Received: 12 August 2020 | Accepted: 28 June 2021

DOI: 10.1111/rssc.12517

ROYAL STATISTICAL SOCIETY
Series C Applied Statistics

ORIGINAL ARTICLE

Beyond unidimensional poverty analysis using distributional copula models for mixed ordered-continuous outcomes

Maike Hohberg[1†] | **Francesco Donat**[2‡] | **Giampiero Marra**[3] | **Thomas Kneib**[1]

[1]Chair of Statistics, University of Goettingen, Gottingen, Germany

[2]Single Resolution Board, Brussels, Belgium

[3]Department of Statistical Science, University College London, London, UK

Correspondence
Maike Hohberg, Chair of Statistics, University of Goettingen, Goettingen, Germany.
Email: mhohber@uni-goettingen.de

Abstract

Poverty is a multidimensional concept often comprising a monetary outcome and other welfare dimensions such as education, subjective well-being or health that are measured on an ordinal scale. In applied research, multidimensional poverty is ubiquitously assessed by studying each poverty dimension independently in univariate regression models or by combining several poverty dimensions into a scalar index. This approach inhibits a thorough analysis of the potentially varying interdependence between the poverty dimensions. We propose a multivariate copula generalized additive model for location, scale and shape (copula GAMLSS or distributional copula model) to tackle this challenge. By relating the copula parameter to covariates, we specifically examine if certain factors determine the dependence between poverty dimensions. Furthermore, specifying the full conditional bivariate distribution

†The author received funding from the Ministry for Science and Culture of Lower Saxony as a part of the project "Reducing Poverty Risks in Developing Countries" and the German Science Foundation within the research project KN 922/9-1.

‡This paper should not be reported as representing the views of the Single Resolution Board. The views expressed are those of the authors and do not necessarily reflect those of the Board.

allows us to derive several features such as poverty risks and dependence measures coherently from one model for different individuals. We demonstrate the approach by studying two important poverty dimensions: income and education. Since the level of education is measured on an ordinal scale while income is continuous, we extend the bivariate copula GAMLSS to the case of mixed ordered-continuous outcomes. The new model is integrated into the `GJRM` package in `R` and applied to data from Indonesia. Particular emphasis is given to the spatial variation of the income–education dependence and groups of individuals at risk of being simultaneously poor in both education and income dimensions.

KEYWORDS

bivariate distributional regression, copula regression, GAMLSS, multidimensional poverty

1 | INTRODUCTION

Although poverty is widely regarded a multidimensional phenomenon and poverty measures moving beyond a single monetary dimension—such as the Multidimensional Poverty Index (MPI, Alkire et al., 2012)—have emerged, little progress has been made on *analysing* poverty as a multidimensional concept. To study poverty at the micro level, univariate linear regression is the standard tool of the empirical economist. Despite their widespread use, however, univariate models for poverty analyses require either studying each poverty dimension separately in different equations, or using as response variable an index that subsumes all dimensions in a single number (e.g. Alkire & Fang, 2019). Both approaches neglect the interdependence between poverty dimensions and ignore that the dependence itself should be part of the analysis. In fact, the level of poverty and well-being depends on the strength of the dependence (Duclos et al., 2006): for example, lower tail dependence can explain persisting poverty, where performing low in one dimension is strongly associated with a low outcome in the other dimensions.

To overcome such limitations, multivariate regression models can be used to tackle multidimensionality in poverty analyses. The relationship between two or more outcomes can also be modelled using copulas which have been proven to be useful and flexible tools in this regard (see Nelsen, 2006, for an introduction to copula theory). A second issue in poverty analysis concerns distributional aspects. Especially for program targeting and risk factor analysis, it is important that poverty studies move beyond the simple mean effects. In fact, concepts like vulnerability to poverty—a forward-looking measure of individuals' exposure to poverty—encompass both the location and the scale of the target distribution. Previous studies on vulnerability to poverty used a step-wise procedure to explicitly make the scale parameter dependent on covariates (see Calvo & Dercon, 2013; Günther & Harttgen, 2009; Thi Nguyen et al., 2015; Zereyesus et al., 2017, for recent works). Another example is inequality, which has become growingly relevant for both the political agenda and for projects implemented in developing countries. The World

Bank, for example, centres its shared prosperity initiative around the goal to reduce inequality (World Bank, 2018). Hence, it is necessary to analyse not only effects on the mean but also on the other parameters characterizing the distribution of the outcomes of interest. Generalized additive models for location, scale and shape (GAMLSS, Rigby & Stasinopoulos, 2005) are able to capture the effects of covariates on the whole conditional distribution of a single poverty dimension.

Both issues of multidimensionality and distributional aspects can be addressed with a combination of GAMLSS and multivariate copula models, also referred to as copula GAMLSS. These models are implemented in the `R` package `GJRM` (Marra & Radice, 2019) and comprise a wide range of potential marginal distributions (continuous, binary and discrete) and copulas. A Bayesian version of this model class is implemented in the software `BayesX` (Belitz et al., 2015) while Klein and Kneib (2016) provide the related literature. The advantage of embedding copula regression into GAMLSS is that each parameter of the marginals and the copula association parameter can be modelled to depend flexibly on covariates. This allows us to not only measure the strength of the dependence, which has been the focus of previous literature on interrelated poverty dimensions, but also to analyse which factors related to household location and composition drive this dependence. This latter aspect has not been previously considered in poverty studies.

When studying poverty, it often occurs that one dimension is reported in ordered categories whereas the other is reported as continuous. For example, two possible dimensions of interest could be income (measured on the continuous scale) and the highest level of education, which is often assessed in ordered categories such as 'no schooling', 'elementary school', 'high school' and 'higher education'. This is a very relevant case, especially in economics and poverty research where several outcomes are measured on the ordinal scale (health, education, subjective well-being, to name just but a few).

The aim of this paper is twofold. First, to theoretically extend copula GAMLSS to a mixed ordered-continuous case. Second, to practically demonstrate how multidimensional poverty analysis can benefit from flexible models that allow for covariate effects on the interdependence between the poverty dimensions.

For the theoretical part, we rely on the latent variable approach that relates the ordered categories to an underlying continuous variable. A similar approach was used by de Leon and Wu (2011) (albeit not within a GAMLSS context) who employed Gaussian copulas to account for the dependence between the outcomes, but without relating the copula parameter to covariates. More recently, Vatter and Chavez-Demoulin (2015) developed a framework for copula regression within generalized additive models (GAM) for two continuous outcomes, while a model with two ordinal outcomes was suggested in Donat and Marra (2018). To relate all distributional parameter of the marginals and the association parameter to covariates, we follow the proposal of Marra and Radice (2017). Their approach estimates the copula dependence and the marginal distribution parameters simultaneously within a penalized likelihood framework using a trust region algorithm. The framework developed in this paper is also in line with Klein et al. (2019) who, however, studied binary and continuous outcomes. The new model is incorporated into the `R` package `GJRM` (Marra & Radice, 2019).

Regarding the application to multidimensional poverty, there is an extensive literature dealing with the measurement of multidimensional poverty. Yet, the methods proposed for analysing multidimensional poverty, including its determinants and poverty profiles, are rather limited. For example, Alkire et al. (2004) suggested employing generalized linear models using a single number index as the response variable. To demonstrate how a more comprehensive poverty analysis can be conducted by researchers, the empirical study in this paper applies copula GAMLSS in this context. Our application deals with two interrelated important poverty

dimensions: income and education. In many developing countries, there is potentially a vicious cycle of poor education and low income. This cycle is also called poverty trap and is a long-established concept in economics: capable children stay under-educated due to their parents' restricted resources and hence remain poor when grown-up (Barham et al., 1995). Understanding what determines the interdependence between poverty dimensions helps designing strategies to interrupt this cycle. To this end, we model the income–education dependency in Indonesia and draw an in-depth picture of monetary and education poverty across the population. We address the following questions: (1) Which factors determine the distributions of household income per capita and individual education and their interdependence? (2) How does this dependence differ spatially across Indonesia? (3) What are the probabilities of being poorly educated *and* income poor for different population's sub-groups? We will answer these questions using a rich dataset from Indonesia which is made publicly available by the RAND corporation (RAND, 2017).

The dependencies between different poverty dimensions have been widely addressed in the economics literature during the last two decades. However, the literature on using copulas to model multidimensional poverty is scarce and, to the best of our knowledge, restricted to the *measurement* of the strength of such dependence. Existing approaches do not place the model into a regression framework and hence neither relate the copula association parameter nor the other parameters characterizing the marginal distributions to covariates. For example, Quinn (2007) quantified the dependence between income and an ordinal health measure in four industrial countries. Decancq (2014) used copula models to measure dependence over time between income, health and schooling (all assumed to be continuous) in Russia. A similar approach was used by Perez and Prieto (2015) to study the dependence between income, material needs and work intensity in Spain. Kobus and Kurek (2018) analysed the distributions of health and education. In contrast to Quinn (2007) and this paper, that make use of a latent variable approach to represent the ordered categories of education, Kobus and Kurek (2018) overcame the unidentifiability issue when using copulas with discrete marginals by concordance ordering. In a Bayesian context, Tan et al. (2018) re-constructed the MPI using a one-factor copula model and data from East-Timur. These examples emphasize once more the importance of extending copula GAMLSS also to the case of mixed ordered-continuous outcomes when these models are applied to poverty analyses.

The remainder of the paper is organized as follows: Section 2 introduces a bivariate copula GAMLSS for mixed ordered-continuous outcomes. Section 3 presents the estimation procedure. Finally, Section 4 studies poverty dimensions with copula GAMLSS using data from Indonesia and discusses practical approaches to model selection. Section 5 concludes the paper.

2 | MODEL DEFINITION

2.1 | Bivariate mixed ordered-continuous model

The model considered in this paper deals with a pair of random variables, $(Y_1,\ Y_2)'$, with support $\mathcal{R} \times \mathbb{R}$, where $(\mathcal{R},\ \preceq)$ is a totally ordered set under the ordering relation $\preceq$. The elements of $\mathcal{R}$ are denoted by r and represent the levels of the categorical variable Y_1, namely $\mathcal{R} := \{1, \ldots, r, \ldots, R+1\}$ with $R+1 < \infty$. The variable Y_2 is assumed to be continuous. In the case study of Section 4, response Y_2 will represent the income and Y_1 the highest level of education attained by each individual surveyed.

We are interested in building up a statistical model for the joint distribution of the response variables $(Y_1, Y_2)'$ where their dependence structure is represented by means of a copula specification. The bivariate cumulative distribution function can then be written as

$$F_{12}(r, y_2) = C(F_1(r), F_2(y_2)) \in [0, 1], \tag{1}$$

where the copula is the map $C : [0, 1]^2 \longrightarrow [0, 1]$, with $F_1(r) := \mathbb{P}(Y_1 \leq r)$ and $F_2(y_2) := \mathbb{P}(Y_2 \leq y_2)$ being the marginal distributions. A significant advantage of the copula representation is that it decomposes the joint distribution into two marginals distributions, that may come from different families, and a copula function C that binds them together. The dependence structure of the two marginals is captured by an association parameter γ that is specific to the copula employed as described below.

If both F_1 and F_2 are continuous, Sklar's theorem ensures that the copula function is uniquely determined (Sklar, 1959). However, since Y_1 is categorical in our case, the uniqueness of the copula does not apply directly. We address this limitation by representing the ordinal outcome as a coarse version of a latent continuous random variable.

Let $Y_1^* \in \mathbb{R}$ denote the unobserved (or latent) continuous variable that drives the decision for the observed categories in $\mathcal{R}$. This continuous latent variable can be modelled as

$$Y_1^* = \mathbf{x}_1' \boldsymbol{\beta}_1 + \epsilon_1, \tag{2}$$

where $\boldsymbol{\beta}_1$ is a vector of regression coefficients, $\mathbf{x}_1$ is a vector of covariates and ε_1 is an error term with density f_1^* and cumulative distribution function (CDF) F_1^*. Later on, the latent variable in Equation (2) will be placed into the more sophisticated GAMLSS framework, but this model formulation with only linear effects shall serve as a starting point. In line with McKelvey and Zavoina (1975), the following observation rule linking the latent to the observed variable is applied:

$$\{Y_1 = r\} \Longleftrightarrow \{\theta_{r-1} < Y_1^* \leq \theta_r\}, \quad r = 1, \ldots, R+1, \tag{3}$$

where θ_r is a cut point on the latent continuum related to the level r of Y_1. We observe category r if the latent variable is between the cut-offs θ_{r-1} and θ_r. There is a total of $R + 2$ cut points: $-\infty = \theta_0 < \theta_1 < \ldots < \theta_{R+1} = \infty$. However, only R of them are estimable, namely $\{\theta_1, \ldots, \theta_R\}$.

To guarantee the monotonicity of the cut points, we apply the transformation $\theta_1^* := \theta_1$ and $\theta_r = (\theta_r^*)^2 + \theta_{r-1}$ for any $r > 1$ (Donat & Marra, 2017). This implies that $\theta_r \geq \theta_{r-1}$ for any $r \in \mathcal{R}$ and $\theta_r \in \mathbb{R}$. However, the equality $\theta_r = \theta_{r-1}$ can be problematic in practice because it results in estimated parameters at the boundary of the parameter space. This happens, for example, wherever a given level of Y_1 has no observations in the sample (Haberman, 1980).

From Equations (2) and (3), we derive the cumulative link model

$$\mathbb{P}(Y_1 \leq r) = \mathbb{P}(Y_1^* \leq \theta_r) = \mathbb{P}(\epsilon_1 \leq \theta_r - \mathbf{x}_1' \boldsymbol{\beta}_1) := F_1^*(\underbrace{\theta_r - \mathbf{x}_1' \boldsymbol{\beta}_1}_{:=\eta_{1r}}), \tag{4}$$

where η_{1r} is the predictor associated with the ordinal categorical response in the model. It depends on the observed level r of Y_1 through cut point θ_r. In Section 2.2 the predictor η_{1r} will be replaced with a generalized additive form. With this information in hand, Equation (1) can be equivalently written as

$$F_{12}(r, y_2) = F^*_{12}(\eta_{1r}, y_2) = C(F_1(r), F_2(y_2)) = C(F^*_1(\eta_{1r}), F_2(y_2)). \quad (5)$$

Since both marginals are now continuous, the applicability of Sklar's theorem in ensured for the latent model formulation. Note, however, that this is only but one possibility to deal with the non-identifiability of the copula function and that various underlying continuous models can lead to exactly the same copula for the observed data.

Finally, deriving the analytical form of the density function f^*_{12} yields

$$f^*_{12}(\eta_{1r}, y_2) = \begin{cases} \dfrac{\partial C(F^*_1(\eta_{1r}), F_2(y_2)) f_2(y_2)}{\partial F_2(y_2)} & \text{for} \quad r = 1 \\ \left(\dfrac{\partial C(F^*_1(\eta_{1r}), F_2(y_2))}{\partial F_2(y_2)} - \dfrac{\partial C(F^*_1(\eta_{1r-1}), F_2(y_2))}{\partial F_2(y_2)} \right) f_2(y_2) & \text{for} \quad 1 < r \leq R+1. \end{cases}$$

This will form the basis for the derivation of the penalized log-likelihood function in Section 3.2.

2.2 | Copula GAMLSS

The bivariate copula model in Equation (1) is embedded into the distributional regression framework to model flexibly both the dependence parameter and the marginal distributions. To this end, the response vector $\boldsymbol{y}_i = (y^*_{1i}, y_{2i})'$, $i = 1, ..., n$, is assumed to follow a parametric distribution where potentially all parameters, except of the cut-points, are related to a regression predictor and consequently to covariates. We write the joint conditional density as $f^*_{12}(\vartheta_{1i}, \ldots, \vartheta_{Ki} \mid \boldsymbol{\nu}_i)$, where the vector $\boldsymbol{\nu}_i$ collects any covariates associated with the parameters ϑ_{ki}, $k = 1, ..., K$, of density f^*_{12}. Accordingly, the parameter vector $\boldsymbol{\vartheta}_i = (\theta^*_1, \ldots, \theta^*_R, \vartheta_{1i}, \ldots, \vartheta_{Ki})'$ includes the transformed cut-points $\{\theta^*_r\}$, the location parameter of the first marginal distribution, all other distributional parameters related to the second marginal distribution, and the copula parameter γ_i. The subscript i attached to the parameters is made explicit to stress their potential dependence on individual-level covariates. For the ordinal response, the logit and probit link functions can be applied, while the scale parameter for density f_1 is set to one in order to achieve model identification. The second marginal distribution can be selected from a wide range of options that are available in `GJRM` and listed in Marra and Radice (2017). At the current stage, some of them are not implemented for the mixed ordered-continuous case, but will be made available in the near future. In this paper we only consider one-parameter copulas; some available options are summarized in the supplementary file (Appendix A) although rotated versions are also implemented in `GJRM`. Since the copula parameter γ_i is not directly comparable over different models, we relate it to the Kendall's τ which can be used for interpreting the dependence. It has a one-to-one relationship to the copula parameter for most one-parameter copulas. For optimization and modelling purposes, an appropriate transformation of the copula parameter, γ^*_i, is used in the estimation algorithm as highlighted in the last column of Table A.1 in the supplementary file.

In the spirit of the GAMLSS approach, each distributional parameter of the continuous marginal distribution and the copula parameter is related to an additive predictor via

$$\vartheta_{ki} = h_k(\eta_i^{\vartheta_k}) \quad \text{and} \quad \eta_i^{\vartheta_k} = g_k(\vartheta_{ki}), \quad (6)$$

where $\eta_i^{\vartheta_k} \in \mathbb{R}$ is the predictor associated with distributional parameter ϑ_{ki}, and $h_k = g_k^{-1}$ is a response function mapping the real line into the domain of ϑ_{ki}. The choice of the link functions g_k depends on the distribution of the continuous marginal. For example, the lognormal distribution has location parameter μ_2 and scale parameter σ_2, with the index denoting that they belong to the second marginal. Both parameters are connected via a log link to the predictor. For the copula parameter, the link function depends on the choice of the copula (see Table A.1 in the supplementary file). For example, the inverse hyperbolic tangent is applied for the Gaussian copula. For the ordinal response, the predictor η_{1ri} in Equation (4) can now be represented as $\eta_{ri}^{\mu_1} = \theta_r - \eta_i^{\mu_1}$, where $\eta_i^{\mu_1}$ is specified as in Equation (6) using a logit or probit link.

The general predictor $\eta_i^{\vartheta_k}$, for each parameter of the marginals and the copula parameter, takes on the additive form

$$\eta_i^{\vartheta_k} = \sum_{j=1}^{J_k} s_j^{\vartheta_k}(\boldsymbol{v}_i),$$

where functions $s_j^{\vartheta_k}(\boldsymbol{v}_i),\ j = 1, \ldots, J_k$, can be chosen to model a range of different effects of (a subset) of explanatory variables $\boldsymbol{v}_i$. In particular:

- Linear effects are represented by setting $s_j^{\vartheta_k}(\boldsymbol{v}_i) = v_{ji}^{\vartheta_k}\beta_j^{\vartheta_k}$, where $v_{ji}^{\vartheta_k}$ is a singleton element of $\boldsymbol{v}_i$ and $\beta_j^{\vartheta_k}$ a regression coefficient to be estimated. For the second marginal, an intercept term is also included to represent the overall level of the predictor; this obtained via $s_j^{\vartheta_k}(\boldsymbol{v}_i) = \beta_0^{\vartheta_k}$. For the ordinal equation the intercept is not estimated separately because it is already accounted for by the cut-point θ_r.
- For continuous covariates, nonlinear effects are achieved by including smooth functions $s_j^{\vartheta_k}(\boldsymbol{v}_i)$ represented by penalized regression splines. Ruppert et al. (2003) and Wood (2017) provide various definitions and options for computing basis functions and related penalties.
- An underlying spatial pattern can be accounted for by specifying spatial information such as geographical coordinates or administrative units in $\boldsymbol{v}_i$. Smoothing penalties can account for the neighbourhood structure and ensure that effects are similar for adjacent regions. Rue and Held (2005) interpret this penalty as the assumption that the vector of spatial effects for all regions follows a Gaussian Markov random field.
- If the data are clustered, random effects can be included in the model specification by setting $s_j^{\vartheta_k}(\boldsymbol{v}_i) = \beta_{jc_i}^{\vartheta_k}$, with c_i denoting the cluster the observations are grouped into.

In Section 4, we state explicitly what the predictors for each parameter look like in the context of our application.

3 | ESTIMATION

3.1 | Maximum penalized likelihood

From the analytical expression of the bivariate density f_{12}^* given in Section 2.1, the model's log-likelihood function is derived as

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(\sum_{r\in\mathcal{R}} \mathbb{1}_{\{y_{1i}=r\}}\left(\log\{F_{12.2}(\eta_{1ri}, y_{2i}) - F_{12.2}(\eta_{1r-1i}, y_{2i})\}\right) + \log\{f_2(y_{2i})\}\right), \tag{7}$$

where $\mathbb{1}_{\{\cdot\}}$ is a Boolean operator that takes on value 1 if condition $\{\cdot\}$ is verified, and 0 otherwise. We define

$$F_{12.2}(\eta_{1ri}, y_{2i}) := \frac{\partial C(F_1^*(\eta_{1ri}), F_2(y_{2i}))}{\partial F_2(y_{2i})} \quad \text{with} \quad F_{12.2}(\eta_{1,1-1,i}, y_{2i}) = 0.$$

The log-likelihood function is maximized with respect to the complete vector of regression coefficients $\boldsymbol{\beta} = (\theta_1^*, \ldots, \theta_R^*, \boldsymbol{\beta}^{\vartheta_1}, \ldots, \boldsymbol{\beta}^{\vartheta_K})'$. Each vector of regression coefficients $\boldsymbol{\beta}^{\vartheta_k}$ includes the coefficients for one parameter ϑ_k.

Embedding the model into the distributional regression framework with highly flexible predictors, including regression spline components, typically requires penalization to avoid overfitting. The penalized log-likelihood $\ell_p(\boldsymbol{\beta})$ with ridge-type penalty can be written as

$$\ell_p(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{S}_\lambda\boldsymbol{\beta}, \tag{8}$$

where $\boldsymbol{S}_\lambda$ is a block diagonal matrix consisting of the penalties associated with each model parameter. For un-penalized parameters (like the cut points or categorical covariates) the corresponding block of $\boldsymbol{S}_\lambda$ is set to $\mathbf{0}$. Penalty matrices are associated with smoothing parameters $\lambda = (\lambda_1^{\vartheta_1}, \ldots, \lambda_{J_K}^{\vartheta_K})'$.

3.2 | Parameter estimation using the trust region algorithm

Marra and Radice (2017) proposed maximizing the penalized likelihood in Equation (8) using a trust region algorithm with integrated automatic selection of the smoothing parameters. The trust region algorithm is more stable and faster than line search methods; it also performs better when a function exhibits long plateaus and the current iterate $\boldsymbol{\beta}^{[a]}$ is in that region (Marra & Radice, 2017). The problem with plateaus is that line search methods may search the next step $\boldsymbol{\beta}^{[a+1]}$ far away from the current iterate hence reducing the efficiency of the algorithm. It can also happen that the search is so far away from the current iterate that the evaluation of the objective function is indefinite or not finite. In a trust region algorithm, this is avoided by resolving sub-problem (9) before evaluating the penalized likelihood (8). Therefore, if a problem arises for a given iterate $\boldsymbol{\beta}^{[a+1]}$, the proposed step $\boldsymbol{p}^{[a+1]}$ is rejected, the radius of the trust region adjusted automatically and the optimization computed again.

As in Radice et al. (2016), Marra and Radice (2017) and Klein et al. (2019), the estimation proceeds in two steps:

Step 1. At iteration a, Equation (8) is maximized for a given parameter vector $\boldsymbol{\beta}^{[a]}$ holding $\lambda^{[a]}$ fixed at a vector of values. A trust region algorithm is applied as follows

$$\begin{aligned} \boldsymbol{\beta}^{[a+1]} &= \boldsymbol{\beta}^{[a]} + \underbrace{\underset{\boldsymbol{p}:\|\boldsymbol{p}\|\leq\Delta^{[a]}}{\operatorname{argmin}}\ \breve{\ell}_p(\boldsymbol{\beta}^{[a]})}_{:=\boldsymbol{p}^{[a+1]}}, \\ \breve{\ell}_p(\boldsymbol{\beta}^{[a]}) &:= -\{\ell_p(\boldsymbol{\beta}^{[a]}) + \boldsymbol{p}'\boldsymbol{g}_p^{[a]} + \frac{1}{2}\boldsymbol{p}'\boldsymbol{H}_p^{[a]}\boldsymbol{p}\}, \end{aligned} \tag{9}$$

where the Euclidean norm is denoted by $\|\cdot\|$ and $\Delta^{[a]}$ is the radius of the trust region. The radius is adjusted in each iteration (see Geyer, 2015, for details). The gradient vector at iteration a is given by $\boldsymbol{g}_p^{[a]} = \boldsymbol{g}^{[a]} - \boldsymbol{S}_\lambda \boldsymbol{\beta}^{[a]}$ and $\boldsymbol{H}_p^{[a]} = \boldsymbol{H}^{[a]} - \boldsymbol{S}_\lambda$ is the Hessian matrix. They are both penalized by matrix $\boldsymbol{S}_\lambda$.

The vector $\boldsymbol{g}^{[a]}$ consists of

$$\boldsymbol{g}^{[a]} = \left(\left.\frac{\partial \ell(\boldsymbol{\beta})}{\partial \theta_1^*}\right|_{\theta_1^*=\theta_1^{*[a]}}, \ldots, \left.\frac{\partial \ell(\boldsymbol{\beta})}{\partial \theta_R^*}\right|_{\theta_R^*=\theta_R^{*[a]}}, \left.\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\vartheta_1}}\right|_{\boldsymbol{\beta}^{\vartheta_1}=\boldsymbol{\beta}^{\vartheta_1^{[a]}}}, \ldots, \left.\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\vartheta_K}}\right|_{\boldsymbol{\beta}^{\vartheta_K}=\boldsymbol{\beta}^{\vartheta_K^{[a]}}} \right)'$$

and the elements of the Hessian matrix are

$$\boldsymbol{H}^{[a]l,m} = \left.\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^l \partial \boldsymbol{\beta}^{m\prime}}\right|_{\boldsymbol{\beta}^l=\boldsymbol{\beta}^{l[a]}, \boldsymbol{\beta}^m=\boldsymbol{\beta}^{m[a]}}, \quad l, m = \vartheta_1, \ldots, \vartheta_K.$$

The second-order partial derivatives of the log-likelihood with respect to cut points $\theta_1^*, \ldots, \theta_R^*$ are derived similarly. At each iteration step, the minimization of Equation (9) uses a quadratic approximation of $\ell_p(\boldsymbol{\beta}^{[a]})$, and the solution $\boldsymbol{p}^{[a+1]}$ is chosen such that it falls within a trust region with centre $\boldsymbol{\beta}^{[a]}$ and radius $\Delta^{[a]}$.

Step 2. Holding the parameter vector value fixed at $\boldsymbol{\beta}^{[a+1]}$, the following problem is solved

$$\lambda^{[a+1]} = \underset{\lambda}{\operatorname{argmin}} \|\boldsymbol{M}^{[a+1]} - \boldsymbol{A}^{[a+1]}\boldsymbol{M}^{[a+1]}\|^2 - Kn + 2\,\mathrm{tr}(\boldsymbol{A}^{[a+1]}), \tag{10}$$

where, after defining $\boldsymbol{\mathcal{I}}^{[a+1]} = -\boldsymbol{H}^{[a+1]}$, the key quantities are

$$\boldsymbol{M}^{[a+1]} = \sqrt{\boldsymbol{\mathcal{I}}^{[a+1]}}\boldsymbol{\beta}^{[a+1]} + \sqrt{\boldsymbol{\mathcal{I}}^{[a+1]}}^{-1}\boldsymbol{g}^{[a+1]},$$
$$\boldsymbol{A}^{[a+1]} = \sqrt{\boldsymbol{\mathcal{I}}^{[a+1]}}(\boldsymbol{\mathcal{I}}^{[a+1]} + \boldsymbol{S}_\lambda)^{-1}\sqrt{\boldsymbol{\mathcal{I}}^{[a+1]}},$$

$\mathrm{tr}(\boldsymbol{A}^{[a+1]})$ is the number of effective degrees of freedom (edf) of the penalized model while K is the number of penalized parameters in vector $\boldsymbol{\vartheta}$. The expression in Equation (10) is solved using the method proposed by Wood (2004). The gradient vector $\boldsymbol{g}$ and the Hessian $\boldsymbol{H}$ are obtained as a side product in step 1. Both are analytically derived in a modular fashion for each parameter, see the supplementary file (Appendix C) for details.

Steps 1 and 2 are iterated until they no longer improve the objective function, that is until the following criterion is met:

$$\frac{|\ell(\boldsymbol{\beta}^{[a+1]}) - \ell(\boldsymbol{\beta}^{[a]})|}{0.1 + |\ell(\boldsymbol{\beta}^{[a+1]})|} < 1e^{-0.7}.$$

To obtain the starting values for the parameter vector, a generalized additive model is fitted using `gam()` (Wood, 2017) or a GAMLSS using the `gamlss()` function within the `GJRM` package. A transformed Kendall's τ between the responses is used as a starting value for the copula parameter. Further details on the trust region algorithm and smoothing parameter selection can be found in the supplementary file (Appendix B) while asymptotic considerations on the proposed maximum penalized likelihood estimator are reported in Appendix D.

3.3 | Confidence intervals

At convergence, reliable point-wise confidence intervals are constructed based on Bayesian large sample approximation as shown in Wood (2017) for generalized additive models (GAM):

$$\widehat{\boldsymbol{\beta}} \overset{a}{\sim} N(\boldsymbol{\beta}, -\boldsymbol{H}_p(\widehat{\boldsymbol{\beta}})^{-1}).$$

The result for the Bayesian covariance matrix $\boldsymbol{V}_{\boldsymbol{\beta}} = -\boldsymbol{H}_p^{-1}$ is an alternative to the frequentist covariance matrix $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = -\boldsymbol{H}_p^{-1}\boldsymbol{H}\boldsymbol{H}_p^{-1}$. For unpenalized models, the two matrices are equal. The application of the Bayesian framework to the GAM or copula GAMLSS context follows the notion that penalization in the estimation implicitly assumes certain prior beliefs about the model's features (Wahba, 1978). In this view, a normal prior for the parameter vector $\boldsymbol{\beta}$, $f_{\boldsymbol{\beta}} \propto \exp(-1/2\boldsymbol{\beta}'\boldsymbol{S}_{\lambda}\boldsymbol{\beta})$, means that wiggly models are less likely to occur than smoother ones (Wood, 2006). Marra and Wood (2012) give a full justification for using $\boldsymbol{V}_{\boldsymbol{\beta}}$ and show that it gives close to across-the-function frequentist coverage probabilities. In fact, it includes both bias and variance components in a frequentist sense, which is not the case for $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$.

To obtain intervals for non-linear functions of the model parameters (e.g. the Kendall's τ), Radice et al. (2016) propose to simulate from the posterior distribution of $\boldsymbol{\beta}$ using the following procedure:

Step 1 Draw M random vectors $\tilde{\boldsymbol{\beta}}_m$, $m = 1, \ldots, M$, from $\mathcal{N}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{V}}_{\boldsymbol{\beta}})$.
Step 2 Calculate M realizations of the function under consideration, say $R(\tilde{\boldsymbol{\beta}_m})$.
Step 3 Calculate the $(\zeta/2)$th and $(1-\zeta/2)$th quantile of the realizations where ζ is typically set to 0.05. The confidence interval is then constructed as $CI_{1-\zeta} = [R(\tilde{\boldsymbol{\beta}}_m)_{\zeta/2}, R(\tilde{\boldsymbol{\beta}}_m))_{1-\zeta/2}]$.

A value of M equal to 100 typically produces reliable results although it can be increased if more precision is required.

3.4 | Simulation study

To evaluate the effectiveness and implementation of the proposed methodology, we conducted a simulation study with four scenarios that differ in terms of the continuous marginal distribution and the copula specification. All four scenarios are assessed using sample sizes of $n = 1000$, 3000 and 10,000. Data generating process and detailed results are both reported in the supplementary file (Appendix E). The results show that our approach is able to capture the effect of both linear and non-linear covariates fairly well, and that the performance improves significantly as the sample size increases. In addition to recovering the simulated coefficients, we calculated the Akaike's information criterion (AIC) for every simulation of the bivariate model and the corresponding independence model. The share of runs in which the bivariate model had a smaller AIC was 1. This provides further evidence of the ability of our model to identify dependence between the responses, if this is indeed required by the data generating process. We refrained from running detailed simulations regarding the selection of marginal distributions and/or copulas since these have been considered before in the literature on copula GAMLSS, albeit limited to the case of two continuous marginal distributions (e.g. Marra & Radice, 2017; Radice et al., 2016).

4 | MULTIDIMENSIONAL POVERTY IN INDONESIA

4.1 | The IFLS dataset

In this section we analyse poverty dimensions in a bivariate copula model and we identify (1) the determinants of the income–education relation, (2) its spatial distribution and (3) groups at risk of being both consumption and education poor. To do so, we rely on the most recent wave (IFLS 5) of the Indonesian Family Life Survey (IFLS). The IFLS is a publicly available, longitudinal survey on individual, household and community level that is designed to study the health and socioeconomic situation of Indonesia's population. The first wave was implemented in 1993 and covered individuals from 7224 households representing 83% of the population from 13 of 27 Indonesian provinces (Strauss et al., 2016). The sample was drawn by stratifying the population on provinces and urban/rural areas before randomly selecting enumeration areas and households within the strata. Due to a large number of split-off households the sample grew up to 16,204 households interviewed in IFLS 5.

In the IFLS, individuals of an IFLS-household older than 15 years were asked to fill in an 'adult individual book' containing questionnaires on subjects such as income, education, employment and subjective health. We use the level of education as the ordinal response variable, and income as the continuous response. Education is proxied by the highest educational institution attended and can take on five different levels: 1 'no schooling', 2 'primary school', 3 'middle school', 4 'high school' and 5 'tertiary education'. In analyses for developing countries, income is often proxied by expenditures for consumption. Expenditures are calculated at the household level and then divided by the number of household members. Our income variable is thus precisely expenditures per capita. Due to different price levels in the provinces, we used the province specific minimum wage to adjust expenditures across provinces. Data on the individuals from the 'adult individual book' are extracted and merged with relevant information on the household head, such as gender and education, and complemented with information on the household's location, such as province or whether the household lives in an urban area. We only included complete cases and individuals from the age of 18 as most of them already attained or are studying towards their highest education level. The final dataset contains 32,884 individuals.

4.2 | Model building

Applying flexible bivariate copula GAMLSS requires the researcher to decide on the specification of multiple parameters, on the form of the continuous marginal distribution and of the copula.

4.2.1 | Continuous marginal distribution

In line with, Klein et al. (2015) and Marra and Radice (2017), we propose to use normalized quantile residuals for selecting the continuous marginal. This allows us to assess graphically the appropriateness of the chosen distribution by examining separate univariate models. A normalized quantile residual $\widehat{q}_{mi}$, $m = 1, 2$, for the second (continuous) marginal is defined as:

$$\widehat{q}_{2i} = \Phi^{-1}\{\widehat{F}_2(y_{2i})\} \quad \text{for } i = 1, \ldots, n,$$

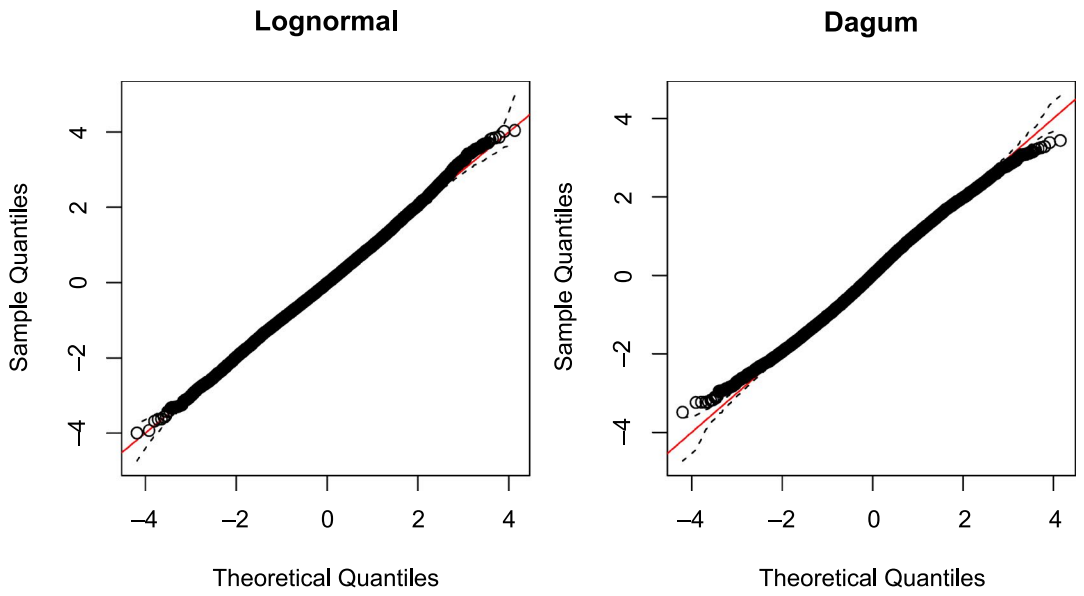


FIGURE 1 Normal QQ-plots for the univariate income model and different distributions with 95% reference bands [Colour figure can be viewed at wileyonlinelibrary.com]

where $\widehat{F}_2(\cdot)$ is the estimated marginal CDF for the continuous response component, and $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution. If $\widehat{F}_2$ is close to the true distribution, $\widehat{q}_{2i}$ approximately follows the standard normal distribution. Quantile residuals are fairly robust to the specification of the distribution parameters (Klein et al., 2015; Marra & Radice, 2017).

We fit univariate models and select the model by inspecting the corresponding QQ-plots. Good distribution candidates for income and expenditure are generally the lognormal distribution, the Singh–Maddala distribution and the Dagum distribution (e.g. Kleiber & Kotz, 2003). Figure 1 shows the QQ-plots for the univariate income model and all potential covariates using the lognormal and the Dagum distributions. Fitting the model with the Singh–Maddala distribution leads to convergence failure, which may signal an inappropriate choice of the marginal distribution. The QQ-plots suggest an appropriate fit for the lognormal distribution. We will therefore use this distribution as first building block of the model specification. Note that once the final bivariate model is built, the QQ-plot for the continuous margin is re-examined. However, we find that the plot (shown in the supplementary file, Appendix F) looks almost identical to Figure 1 (left panel). This indicates that a good fit for the continuous margin of the proposed copula model has been obtained.

4.2.2 | Variable selection

For the specification of the link function of the ordinal response, as well as for the variable selection in the bivariate model and the choice of the copula function, the AIC and the Bayesian information criterion (BIC) can be used. These are defined as

$$\begin{aligned}\text{AIC} &:= -2\ell(\widehat{\boldsymbol{\beta}}) + 2\text{edf},\\ \text{BIC} &:= -2\ell(\widehat{\boldsymbol{\beta}}) + \log(n)\text{edf},\end{aligned}$$

where $\ell(\widehat{\boldsymbol{\beta}})$ is the log-likelihood of the bivariate model evaluated at the penalized parameter estimate and $\text{edf} = \text{tr}(\widehat{\boldsymbol{A}})$ as defined in Section 3.2. Theoretical knowledge about the problem at hand facilitates the variable selection procedure by pre-selecting candidate predictors. Radice et al. (2016) also suggest to start with a model specification where all distributional and the association parameter depend on all covariates. In case the algorithm does not converge, an instance that often indicates that the sample size is too small for the model's complexity, they recommend trying out a series of more parsimonious specifications. To test smooth components for equality to zero, we have adapted the results of Wood (2017) to the current context.

To fit the bivariate model, we specify an equation for each distributional and the copula parameter as follows. We start with a set of variables selected according to economic reasoning. Note that in income or expenditure equations, household size is often used as a covariate in addition to number of children and elderly. However, we do not wish to separate the child effect in a 'pure' child effect and children as additional household member effect. Moreover, the outcome variable is already adjusted for household size. Religion is included because it defines minority groups. While education at individual level is part of the response vector, the level of education of the household head is included as a control in the predictor for capita income. For the bivariate model, we fit a full specification for the location parameter of each marginal distribution, namely $\mu_{1\,educ}$ and $\mu_{2\,inc}$, and perform variable selection using the AIC for the scale parameter of the second marginal, $\sigma_{2\,inc}$, and for the copula parameter, γ. The copula parameter is linked via the inverse hyperbolic tangent to the predictor. A backwards selection procedure is applied for $\sigma_{2\,inc}$ given a full specification for γ, and subsequently a second backwards selection is performed on γ given the reduced model for $\sigma_{2\,inc}$. This excludes only four variables for the scale predictor and two variables for the copula parameter specification. We therefore write the final model specification as:

$$
\begin{aligned}
\eta^{\mu_1}_{educ} &= \theta_r - \{s(age) + \beta^{\mu_1}_1 \cdot hhmarstat + \beta^{\mu_1}_2 \cdot hhmale + \beta^{\mu_1}_3 \cdot urban + \\
&\qquad \beta^{\mu_1}_4 \cdot num\_child + \beta^{\mu_1}_5 \cdot elderly + \beta^{\mu_1}_6 \cdot relig\} \\
\eta^{\mu_2}_{inc} &= \beta^{\mu_2}_0 + s(age) + \beta^{\mu_2}_1 \cdot hhmarstat + \beta^{\mu_2}_2 \cdot hhmale + \beta^{\mu_2}_3 \cdot urban + \\
&\qquad \beta^{\mu_2}_4 \cdot num\_child + \beta^{\mu_2}_5 \cdot elderly + \beta^{\mu_2}_6 \cdot relig + \beta^{\mu_2}_7 \cdot hheduc + s(prov) \\
\eta^{\sigma_2}_{inc} &= \beta^{\sigma_2}_0 + s(age) + \beta^{\sigma_2}_1 \cdot hhmarstat + \beta^{\sigma_2}_2 \cdot num\_child + \beta^{\sigma_2}_3 \cdot relig + s(prov) \\
\eta^{\gamma} &= \beta^{\gamma}_0 + s(age) + \beta^{\gamma}_1 \cdot hhmarstat + \beta^{\gamma}_2 \cdot urban + \\
&\qquad \beta^{\gamma}_3 \cdot num\_child + \beta^{\gamma}_4 \cdot elderly + \beta^{\gamma}_5 \cdot hheduc + s(prov).
\end{aligned}
$$

Continuous variables enter the equations with smooth non-parametric effects `s()` represented via thin plate regression splines with 10 bases and second-order derivative penalties. Spatial effects of the provinces and their neighbourhood structure are modelled using Gaussian Markov random fields. We choose to model the spatial effect at the province level since minimum wages are set by each province; this affects individual wages and thus the expenditure measure as well (Hohberg & Lay, 2015). Table F.1 in the supplementary file displays full names and levels of the selected variables.

4.2.3 | Ordinal model

The ordinal outcome education is fitted using an ordered model. Table 1 compares the AIC and BIC between a probit and a logit link of the bivariate model using the lognormal as the continuous

TABLE 1 AIC and BIC of bivariate ordered-continuous model using the logit and probit links

	AIC	BIC
Logit	1,076,985	1,078,081
Probit	1,077,371	1,078,466

marginal and a Gaussian copula. Both AIC and BIC favour the logit model for the first marginal. This choice is mostly stable over the range of tested copulas.

4.2.4 | Choice of the copula

For the copula selection, a good starting point would be the use of a Gaussian copula and then consider all consistent alternatives depending on the direction of the dependence (Klein et al., 2019; Radice et al., 2016). Again, AIC and BIC can help choosing among several candidate copulas.

Starting off with the Gaussian copula, we obtain an average value for the copula parameter (with 95% confidence interval in brackets) of $\gamma = 0.163$ (0.103, 0.223). Building on this finding, we test a range of suitable alternative candidates. After checking convergence, we can only eliminate the un-rotated Joe and Clayton copula. The remaining candidates are compared using the AIC and BIC; see Table 2 for the results. The AIC and BIC indicate that a Gaussian copula should be used for our model, and all copula models should be favoured over the independence model.

TABLE 2 AIC and BIC for different copula specifications

	AIC	BIC
Gaussian	1,076,985	1,078,081
F	1,077,003	1,078,109
FGM	1,077,039	1,078,140
PL	1,076,997	1,078,101
AMH	1,077,054	1,078,160
C0	1,077,192	1,078,296
C180	1,077,094	1,078,142
J0	1,077,171	1,078,161
J180	1,077,246	1,078,348
G0	1,077,052	1,078,107
G180	1,077,095	1,078,173
T3	1,078,531	1,079,634
T5	1,077,608	1,078,711
T10	1,077,147	1,078,248
Independence	1,077,585	1,078,390

Note: Abbreviations correspond to Frank, Farlie-Gumbel-Morgenstern, Plackett, Ali-Mikhail-Haq, Clayton, rotated Clayton (180 degrees), Joe, rotated Joe (180 degrees), Gumbel, rotated Gumbel (180 degrees), Student's t with 3, 5, and 10 degrees of freedom, respectively.

Using the Gaussian copula suggests that the dependence between per capita expenditures and education is symmetric with asymptotically independent extremes.

4.3 | Model evaluation

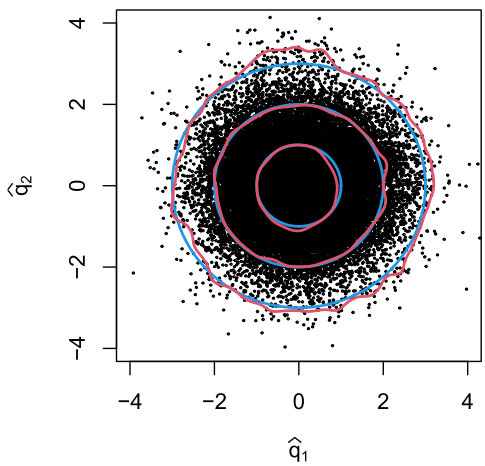
After deciding on the marginal distributions, the copula and the covariates by comparing different candidates, we check the final bivariate model. To this end, we use a multivariate generalization of the quantile residuals introduced in Section 4.2 that was proposed by Kalliovirta (2008).

Multivariate quantile residuals for two continuous responses are defined as

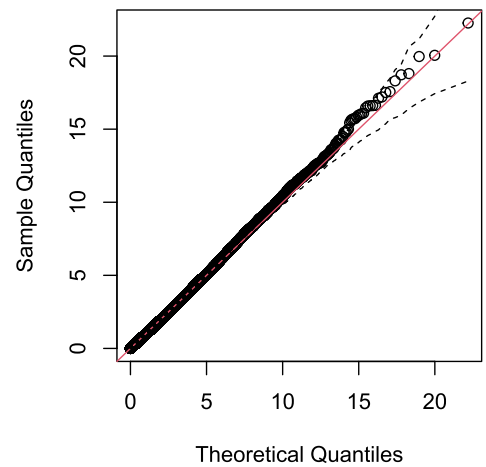
$$\hat{\boldsymbol{q}}_i = \begin{pmatrix} \hat{q}_{1i} \\ \hat{q}_{2i} \end{pmatrix} = \begin{pmatrix} \Phi^{-1}(\hat{F}_1(y_{1i})) \\ \Phi^{-1}(\hat{F}_{2|1}(y_{2i}|y_{1i})), \end{pmatrix}$$

where $\hat{F}_{2|1}$ is the (estimated) conditional CDF of Y_2 given Y_1. In our case, the first marginal is discrete such that we resort to randomized quantile residuals where uniformly distributed random variables on the interval corresponding to cumulated probabilities are plugged into $\Phi^{-1}(\cdot)$. If the model is correctly specified, then $\hat{\boldsymbol{q}}_i$ approximately follows a bivariate standard normal distribution. This approximation follows from the consistency of the parameter vector $\hat{\boldsymbol{\beta}}$ (Radice et al., 2016, and the supplementary file).

The contour plot for the bivariate model in Figure 2(a) shows the density of the quantile residuals $\hat{\boldsymbol{q}}_i$ by means of a multivariate kernel density estimator. This estimated density is compared to the density of the standard normal distribution. The contour lines of both densities are close to each other indicating a good fit of the bivariate copula model.



(a) Contour plot of multivariate quantile residuals. The red lines indicate the density of the quantile residuals estimated by a multivariate kernel density estimator. The blue circles are the contour lines of the density of the standard normal distribution with radius 1, 2 and 3.



(b) QQ-plot depicting the sum of the squared elements of the multivariate quantile residuals with 95% reference bands.

FIGURE 2 Multivariate quantile residuals of the bivariate model [Colour figure can be viewed at wileyonlinelibrary.com]

In Figure 2(b), the sum of the squared elements of the multivariate quantile residuals is considered, that is $\hat{\boldsymbol{q}}_i'\hat{\boldsymbol{q}}_i = \hat{q}_{1i}^2 + \hat{q}_{2i}^2$, where $\hat{\boldsymbol{q}}_i$ is the multivariate quantile residual for the ith individual. Since $\hat{q}_{1i} \overset{a}{\sim} \mathcal{N}(0, 1)$ and $\hat{q}_{2i} \overset{a}{\sim} \mathcal{N}(0, 1)$, it follows that $\hat{\boldsymbol{q}}_i'\hat{\boldsymbol{q}}_i \overset{a}{\sim} \chi^2(2)$ which is assessed in the QQ-plot in Figure 2(b). The reference bands are obtained by repeatedly simulating from a $\chi^2(2)$ distribution. We draw 100 samples for each individual and compute the 2.5% and 97.5% quantiles across the sorted samples. The plot supports our model choice.

4.4 | Results

This section demonstrates the results that poverty researchers can derive from fitting a copula GAMLSS model for the joint analysis of two inter-related poverty dimensions. We summarize briefly the covariate effects on the marginals before taking a closer look at the dependence structure and risk groups. In this way, we are able to answer questions on how dependence and risk are affected by a household's location or composition.

4.4.1 | Effects of covariates on the marginal distributions

The full list of effects on each parameter of the marginals and on the copula parameter is included in the supplementary file (Appendix F). It should be stressed that the effects are subject to a *ceteris paribus* interpretation. For example, more schooling is associated with more income, with tertiary education having the highest effect. Households with more children or elderly people living in the household have on average a lower income per capita and less education (except for the effect of two or three elderly on education which is positive). Urban households are associated with both better income and better education. A little surprising is that living in a household where the head is married is correlated with a reduction in income and education compared to not (yet) married households (Table F.3 and F.2 in the supplementary file). One possible explanation is that non-married households compared to married households include a larger share of young, single persons that do not need to share their income and have a comparatively higher level of education. Non-Muslim households are associated with higher income per capita and more education (only Christians) although they represent a minority in Indonesia. For a male household head, the effects are not as expected since it is negative for income but positive for education. For the second parameter of the continuous margin, that is the scale parameter for the income equation, the number of children and a marital status other than not married have negative effects while the effects of elderly and other religions compared to Islam are positive. All of the covariates, except of the dummies for a male household head and belonging to the Hindu religion, have a positive effect on the copula parameter although not all effects are significant.

Age is modelled in a non-linear way and Figure 3 displays the smooth effects of age on each parameter. Education attainments are lower for higher ages which can be explained by the education expansion that Indonesia has undergone since the 1970s and younger individuals benefited from. For example, between 1974 and 1978 over 61,000 primary schools were built. In 1984, compulsory education was set to 6 years which was extended to 9 years in 1994 (Akita, 2017; Duflo, 2004). The effect of income on the location parameter is inverted u-shaped until the age of 60 with a peak around the age of 40. After 80, the confidence intervals become very wide due to a lower number of observations in this age span and the effect is thus less clear. The effects on the scale parameter are around zero. For the copula parameter, the age effect indicates that the

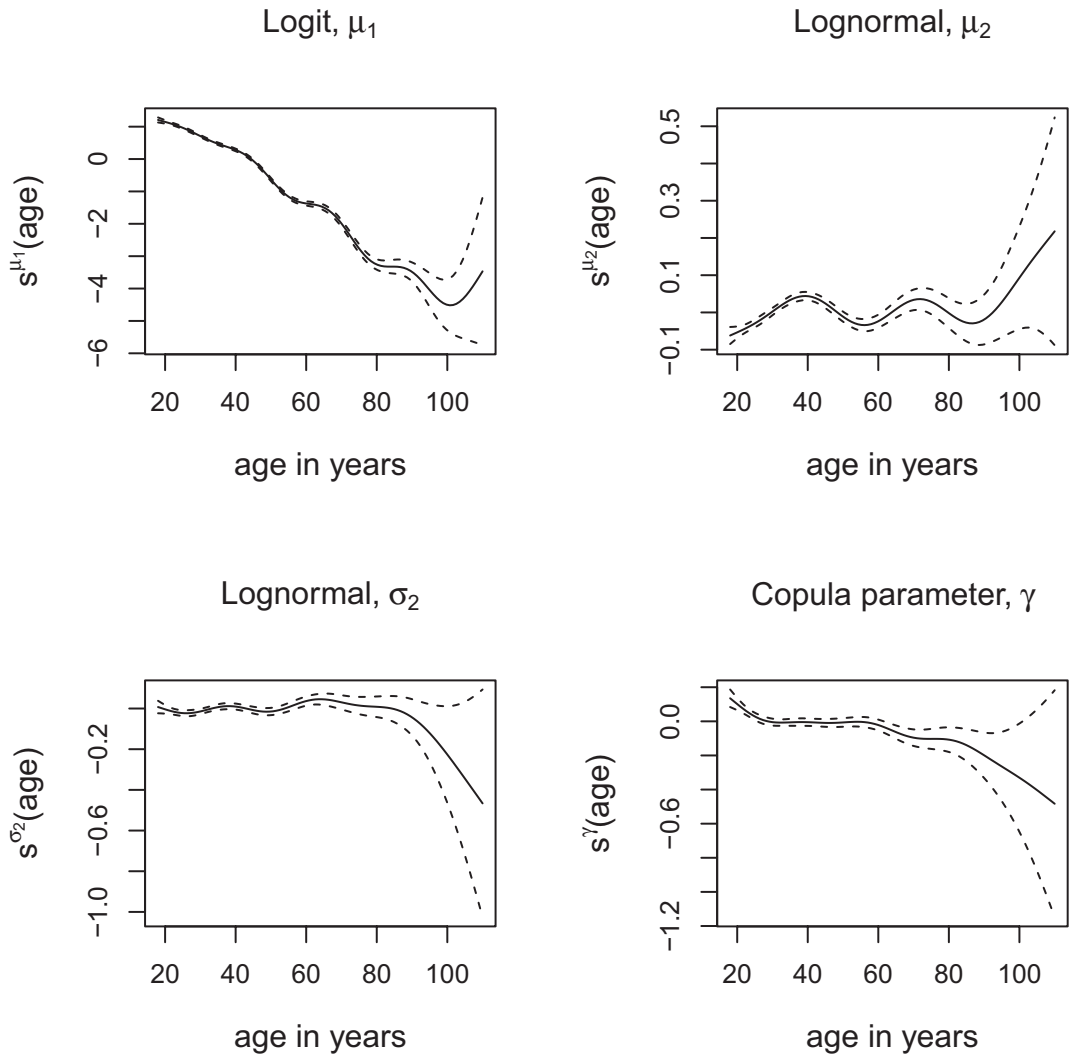


FIGURE 3 Estimated smooth functions of age and respective point-wise 95% confidence intervals

dependence is decreasing for individuals up to their mid 30s and stays around zero afterwards until it decreases again after the age of 60.

Figure 4 shows the effect of the underlying spatial pattern on the parameters μ_2 and σ_2 of the continuous response, and on the copula parameter γ. Households located in provinces of Java seem to have higher income per capita compared to the observed provinces in Sumatra, Borneo and Sulawesi. Provinces with a negative effect on the location parameter have higher effects on the scale parameter except for Borneo whose scale effect is negative.

4.4.2 | Dependence structure

Dependence structure can be represented via contour plots for specific covariate combinations. For example, we focus here on the location of the household (urban/rural and province) as they are the significant drivers of the copula parameter, while the remaining plots can be found in

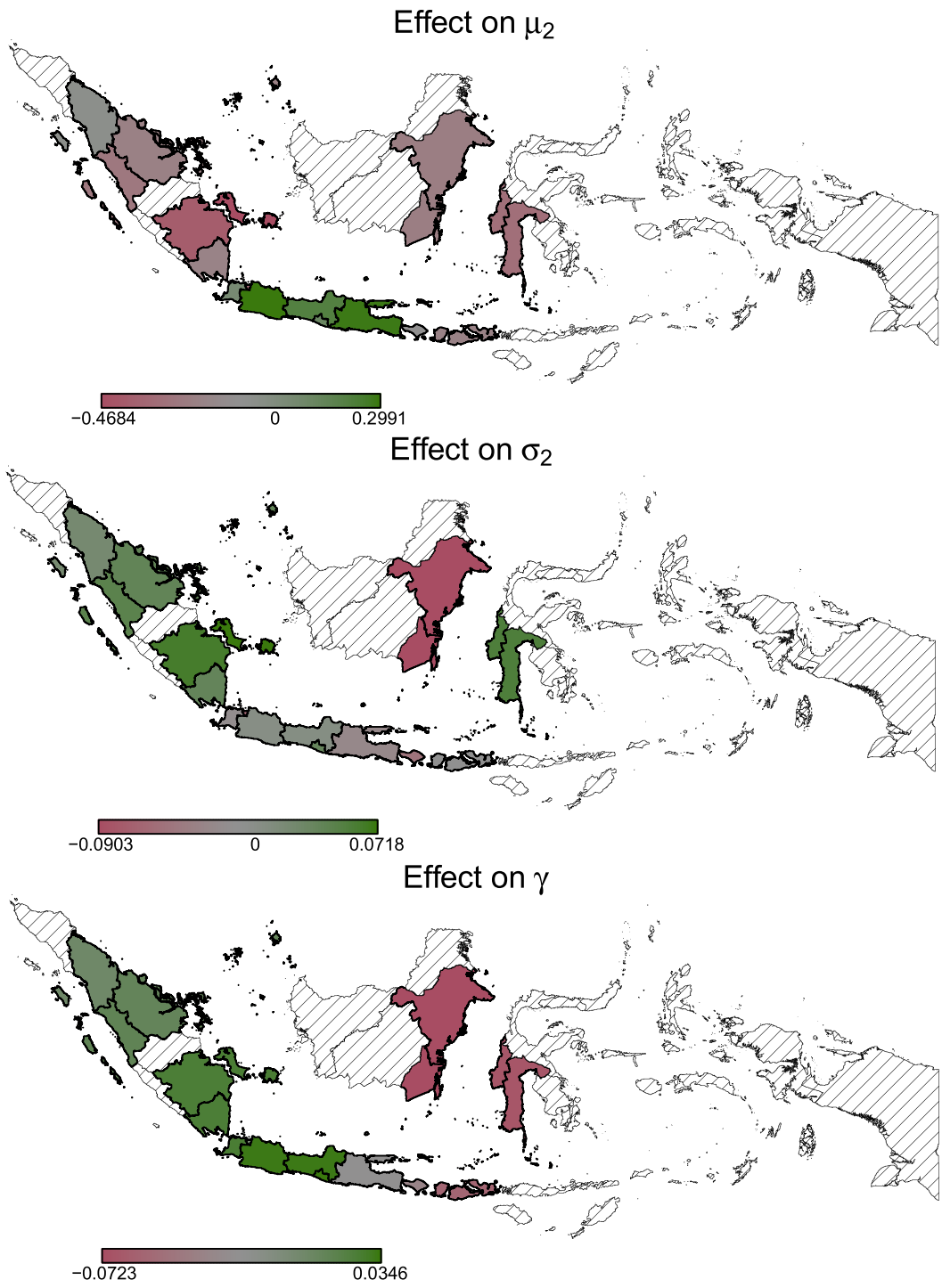


FIGURE 4 Spatial effects of the provinces on the distribution parameters of income and on the copula parameter [Colour figure can be viewed at wileyonlinelibrary.com]

the supplementary file (Appendix F). For comparison purposes, we include provinces with the highest frequencies in the dataset but select only one of Java's provinces. To compare the dependence structure across different locations, we create an example of typical individual whose

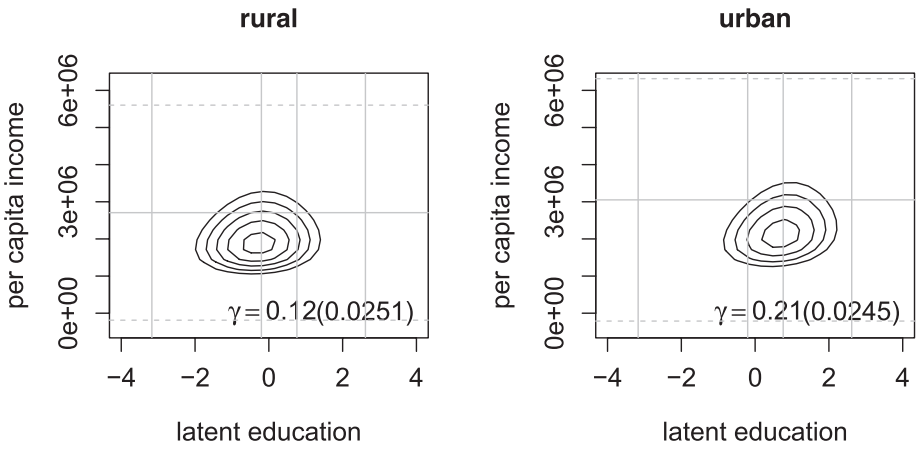
characteristics, other than the one under consideration, are set to their mean value or to their most frequent observation. The only exception is the education of the household head which is set to the second most frequent observation. In fact, for a household head with 'high school' degree the dependence is a bit more pronounced, hence we selected this level for demonstration purposes. This is the covariates' combination that we refer to as an 'example individual' henceforward.

Figure 5(a) shows that the dependence is stronger for individuals in urban households compared to rural households. One reason might be that average education levels are lower in rural areas (horizontal axis) while, at the same time, high paid job opportunities are limited in a rural environment, resulting in more equal incomes compared to an urban environment. Figure 5(b) compares the dependence structure across selected provinces. The dependence seems weakest in the province of Nusa Tenggara Barat, which is one of the poorest provinces in Indonesia. It is surprising that the average per capita consumption (horizontal straight line) of Jakarta is about the same level than that one of Nusa Tenggara Barat. Most likely this is due to the high price level in Jakarta and the deflation measure applied which scaled down our expenditure measure. Although the copula coefficient for Jakarta is similar to Jawa Timur, the latter has more variation in incomes and the contour levels lie further apart. Considering all—and not just the selected provinces—the value of the copula parameter for the example individual ranges from 0.16 for Kalimantan Timur and Kalimantan Selatan up to 0.27 for Yogyakarta.

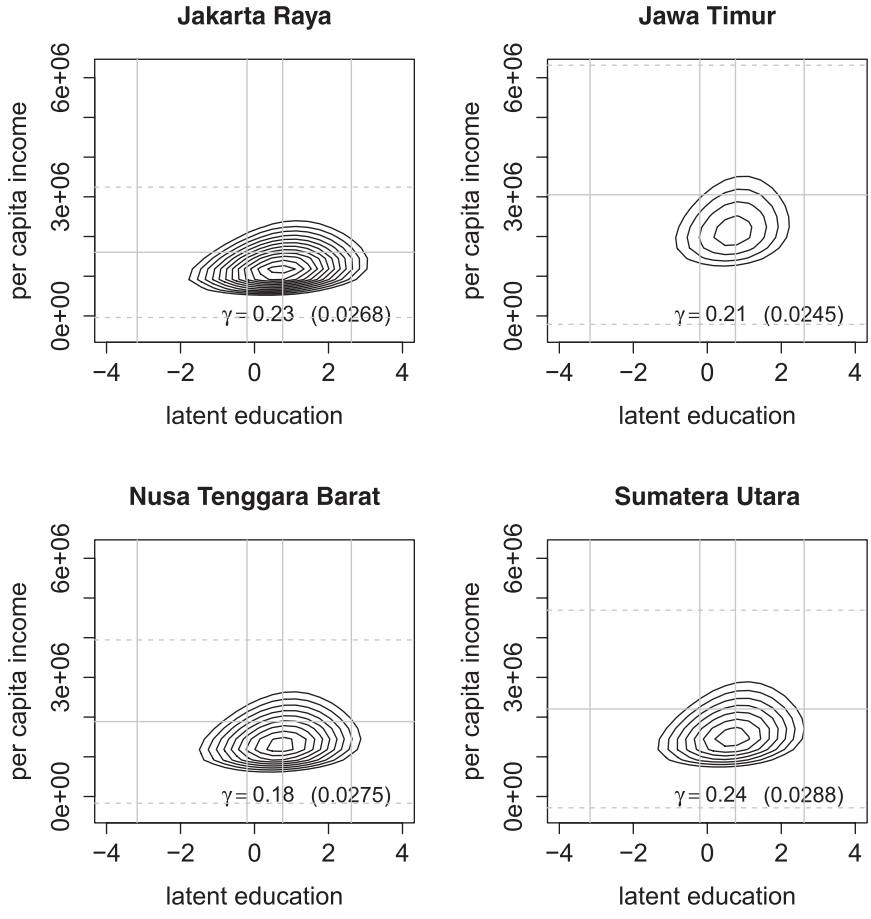
A policy maker might be interested to know in which locations each individual in the dataset, and not the example individual, have higher dependencies in order to efficiently design policy strategies. Although the lower panel of Figure 4 shows the effect of the provinces on the copula parameter, it might be more helpful for interpretation to transform it into the Kendall's τ, an association measure that takes on values in the interval $[-1, 1]$. Each individual with his/her specific covariates' combination is related to an individual-specific τ. One way to present the differences across provinces is to average the τ over all individuals in a particular province. This is shown in Figure 6. The Kalimantan Selatan (South Borneo) is the province with the lowest average of Kendall's τ with a value of 0.0468 and Kepulauan Riau (Riau Islands, northwest of Borneo) has a value of 0.1467 which is the highest average value that also indicates spatial heterogeneity in the strength of the dependence. The provinces of Sumatra seem to have higher dependence between income and education than provinces in Borneo or Sulawesi. Interestingly, for Java and its neighbouring smaller islands on the east, the dependence seems to decrease from west to east. It is worth noting that Figure 6 shows the association between the latent education variable and the continuous income outcome. We also checked the association on the ordinal level of education attained via a simulation-based approach. To this end, we first simulated 1000 response pairs for each individual from its estimated covariate-specific latent bivariate distribution. Using the estimated cut-points, the simulated latent response is transformed into the ordinal category. Kendall's tau is then calculated between the ordinal category and the simulated continuous outcome and averaged over provinces. The resulting map is almost identical map and therefore not included here.

4.4.3 | Joint probabilities

Other results that we can derive from a copula GAMLSS model are the joint probabilities for different sub-groups of individuals. In particular, we calculate the probability for the example



(a) Contour plots for an example individual in an urban or rural household in the province of Jawa Timur.



(b) Contour plots for an example individual in an urban household in different provinces.

FIGURE 5 Contour plots for (education, income)' and a Gaussian copula by households' location. Contour lines of densities are at levels from 0.00000005 to 0.00000025 in 0.00000001 steps. The vertical straight lines represents the cut-off values for the education categories, horizontal straight lines are the consumption average, and dashed horizontal line are at two standard deviations around this average

individual of being poor in both the education and income dimensions. As an example, we focus here on household location and household composition. To define poverty, we classify individuals that have attained at most primary education as education poor, and we set a relative poverty line of 60% of the median of the per capita expenditures' unique values. Note that Indonesia also has a national absolute poverty line that is, however, based on a different expenditure measure than the one we constructed from the IFLS data. This motivated us to use a relative poverty line as defined above. An individual that is poor in both dimensions has expected values below each of these thresholds.

One of the sub-groups is set as the base category and the other sub-groups are compared to this base category. Figure 7 shows that the probability for being poor in both dimensions is two times higher for the example individual in a rural household compared to the same individual in an urban household. Compared to Jawa Timur the joint probabilities of Jakarta Raya, Nusa Tenggara Barat and Sumatera Utara are about eight times, six times and four times higher respectively. Not surprisingly, the risk of being poor in both dimensions increases with the number of children and elderly in the household.

4.4.4 | Vulnerability to poverty

The higher the dependence between education and income, the higher the chances that we miss some individuals at risk by looking only at the marginal distributions of each poverty dimension. To identify the individuals at risk, we calculate the probabilities of being poor in the two dimensions for each individual in the dataset, first using the baseline independence model and then

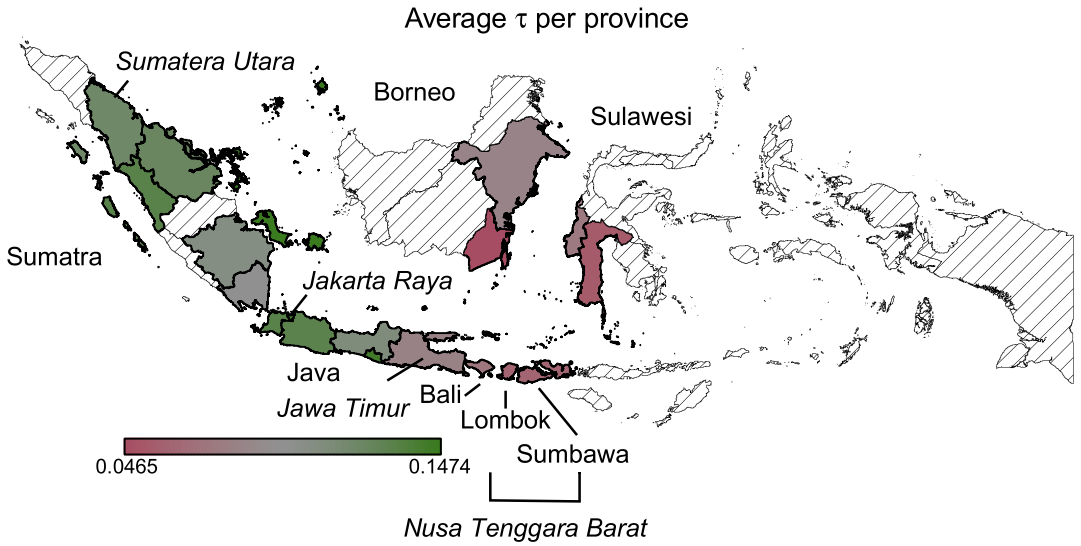


FIGURE 6 Kendall's τ for each individual averaged within provinces [Colour figure can be viewed at wileyonlinelibrary.com]

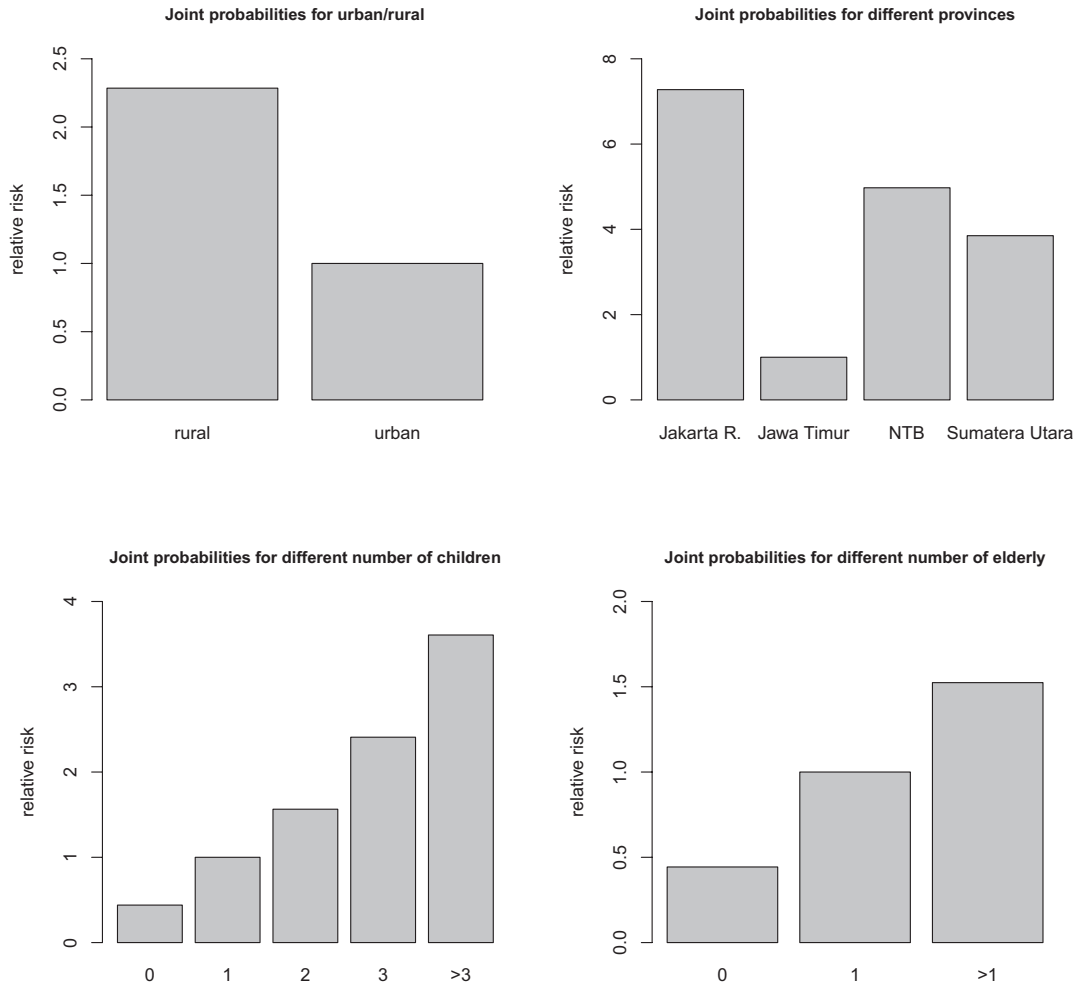


FIGURE 7 Relative joint poverty risks of an example individual differentiated by household location and household composition. Baseline categories are urban, Jawa Timur, one child and one elderly, respectively, and are set to one. The abbreviation NTB denotes the province of Nusa Tenggara Barat

using the preferred copula model. A vulnerability threshold is arbitrarily set at a probability of 0.1 for simplicity. All individuals with a probability of being poor above this threshold are declared as vulnerable. We then compare the individuals that are identified as vulnerable by the copula and the independence models to their actual poverty status, and we calculate the sensitivity or true positive rate. Results are displayed in Table 3. We find that the copula model has better specificity for two-dimensional poverty than the independence model although the difference is fairly small.

TABLE 3 Sensitivity of the copula and independence models

	Copula	Independence
Income dimension	0.88	0.88
Both dimensions	0.66	0.63

Note: Individuals that are declared vulnerable are compared to their actual poverty status.

4.4.5 | Discussion of the results

The main advantage of applying a copula GAMLSS in the context poverty analysis is that we are able to analyse the dependence between poverty dimensions in greater details compared to studies that are limited to the measurement of such dependence. Even though the estimated dependence is not very strong, once we control for covariates in the marginals, we consider this as an interesting outcome. We could further identify heterogeneities in the strength of the dependence between income and education, and poverty risk with respect to a household's location. For instance, an example individual in a rural household exhibits lower dependencies compared to the same individual in an urban household. One explanation might be that opportunities in rural areas are limited and thus the individual's education level has a smaller influence on per capita expenditures.

There is also a strong spatial heterogeneity between provinces. High dependencies can be found in the north-west of the country while the values decrease for the more central provinces. The argument that such low dependence is due to limited opportunities could also apply to the province of Nusa Tenggara Barat in which an example individual showed little average dependence and lower dependence compared to other provinces. Nusa Tenggara Bararat is one of the poorest provinces of the country, with a low GDP per capita and in which agriculture and fishery are the most important industries. On the other hand, provinces such as Jakarta and Kalimantan Timur, that have the highest per capita GDP out of all the provinces in Indonesia, have higher probabilities of being poor in both dimensions compared to most other provinces. These probabilities are calculated for an example individual that has average characteristics and a high school degree. Possible reasons for the discrepancy between rich provinces and high relative poverty risks might be that in Jakarta income and consumption are very unequally distributed, whereas in East Kalimantan the high GDP is a result of high natural resource exploitation which yields little benefit for the population. For example, Bhattacharyya and Resosudarmo (2015) found that growth in the mining sector has had no effect on poverty and inequality in Indonesia.

5 | CONCLUSION

Although poverty is conventionally regarded a multidimensional phenomenon, regression analyses in this research areas either examine each poverty dimensions separately or use a scalar index, such as the Multidimensional Poverty Index, as response variable. Both approaches neglect the dependence between poverty dimensions. This paper presented an alternative statistical model for an in-depth poverty study that accounts explicitly for this dependence and its determinants. This is an important feature of the model because it is critical to understand what drives the dependence between poverty dimensions: high dependencies can in fact explain persisting poverty.

For this kind of poverty analysis, we proposed a bivariate copula GAMLSS which relates each distributional parameter of the marginals and of the copula parameter to flexible covariate effects. Since poverty analyses often include one monetary measure, such as income or consumption, and some ordinal measure, such as education level or health status, we extended the class of copula GAMLSS to incorporate ordinal responses. This extension has been incorporated in the R-package `GJRM`.

We used data from Indonesia to show how copula GAMLSS can be applied to a poverty analysis. The model identified the number of elderly people and children in the household, the highest

education attained by the household head, and the household's location as risk factors for low income or poor education, or both. The gender of the household head, whether or not he/she belongs to a minority religion and is widowed or separated, has shown less (or opposite) influence than expected. We did not find evidence for strong tail dependencies between education and expenditures after conditioning on the covariates in the marginals and in the copula parameter.

Focusing more on the household's location, we found that an example individual in a rural household exhibits lower dependencies compared to the same individual in an urban household. This is potentially due to limited employment opportunities for highly educated individuals in rural areas. We identified a strong spatial heterogeneity regarding poverty risk and the strength of the dependence. In particular, high dependencies were found in the north-west of Indonesia, while in central provinces we reported lower dependencies. For Jakarta and Kalimantan Timur we found a discrepancy between being rich in terms of GDP and exhibiting high relative poverty risks; this may potentially indicate unequally distributed economic gains.

Thanks to the flexibility of our approach, the analysis of the dependence between poverty dimensions and the other results documented in this paper can be derived consistently by estimating only one model. We advocate therefore to include a copula GAMLSS in the toolbox of poverty researchers alongside the other statistical strategies. Further applications in the context of poverty may comprise analysing the drivers and the spatial patterns of inter-generational poverty persistence or upward social mobility. On the methodological side, future research can be directed at combining copula GAMLSS with experimental or quasi-experimental methods to evaluate Indonesia's poverty policies at the micro level. Extensions beyond the bivariate case are linked to the computational aspects of multivariate copula functions. As the dimension increases, the implementation of these models becomes more numerically and technically demanding and the interpretation of the results will be challenging.

ACKNOWLEDGMENTS

Open access funding was enabled and organized by ProjektDEAL.

ORCID

Maike Hohberg http://orcid.org/0000-0003-2860-1863
Giampiero Marra http://orcid.org/0000-0002-9010-2646

REFERENCES

Akita, T. (2017) Educational expansion and the role of education in expenditure inequality in indonesia since the 1997 financial crisis. *Social Indicators Research*, 130(3), 1165–1186.

Alkire, S. & Fang, Y. (2019) Dynamics of multidimensional poverty and uni-dimensional income poverty: an evidence of stability analysis from China. *Social Indicators Research*, 142, 25–64.

Alkire, S., Foster, J.E., Seth, S., Santos, M.E., Roche, J.M. & Ballon, P. (2004) *Multidimensional poverty measurement and analysis*. Oxford: Oxford University Press.

Alkire, S., Conconi, A. & Roche, J.M. (2012) Multidimensional poverty index 2012: brief methodological note and results. University of Oxford, Department of International Development, Oxford Poverty and Human Development Initiative, Oxford, UK.

Barham, V., Boadway, R., Marchand, M. & Pestieau, P. (1995) Education and the poverty trap. *European Economic Review*, 39(7), 1257–275.

Belitz, C., Brezger, A., Klein, N., Kneib, T., Lang, S. & Umlauf, N. (2015) BayesX - Software for Bayesian inference in structured additive regression models. Version 3.0.2. Available from: http://www.bayesx.org.

Bhattacharyya, S. & Resosudarmo, B.P. (2015) Growth, growth accelerations, and the poor: lessons from Indonesia. *World Development*, 66, 154–165.

Calvo, C. & Dercon, S. (2013) Vulnerability to individual and aggregate poverty. *Social Choice and Welfare*, 41(4), 721–740.

de Leon, A.R. & Wu, B. (2011) Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine*, 30(2), 175–185.

Decancq, K. (2014) Copula-based measurement of dependence between dimensions of well-being. *Oxford Economic Papers*, 66(3), 681–701.

Donat, F. & Marra, G. (2017) Semi-parametric bivariate polychotomous ordinal regression. *Statistics and Computing*, 27(1), 283–299.

Donat, F. & Marra, G. (2018) Simultaneous equation penalized likelihood estimation of vehicle accident injury severity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(4), 979–1001.

Duclos, J.-Y., Sahn, D.E. & Younger, S.D. (2006) Robust multidimensional poverty comparisons. *The Economic Journal*, 116(514), 943–968.

Duflo, E. (2004) The medium run effects of educational expansion: evidence from a large school construction program in Indonesia. *Journal of Development Economics*, 74(1), 163–197.

Geyer, C. J. (2015) Trust: trust region optimization. R package version 0.1-7. URL: https://CRAN.R-project.org/package=trust

Günther, I. & Harttgen, K. (2009) Estimating households vulnerability to idiosyncratic and covariate shocks: a novel method applied in Madagascar. *World Development*, 37(7), 1222–1234.

Haberman, S.J. (1980) Discussion of "Regression models for ordinal data" by Peter McCullagh. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 42(2), 136–137.

Hohberg, M. & Lay, J. (2015) The impact of minimum wages on informal and formal labor market outcomes: evidence from Indonesia. *IZA Journal of Labor & Development*, 4(1), 14.

Kalliovirta, L. (2008) Quantile residuals for multivariate models. *Technical Report 247*, Helsinki Center of Economic Research.

Kleiber, C. & Kotz, S. (2003) *Statistical size distributions in economics and actuarial sciences*. Hoboken: Wiley

Klein, N. & Kneib, T. (2016) Simultaneous inference in structured additive conditional copula regression models: a unifying Bayesian approach. *Statistics and Computing*, 26(4), 841–860.

Klein, N., Kneib, T., Klasen, S. & Lang, S. (2015) Bayesian structured additive distributional regression for multivariate responses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(4), 569–591.

Klein, N., Kneib, T., Marra, G., Radice, R., Rokicki, S. & McGovern, M.E. (2019) Mixed binary-continuous copula regression models with application to adverse birth outcomes. *Statistics in Medicine*, 38(3), 413–436.

Kobus, M. & Kurek, R. (2018) Copula-based measurement of interdependence for discrete distributions. *Journal of Mathematical Economics*, 79, 27–39.

Marra, G. & Radice, R. (2017) Bivariate copula additive models for location, scale and shape. *Computational Statistics & Data Analysis*, 112, 99–113.

Marra, G. & Radice, R. (2019) GJRM: generalised joint regression modelling. R package version 0.2. URL: https://CRAN.R-project.org/package=GJRM

Marra, G. & Wood, S.N. (2012) Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1), 53–74.

McKelvey, R.D. & Zavoina, W. (1975) A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, 4(1), 103–120.

Nelsen, R. (2006) *An introduction to copulas*. New York: Springer.

Perez, A. & Prieto, M. (2015) Measuring dependence between dimensions of poverty in Spain: an approach based on copulas. In: *Proceedings of the conference of the international fuzzy systems association and the European society for fuzzy logic and technology (IFSA-EUSFLAT-15)*, pp. 734–741.

Quinn, C. (2007) Using copulas to measure association between ordinal measures of health and income. *Technical Report 07/24*, University of York.

Radice, R., Marra, G. & Wojtys, M. (2016) Copula regression spline models for binary outcomes. *Statistics and Computing*, 26(5), 981–995.

RAND. (2017) The Indonesia family life survey (IFLS). URL: https://www.rand.org/labor/FLS/IFLS.html

Rigby, R.A. & Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554.

Rue, H. & Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman & Hall.

Ruppert, D., Wand, M.P. & Carroll, R.J. (2003) *Semiparametric regression*. Cambridge series in statistical and probabilistic mathematics. Cambridge: Cambridge University Press.

Sklar, A. (1959) Fonctions de reepartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris*, 8, 229–231.

Strauss, J., Witoelar, F. & Sikoki, B. (2016) *The fifth wave of the Indonesia family life survey: overview and field report*, vol. 1. Santa Monica, CA: RAND

Tan, B.K., Panagiotelis, A. & Athanasopoulos, G. (2018) Bayesian inference for the one-factor copula model. *Journal of Computational and Graphical Statistics* (to appear).

Thi Nguyen, K.A., Jolly, C.M., Bui, C.T.P.N. & Le, T.H.T. (2015) Climate change: rural household food consumption and vulnerability: the case of Ben Tre province in Vietnam. *Agricultural Economics Review*, 16(2), 95–109.

Vatter, T. & Chavez-Demoulin, V. (2015) Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141, 147–167.

Wahba, G. (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(3), 364–372.

Wood, S.N. (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467), 673–686.

Wood, S.N. (2006) On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, 48(4), 445–464.

Wood, S. (2017) *Generalized additive models: an introduction with R*, 2nd edn. Chapman & Hall/CRC texts in statistical science. Boca Raton, FL: CRC Press.

World Bank. (2018) Piecing together the poverty puzzle. *Technical report*. World Bank, Washington, DC.

Zereyesus, Y.A., Embaye, W.T., Tsiboe, F. & Amanor-Boadu, V. (2017) Implications of nonfarm work to vulnerability to food poverty-recent evidence from northern Ghana. *World Development* 91, 113–124.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hohberg M, Donat F, Marra G, Kneib T. Beyond unidimensional poverty analysis using distributional copula models for mixed ordered-continuous outcomes. *J R Stat Soc Series C*. 2021;70:1365–1390. https://doi.org/10.1111/rssc.12517