{G_2}$	**0.0138**
8	Ga	G_1	G_2	G_1	C_2	G_2	0.0117
9	Ga	Ga	G_2	G_1	C_2	G_2	0.0107
10	Ga	T	C_1	G_1	C_2	C_1	0.0102

There are 213 unique models.
The bold row is the 'true' model.

A summary of parameter estimates, including estimates of the latent factors averaged across 100 replications is shown in Table 12 in the Appendix. Notice the empirical coverage of 95%

Table 5: *Summary of posterior mode across 100 replications.*

No	η_1	η_2	η_3	η_4	η_5	η_6	Number of times selected
1	**Ga**	**T**	$\mathbf{C_1}$	$\mathbf{G_1}$	$\mathbf{C_2}$	$\mathbf{G_2}$	**18**
2	G_1	T	C_1	G_1	C_2	G_2	7
3	Ga	G_2	C_1	G_1	C_2	G_2	6
4	G_1	G_2	C_1	G_1	C_2	G_2	6
5	Ga	T	G_2	G_1	C_2	G_2	4
6	Ga	T	C_1	Ga	C_2	G_2	3
7	Ga	G_2	C_1	Ga	C_2	G_2	3
8	Ga	G_2	C_1	G_1	C_2	C_1	3
9	Ga	T	C_1	G_1	G_1	G_2	2
10	Ga	T	C_1	G_1	C_2	C_1	2

There are 50 unique models.
The bold row is the 'true' model.

credible intervals for the latent factors is in the range of 87 % to 100% with an average empirical coverage of around 94.97%. This suggests that inference on the latent factors is reliable.

5.1.2 Comparison Study

Table 6 summarizes the results of model selection based on 100 replications for the cases of $m = 6; n = 1000$, $m = 12; n = 250$ and $m = 12; n = 1000$ respectively. The first case took a total of around 37 minutes, the second case took around 27 minutes while the third case took about 56 minutes to complete one replication of the sampling scheme. As before, the convergence of every chain is monitored via the trace plot and the Robbins-Monro tuning process seems to be well-behaved.

As the number of observations (n) increases, the data is less sparse in the tails allowing the true DGP to be more easily identified. This is shown in the first row of Table 6, when $n = 1000$ the posterior mode is the 'true' DGP in 80 out of 100 replications. Table 6 also shows the effect of increasing dimension. Note that as m increases, the number of models grows exponentially and is given by $|\mathcal{M}| = 6^m$. In spite of this, our algorithm still performs well, selecting the true DGP in 79 out of 100 replications when $n = 1000$. When $n = 250, m = 12$ the selection algorithm only managed to identify the true DGP 6 out of the 100 replications, although this can be explained by the low sample size relative to the size of the model space.

Table 6: *Summary of posterior mode across 100 replications for the three different cases.*

No. of Marginals, m	No. of Observations, n	Number of times selected
6	**1000**	**80**
12	250	6
12	**1000**	**79**

The bold rows indicate that the posterior mode coincides with the 'true' model.

6 APPLICATION TO MULTIDIMENSIONAL POVERTY DATA

Recently, the Oxford Poverty and Human Development Initiative (OPHI) and the United Nations Development Programme (UNDP) developed the Multidimensional Poverty Index (MPI). The MPI has been argued to provide a more complete and comprehensive assessment of poverty and deprivation, beyond traditional measures based on income or wealth. The purpose is to aggregate multiple measures of deprivation in facets such as health, education and living standards into a single number (see Alkire & Santos, 2010; Klugman & Nations, 2010). Although rarely stated explicitly, the rationale behind such an approach is that the multidimensional nature of poverty can be explained using a single measure, albeit one that cannot be observed directly. As such, the construction of poverty indices is amenable to factor analysis in general and to the 1-factor copula model in particular. In this setting conducting inference on the poverty measure is equivalent to conducting inference on the latent factor, necessitating the Bayesian approach that we propose.

We expect the 1-factor copula model to have a number of advantages over both the MPI and a classical linear factor modeling approach. Notwithstanding constraints faced by households suffering deprivation, the data still to some extent represent choices that are rational. This seems to suggest that the MPI, which is constructed using an arbitrary, fixed and linear set of weights may not reflect the utilities and preferences faced by households in the dataset. A classical factor analysis mitigates against this to some extent, since the factor loadings are estimated in a data driven way. However the linear relationship embedded within a classical factor analysis implies the same restrictive relationship between the unobserved factor and each observable dimension of poverty. In contrast, the factor copula model with additional model averaging over different copulas is more flexible and leads to a better fit and more nuanced understanding of the relationships between different dimensions of poverty.

The dataset includes survey data of 11463 households in East Timor for the year 2009. The unit of analysis for all modeling is the household. A factor score is estimated for each household, however since the sample size is large, results are reported as aggregates over the 13 districts of East Timor, revealing the geographic distribution of poverty. The MPI is based on 10 binary indicators (where 1 reflects deprivation in that indicator) that cover three facets of deprivation: namely education, health and living standards as summarised in Table 7. More extensive definitions of the indicators are given in Alkire and Robles (2015) and references therein. Table 7 also gives the weights used for the MPI, each of what we call the 'facets' of poverty are given a combined weight of one third, with each variable within a facet assigned equal weights. As such, the MPI weights are determined by the grouping of variables into education, health and living standards. This grouping can be quite arbitrary for example, sanitation could arguably have an impact on health but is grouped into living standards. This arbitrariness has been pointed out by Rippin (2011) who also criticises the MPI for not taking into account the correlation between indicators. Factor models, including our proposed approach, aggregate the variables in a data driven way and for these approaches we will ignore the arbitrary grouping of variables into facets of education, health and living standards entirely.

Table 8 provides a summary of the proportion of deprived households in each district of East Timor for each indicator. There are in general high levels of deprivation across all the indicators with the highest rate of deprivation in cooking fuel where nationally approximately 97% of the households use dirty cooking fuel. The deprivation rates corresponding to living standards are shown to be substantially higher than the deprivation rates corresponding to education and health. In terms of districts, the capital Dili has the lowest deprivation rates while Oecussi, Ainaro and Ermera seem to face the highest deprivation rates for most indicators.

We perform a classical linear factor analysis (LFA) based on a polychoric correlation matrix of the binary variables and obtain their respective loadings which are reported in Table 9. This model will be used as a benchmark but is also used to obtain initial values for the MCMC scheme used to fit the 1-factor copula model (FCM). The identifying restriction $\tau_{\text{elec}} > 0$ is imposed ensuring that the factor measures deprivation rather than welfare. In this study, we adopted an MCMC sample size of 20000 with a 10% burn-in period. The full estimation process took less than 20 hours. The chains have converged (as shown in Figure 5 in the Appendix) and the Robbins-Monro tuning process (tuning sample size chosen to be 10000) is well behaved (as indicated in Figure 6 in the Appendix).

Facets of poverty	Indicator	Abbreviation	Deprived if	MPI weight
Education	Years of Schooling	sch	no household member has completed six years of schooling	1/6
	Child School Attendance	sch-ch	any school-aged child is not attending school up to class 8	1/6
Health	Child Mortality	mort	any child has died in the family in past 5 years	1/6
	Nutrition	nutr	any adult or child for whom there is nutritional information is malnourished	1/6
Living Standards	Electricity	elec	the household has no electricity	1/18
	Improved Sanitation	sanit	the household's sanitation facility is not improved, or it is improved but shared with other households	1/18
	Improved Drinking Water	water	the household does not have access to safe drinking water or safe drinking water is more than a 30-minute walk from home	1/18
	Flooring	floor	the household has a dirt, sand or dung floor	1/18
	Cooking Fuel	cook	the household cooks with dung, wood or charcoal	1/18
	Assets Ownership	asset	the household does not own more than one of the following assets: radio, TV, telephone, bike, motorbike or refrigerator and does not own a car or truck	1/18

Table 7: *The categories, Indicators, deprivation cutoffs and MPI weights.*

The posterior model probabilities are reported in Table 10 for the ten models with highest posterior probability. Notice that each of these models consist of a variety of copula families. The two models with highest posterior probabilities, which together account for over 60% of the posterior model probability also are composed of various copulas, but only differ between them in the copula assigned to school attendance. It is worth mentioning that the model assigning a Gaussian copula to every pair is never visited by the sampler. This sends a strong indication that the inclusion of non-Gaussian copulas in the candidate set is needed to more accurately model the dependence in the data.

Table 13 summarises the parameter estimates from the fitted factor copula model. The strength of dependence between the poverty index and the two indicators corresponding to health (child mortality and nutrition) are quite weak. This may suggests that these variables are influenced by a strong idiosyncratic component and are not as useful in forming a deprivation measure

Table 8: *Percentages of households (by district) which suffers from deprivation in the each of the 10 indicators.*

District	Education		Health		Living Standards					
	sch	sch-ch	mort	nutr	elec	sanit	water	floor	cook	asset
Aileu	21.28	30.44	20.33	38.17	75.27	71.34	56.36	80.02	99.05	74.20
Ainaro	28.20	33.64	20.27	40.66	81.54	**86.98**	**62.85**	77.46	99.21	72.82
Baucau	24.53	31.66	**11.48**	34.89	68.23	81.83	57.86	77.93	97.66	75.36
Bobonaro	27.09	30.99	20.85	56.52	70.23	61.98	31.33	56.19	98.77	72.24
Cova Lima	14.14	**19.00**	24.22	47.16	63.50	57.71	54.92	37.66	99.30	62.80
Dili	**7.03**	20.12	15.30	**34.55**	**5.29**	**26.85**	**4.04**	**21.17**	**75.75**	**21.75**
Ermera	33.14	**47.82**	26.38	55.28	**85.21**	76.38	44.61	70.87	98.05	80.62
Liquica	25.55	37.34	**26.47**	49.60	65.90	70.40	31.91	65.09	99.19	64.97
Lautem	14.59	22.00	26.00	46.82	62.82	66.00	52.00	60.94	99.65	66.12
Manufahi	23.76	22.44	20.63	40.65	69.96	84.32	60.31	55.97	99.16	68.64
Manatuto	15.48	22.71	18.31	44.18	50.62	54.01	38.87	57.74	**99.66**	72.77
Oecussi	**34.11**	39.22	23.78	**60.56**	81.78	81.11	47.11	76.33	99.22	**81.89**
Viqueque	24.70	28.74	18.88	43.82	66.75	79.69	48.46	**82.30**	99.64	79.10
Overall	22.41	29.63	20.88	45.51	64.30	68.47	44.74	62.45	96.93	68.06

The bold blue entry in each column denotes the district with the lowest deprivation rate for that indicator.
The bold red entry in each column denotes the district with the highest deprivation rate for that indicator.

Table 9: *Factor loadings for the 10 indicator variables from a classical linear factor analysis (LFA) and their corresponding τ values.*

Returns Index	Factor Loadings	τ
sch	0.334	0.2167
sch-ch	0.212	0.1357
mort	0.052	0.0334
nutr	0.018	0.0112
elec	**0.722**	**0.5140**
sanit	0.619	0.4246
water	0.401	0.2626
floor	0.588	0.4004
cook	0.304	0.1964
asset	0.699	0.4931

τ is computed based on bijective relationship between a Gaussian copula's natural parameter to Kendall's Tau.
The bold row has the largest (in magnitude) dependence strength.

that differentiates between households. This result is also in line with the results from the linear factor analysis. However the for other variables, the strength of dependence with the common factor may differ between the factor copula model and linear factor analysis. For example, the strength of dependence between the cooking fuel indicator and the estimated factor is $\tau = 0.5352$ for the factor copula model but only $\tau = 0.1964$ for the linear factor analysis.

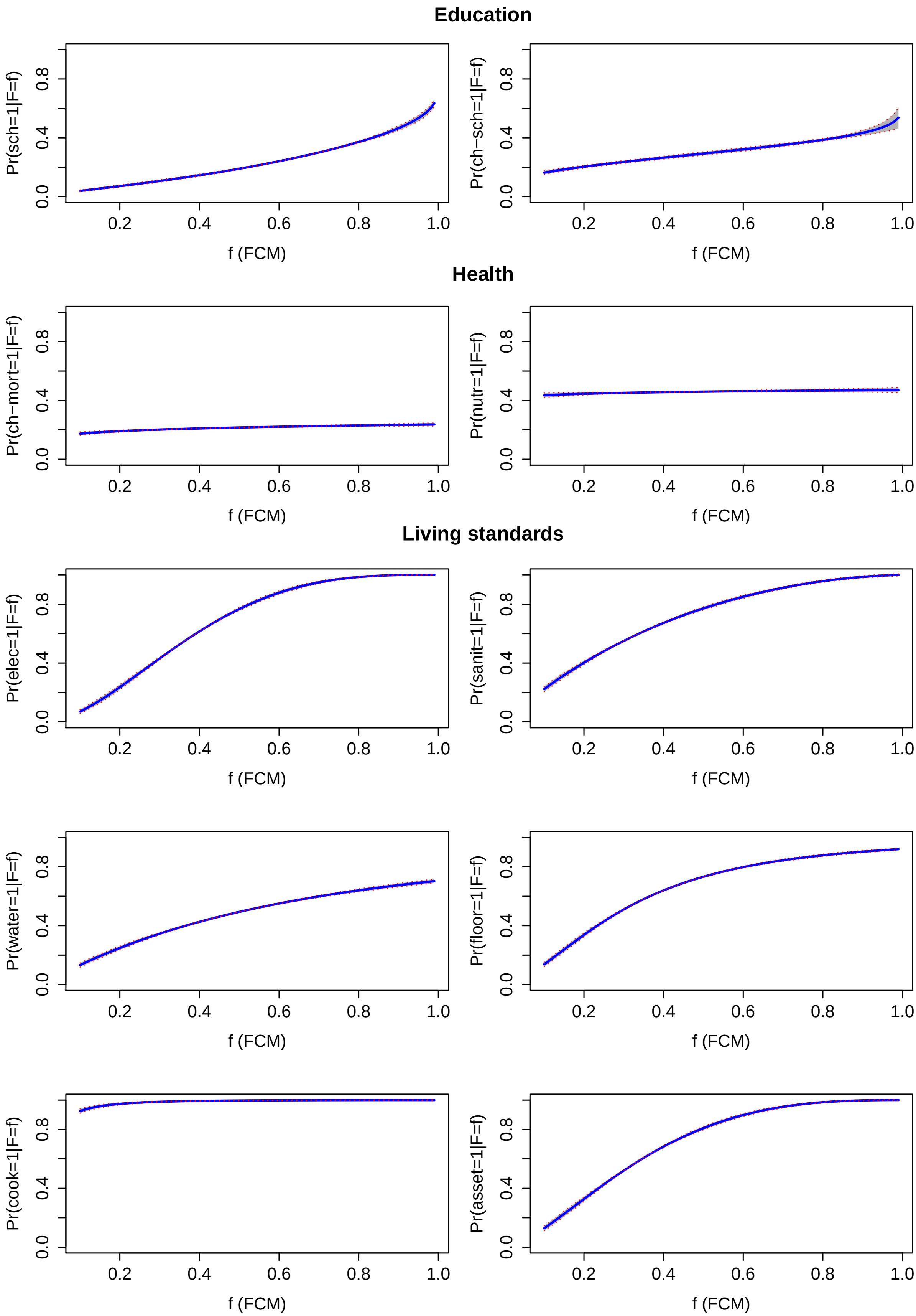

Figure 2: *Predicted marginal probabilities of deprivation in each of the indicators (—) with the corresponding 95 % credible bands (- - -) and regions (grey area) at the given level of f (FCM) for households in East Timor.*

Table 10: *Application to Multidimensional Poverty Data: Posterior Model Probabilities of the top 10 models.*

Rank	η_{sch}	η_{sch-ch}	η_{mort}	η_{nutr}	η_{elec}	η_{sanit}	η_{water}	η_{floor}	η_{cook}	η_{asset}	Model Probability
1	G_2	G_2	C_1	C_1	Ga	Ga	C_1	C_1	T	Ga	0.33000
2	G_2	Ga	C_1	C_1	Ga	Ga	C_1	C_1	T	Ga	0.31028
3	G_2	G_2	C_1	C_1	Ga	T	C_1	C_1	T	Ga	0.06550
4	G_2	Ga	C_1	C_1	Ga	T	C_1	C_1	T	Ga	0.05717
5	G_2	G_2	C_1	C_1	Ga	Ga	C_1	C_1	T	T	0.03372
6	G_2	Ga	C_1	G_2	Ga	Ga	C_1	C_1	T	Ga	0.03206
7	G_2	G_2	C_1	G_2	Ga	Ga	C_1	C_1	T	Ga	0.02956
8	G_2	Ga	C_1	C_1	Ga	Ga	C_1	C_1	T	T	0.02856
9	G_2	Ga	G_2	C_1	Ga	Ga	C_1	C_1	T	Ga	0.01522
10	G_2	Ga	C_1	C_1	T	Ga	C_1	C_1	T	Ga	0.01078

There are 69 unique models.

One of the advantages of averaging over different factor copula models is that quite complicated relationships between variables can be captured that cannot adequately be described by a single number. To demonstrate, Figure 2 shows the conditional probability of deprivation for each variable conditional on the poverty measure. The monotonically increasing behaviour is expected since increasing levels of the latent factor are associated with higher poverty and therefore higher probabilities of deprivation in each dimension. However, the shape of these curves differ dramatically across each variable. The indicators reflecting level of education (both "sch" and "sch-ch") are convex functions and deprivation in these indicators is very sensitive to movements in the households belonging to the poorest class (beyond the 80th percentile of the factor score). The conditional deprivation probabilities for the health category are flatter, suggesting that the probability of deprivation in these dimensions is not as strongly related to the latent factor. Generally for indicators reflecting living standards, the probabilities of deprivation are fairly flat for households with high levels of the factor score. It is noticeable that the probability of deprivation in ‘cook’ (cooking fuel) increases at a very fast rate for households with lower factor scores (the most privileged households) and remains at high beyond the 40th quantile.

To establish that the factor copula model is not simply overfitting the data, we perform the following cross validation exercise. The data are randomly separated into 10 non-overlapping partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{10}$,, and we attempt to predict all responses within each partition using the remaining nine partitions as a test set. Since a full probabilistic model has been specified, it is possible to evaluate different approaches using the sum of out-of-sample log scores. Log scores are obtained by estimating the joint probability mass function of the responses using training data, and evaluating this function at each realisation of the test sample. The total score for the l^{th} partition (or fold) of the data is given by

$$S_l = \sum_{\boldsymbol{y}_i \in \mathcal{P}_l} \hat{p}(\boldsymbol{y}_i; \mathcal{P}_{k \setminus l}), \tag{6.1}$$

where $\hat{p}(a; b)$ denotes the density estimate using information in b but evaluated at a. In the case of the FCM, uncertainty about the factors, parameters and model are integrated out, as is common in Bayesian approaches. In the case of the classical LFA the probability mass function is obtained by taking finite differences of multivariate Gaussian distribution with a correlation matrix formed using the estimates of the factor loadings and idiosyncratic variances which is consistent with assuming that discrete responses are determined by an underlying multivariate Gaussian variable. The cross-validation results for LFA and FCM are shown in Table 11. The average log scores show that the factor copula model performs better than the classical approach for every fold and overall. This result is in agreement with the finding in Table 10 where the factor copula model composed of only bivariate Gaussian copulas is never selected.

Table 11: *Comparison of a 10-fold Cross-Validation Results across LFA and FCM.*

Fold number	Log Scores	
	LFA	FCM
1	-5948.554	-4826.952
2	-5982.138	-4844.994
3	-5906.682	-4784.711
4	-5999.020	-4821.592
5	-6045.345	-4933.000
6	-5939.157	-4789.183
7	-5952.019	-4850.425
8	-5907.916	-4746.259
9	-5982.720	-4876.030
10	-5956.852	-4814.306
Average	-5962.74	-4828.745

Log score is defined such that it only takes negative values.
Model that maximizes the log score indicates a better fitted model to the data.

We compare the average values of the poverty indices by district from the widely used MPI as well as from the LFA and FCM. The results are summarized on the maps as illustrated in the three panels of Figure 3. Although taking district level averages understates the differences between the approaches at a household level, it is nevertheless interesting to note that the FCM leads to different conclusions about the geographical distribution of poverty compared to the MPI, as well as the LFA , albeit to a lesser extent. We note that the FCM suggests that Liquica and Bobonaro have lower levels of deprivation while Ainaro, Manufahi and Baucau have relatively higher levels of deprivation, compared to the MPI. In terms of the two factor approaches the FCM suggests that Manufahi is suffering more severe poverty compared to the LFA.

Since the main focus in researching poverty is to consider households in the tail of the welfare distribution we look beyond a simple district level average. Motivated by a cut-off value of 1/3 that has been adopted for the MPI by OPHI, in Figure 4, we report the proportion of households in each district with a poverty measure (either the MPI, LFA or FCM) in the highest third of all households. By doing so, we are able to distinguish whether the results from Figure 3 are unduly influenced by a small group of households facing severe poverty or a larger group of households facing mild to moderate poverty. Compared to the MPI, the FCM suggests lower poverty for Bobonaro, Dili, Ermera, Liquica and Manatuto, ranging from a percentage difference of around 3.5% to 15% and an approximate 5% higher poverty rate for Manufahi. These differences are significant at a 5 %level. On the other hand, there is also a significant difference between the LFA and the FCM for Dili, where the FCM approach estimated a 4% lower rate of poverty.

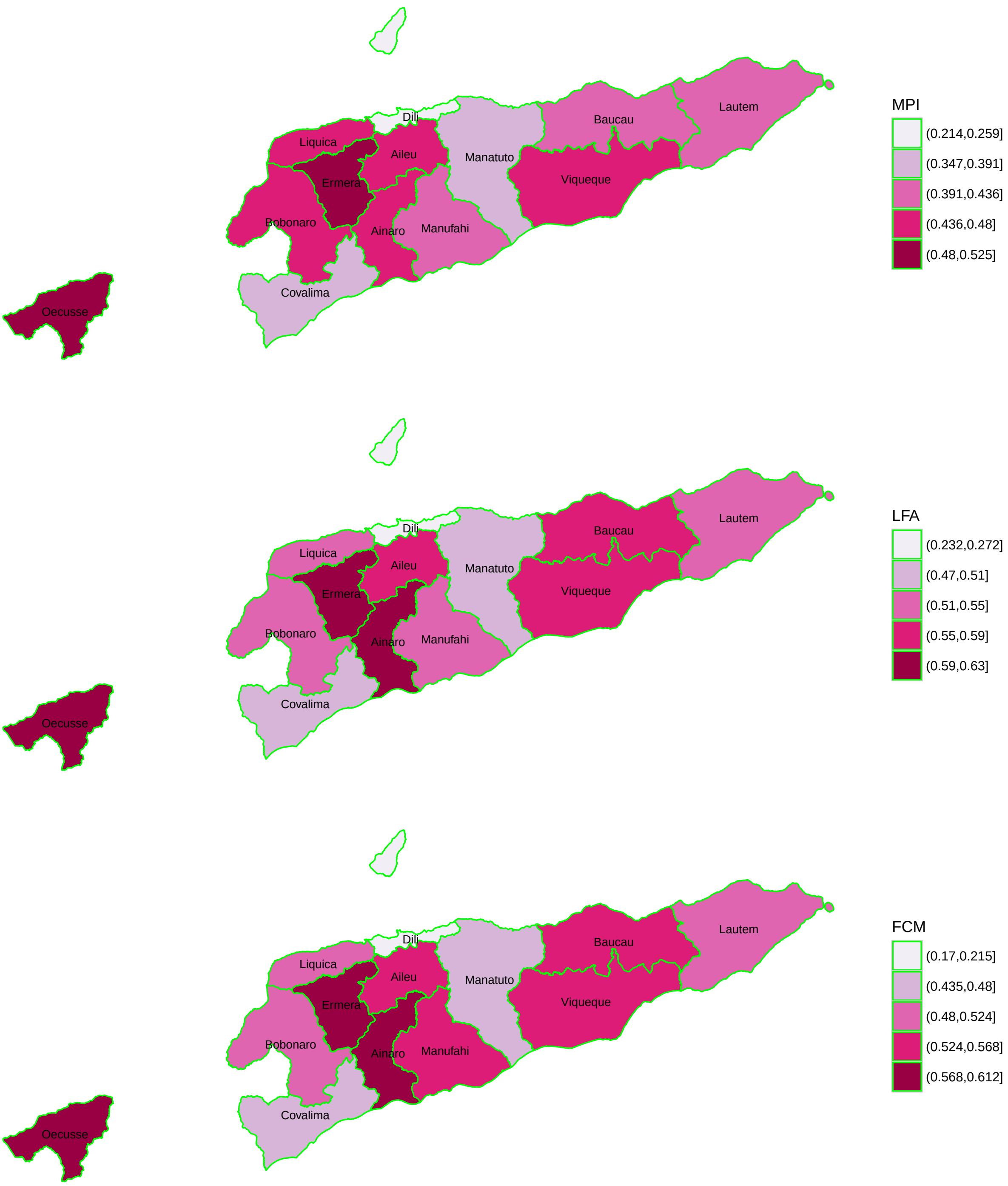

Figure 3: *Poverty measures aggregated to district level for the population of East Timor in 2009, based on the Multidimensional Poverty Index (MPI), Linear Factor Analysis (LFA) and the Factor Copula Model (FCM).*

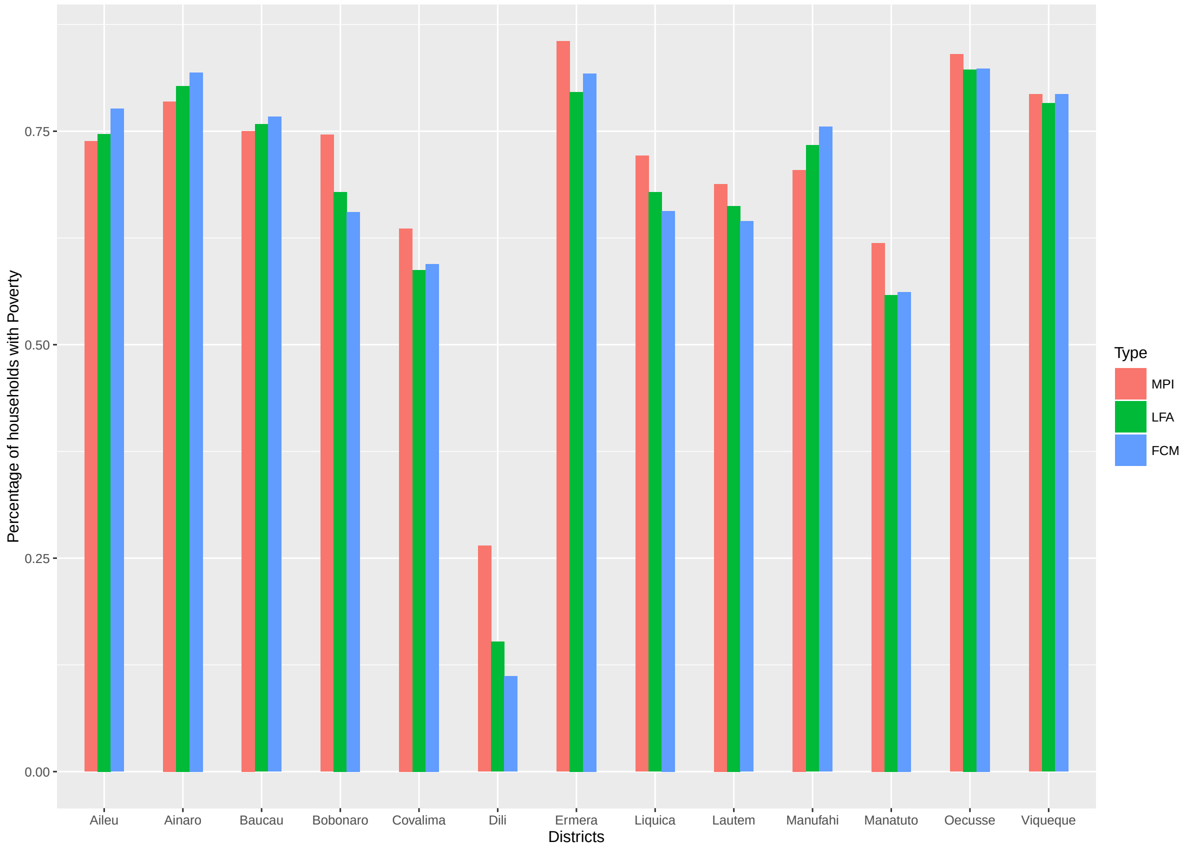

Figure 4: *Percentages of households in East Timor (by districts) which are facing poverty based on MPI, LFA and FCM respectively.*

7 CONCLUSIONS

We have developed an efficient Bayesian approach to simultaneously estimate the marginal parameters, latent factors and bivariate pair copula building blocks in a factor copula model which is applicable in high dimensional setting. We have also demonstrated with a simulation study that with a sufficiently large number of observations, our estimation procedure ensures that the 'true' model can be selected. Based on 100 replications and $n = 1000$ only, we are able to identify the 'true' model in about 80% of replications and the corresponding latent estimates from model averaging have empirical coverage close to 95%. We have displayed a potential application of this model and methodology by constructing a poverty index for East Timor, in a way that is flexible enough to fit the data better than competing methodologies.

To accommodate for multivariate data that involve some time series aspect, it is more realistic to allow for dynamics in the latent factor. The challenge for future research here is to find a way to efficiently sample the latent factor, possible by writing the model in state space form. If such techniques are developed they should be easy to combine with the approach to model selection that we propose. Another extension is to consider the case where underlying variables may vary with more than one factor. The extension of the Bayesian framework to p-factor for $p > 1$ copula models with model selection is another interesting avenue of future research.

ACKNOWLEDGEMENTS

We thank Associate Professor Xibin Zhang from Department of Econometrics and Business Statistics, Monash University, Australia for bringing our attention to the adaptive method using Robbins-Monro process proposed by Garthwaite et al. (2010) which significantly improves the efficiency of our sampling algorithm. We thank Professor Brett Inder from Department of Econometrics and Business Statistics, Monash University, Australia for providing us with the dataset for East Timor. We also thank Professor Mervyn Silvapulle, for providing general comments and feedback.

References

Alkire, S & Robles, G. (2015). Multidimensional poverty index-summer 2015: Brief methodological note and results. *University of Oxford, June*.

Alkire, S & Santos, ME. (2010). Acute multidimensional poverty: A new index for developing countries.

Anderson, TW & Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third berkeley symposium on mathematical statistics and probability* (Vol. 5, *1*, pp.1).

Asparouhov, T & Muthén, BO. (2009). *Exploratory Structural Equation Modeling*.

Carlin, BP & Chib, S. (1995). Bayesian model choice via markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 473–484.

Chamberlain, G & Rothschild, M. (1982). Arbitrage, factor structure, and mean-variance analysis on large asset markets. National Bureau of Economic Research Cambridge, Mass., USA.

Forni, M, Hallin, M, Lippi, M & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, **82**(4), 540–554.

Garthwaite, PH, Fan, Y & Sisson, Sa. (2010). Adaptive Optimal Scaling of Metropolis-Hastings Algorithms Using the Robbins-Monro Process, 1–19. arXiv: 1006.3690

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D Aigner & A Goldberger (Eds.), *Latent variables in socio-economic models* (pp.365–383). Amsterdam: New Holland.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711.

Hastie, D & Green, P. (2012). Model choice using reversible jump Markov chain Monte Carlo. *Statistica Neerlandica*, **66**(3), 309–338.

Hull, JC & White, AD. (2004). Valuation of a cdo and an n-th to default cds without monte carlo simulation. *The Journal of Derivatives*, **12**(2), 8–23.

Joe, H. (2014). *Dependence modeling with copulas*. CRC Press.

Kass, RE, Carlin, BP, Gelman, A & Neal, RM. (1998). Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician*, **52**(2), 93–100.

Klugman, J & Nations, DPU. (2010). *The real wealth of nations: Pathways to human development*. Palgrave Macmillan.

Krupskii, P & Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, **120**, 85–101.

Monro, HRS. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, **22**, 400–407.

Neal, R. (1993). Probabilistic inference using Markov chain Monte Carlo methods. **1**.

Nikoloulopoulos, AK & Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, **80**(1), 126–150.

Oh, D & Patton, A. (2012). Modelling dependence in high dimensions with factor copulas. *Manuscript, Duke University*, (June).

Rippin, N. (2011). A response to the weaknesses of the multidimensional poverty index (mpi): The correlation sensitive poverty index (cspi). *DIE Briefing Paper*, **19**, 2011.

Sargent, TJ, Sims, CA et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, **1**, 145–168.

Sisson, Sa. (2005). Transdimensional Markov Chains. *Journal of the American Statistical Association*, **100**(471), 1077–1089.

Stock, JH & Watson, MW. (1989). New indexes of coincident and leading economic indicators. In *Nber macroeconomics annual 1989, volume 4* (pp.351–409). MIT press.

Stock, J & Watson, M. (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, **97**(460), 1167–1179.

APPENDIX

Parameters	ESS	P.Mean	P.Med	Lower	Upper	True Values	Coverage (%)
$\pi_{1,1}$	2553.7537	0.1843	0.1840	0.1578	0.2110	0.2000	68
$\pi_{1,2}$	3040.1989	0.1972	0.1969	0.1693	0.2250	0.2000	77
$\pi_{1,3}$	3132.8621	0.2007	0.2004	0.1728	0.2290	0.2000	83
$\pi_{1,4}$	3095.9441	0.2050	0.2048	0.1770	0.2334	0.2000	81
$\pi_{1,5}$	2329.8008	0.2128	0.2126	0.1848	0.2412	0.2000	72
$\pi_{2,1}$	2252.7803	0.2068	0.2066	0.1791	0.2350	0.2000	88
$\pi_{2,2}$	3112.4846	0.2065	0.2062	0.1782	0.2349	0.2000	77
$\pi_{2,3}$	3105.5912	0.2007	0.2004	0.1729	0.2290	0.2000	84
$\pi_{2,4}$	3015.4998	0.2009	0.2007	0.1730	0.2294	0.2000	79
$\pi_{2,5}$	2578.2864	0.1851	0.1849	0.1587	0.2121	0.2000	64
$\pi_{3,1}$	1913.1583	0.1823	0.1820	0.1564	0.2089	0.2000	61
$\pi_{3,2}$	2972.3208	0.1970	0.1968	0.1695	0.2250	0.2000	70
$\pi_{3,3}$	2938.2648	0.2037	0.2034	0.1758	0.2323	0.2000	74
$\pi_{3,4}$	2887.5855	0.2062	0.2060	0.1782	0.2346	0.2000	70
$\pi_{3,5}$	2708.5485	0.2107	0.2104	0.1825	0.2391	0.2000	76
σ_4^2	1747.0633	0.9620	0.9573	0.8048	1.1265	1.0000	97
σ_5^2	1076.7366	1.0177	1.0131	0.8562	1.1872	1.0000	95
σ_6^2	1382.7477	1.0133	1.0086	0.8519	1.1842	1.0000	99
μ_4	1790.2399	-0.0481	-0.0479	-0.1510	0.0547	0.0000	88
μ_5	1171.1046	0.0321	0.0324	-0.0738	0.1362	0.0000	91
μ_6	1397.7287	-0.0349	-0.0351	-0.1404	0.0700	0.0000	94
τ_1	1718.5933	-0.4878	-0.4888	-0.5594	-0.4140	-0.5000	94
τ_2	1763.8030	0.4851	0.4860	0.4097	0.5595	0.5000	96
τ_3	1443.9771	-0.4864	-0.4872	-0.5607	-0.4099	-0.5000	95
τ_4	1340.7310	0.4825	0.4834	0.4092	0.5536	0.5000	94
τ_5	892.1476	-0.5092	-0.5098	-0.5796	-0.4369	-0.5000	100
τ_6	1213.4956	0.4946	0.4955	0.4224	0.5657	0.5000	98
f_1	5372.5579	0.5654	0.5689	0.3114	0.8154	0.5776	96
f_2	5655.2301	0.4926	0.4917	0.2343	0.7523	0.4588	94
f_3	3734.2228	0.7896	0.8044	0.5923	0.9547	0.8355	96
f_4	5274.1442	0.4706	0.4681	0.2274	0.7170	0.4155	89
f_5	5304.6357	0.5038	0.5031	0.2532	0.7583	0.4853	93
f_6	4877.4892	0.6601	0.6703	0.4132	0.8864	0.6592	93
f_7	5512.1907	0.5152	0.5156	0.2583	0.7715	0.4903	92
f_8	3047.6649	0.1621	0.1482	0.0361	0.3186	0.0986	95
f_9	5451.9264	0.5132	0.5136	0.2604	0.7657	0.5088	95
f_{10}	3988.8744	0.2624	0.2491	0.0797	0.4724	0.1681	87

To save space, only the first 10 out of the 100 latent variables are reported in this table.
Only common parameters across all the models are reported. The ν parameter for Student's t-copula is not reported.
The lower and upper column indicates the 95% credible interval. The Highest Posterior Density as the credible interval was adopted due to the presence of asymmetry in the posterior densities.

Table 12: *Full parameters estimation for the common parameters from BMA averaged over 100 replications.*

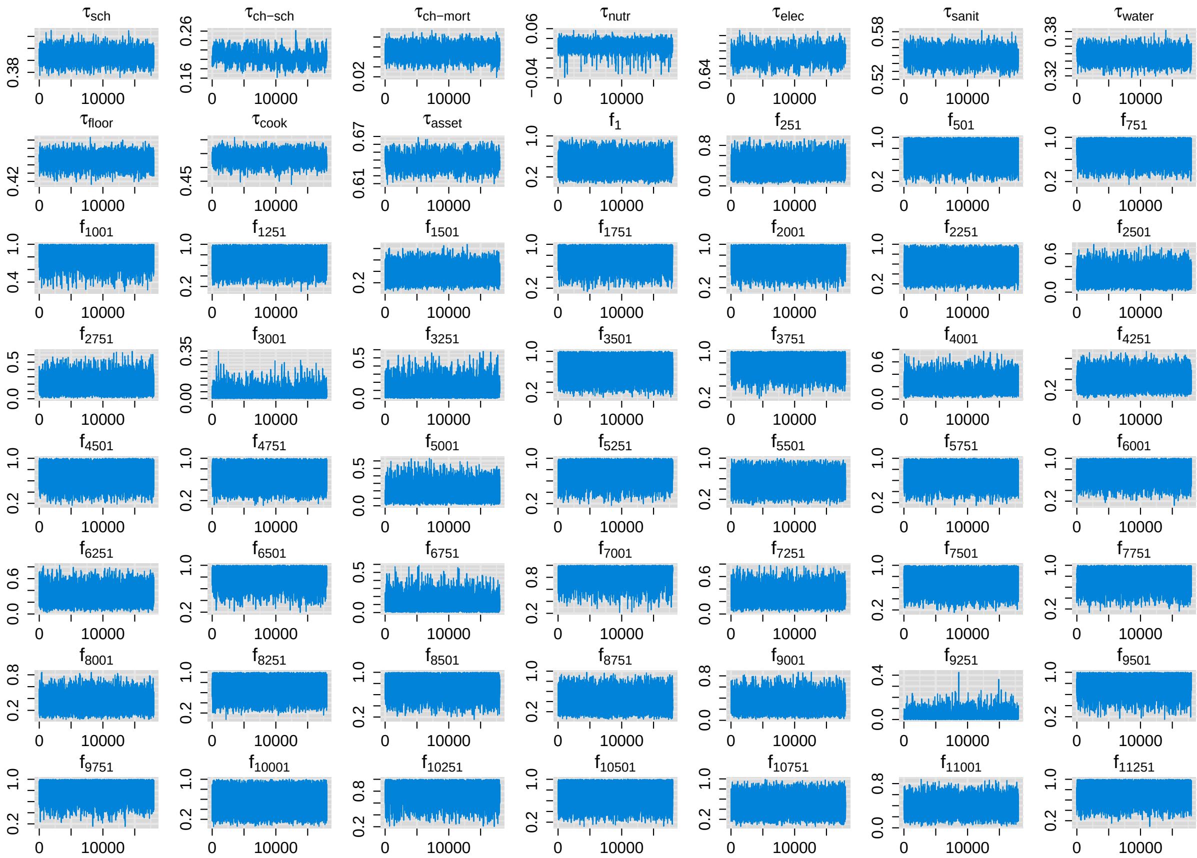

Figure 5: *Application to Multidimensional Poverty Data: trace plots for a subset of the parameters.*

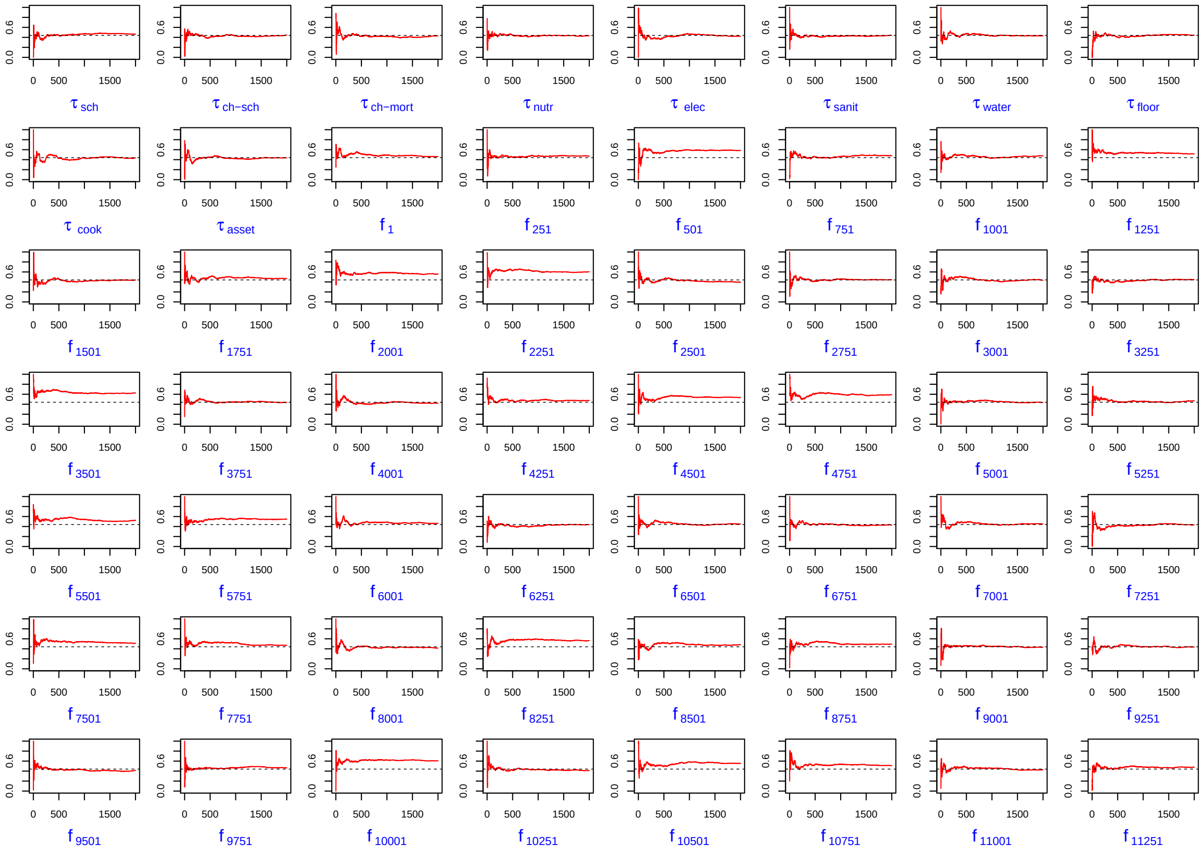

Figure 6: *Application to Multidimensional Poverty Data: running acceptance rate for a subset of the parameters.*

Parameter	ESS	P.Mean	P.Med	Lower	Upper
π_{sch}	8636.5006	0.2242	0.2242	0.2199	0.2288
π_{sch-ch}	10328.9499	0.2964	0.2964	0.2916	0.3013
π_{mort}	11386.2121	0.2089	0.2089	0.2045	0.2131
π_{nutr}	11018.5035	0.4552	0.4551	0.4500	0.4604
π_{elec}	5548.3033	0.6421	0.6421	0.6375	0.6470
π_{sanit}	7148.9075	0.6845	0.6845	0.6796	0.6892
π_{water}	10245.9428	0.4473	0.4473	0.4421	0.4522
π_{floor}	7068.2500	0.6242	0.6242	0.6195	0.6291
π_{cook}	7063.9373	0.9696	0.9696	0.9678	0.9714
π_{asset}	5794.9712	0.6798	0.6798	0.6752	0.6843
τ_{sch}	1746.5773	0.4081	0.4081	0.3883	0.4294
τ_{sch-ch}	155.7530	0.2054	0.2054	0.1807	0.2299
τ_{mort}	2653.7793	0.0700	0.0700	0.0457	0.0924
τ_{nutr}	1120.3968	0.0236	0.0242	0.0061	0.0416
τ_{elec}	886.0182	0.6647	0.6646	0.6475	0.6812
τ_{sanit}	1389.3169	0.5496	0.5498	0.5327	0.5660
τ_{water}	1993.9924	0.3472	0.3472	0.3313	0.3644
τ_{floor}	1905.7054	0.4450	0.4452	0.4305	0.4592
τ_{cook}	1521.6322	0.5362	0.5368	0.4965	0.5734
τ_{asset}	1091.1833	0.6381	0.6380	0.6211	0.6545
f_1	7120.1685	0.4363	0.4250	0.1545	0.7369
f_2	4915.7947	0.2877	0.2725	0.0748	0.5242
f_3	5800.3564	0.3576	0.3476	0.1155	0.5969
f_4	4861.4555	0.3075	0.2945	0.0689	0.5384
f_5	4035.0750	0.1612	0.1454	0.0107	0.3391
f_6	4662.6001	0.2506	0.2338	0.0505	0.4778
f_7	3644.1886	0.1622	0.1455	0.0103	0.3526
f_8	4201.0603	0.2184	0.2049	0.0364	0.4306
f_9	2666.9766	0.0952	0.0763	0.0000	0.2452
f_{10}	8330.5007	0.6685	0.6741	0.3504	0.9998

To save space, only the first 10 out of the 11463 latent variables are reported in this table.
Only common parameters across all the models are reported. The ν parameter for Student's t-copula is not reported.
π_j represents the estimated proportion of households which are deprived in indicator j.
The lower and upper column indicates the 95 % credible interval. The Highest Posterior Density as the credible interval was adopted due to the presence of asymmetry in the posterior densities.

Table 13: *Application to Multidimensional Poverty Data: full parameter estimation for the common parameters from BMA.*