A D-vine copula-based quantile regression model with spatial dependence for COVID-19 infection rate in Italy

Pierpaolo D'Urso [a], Livia De Giovanni [b], Vincenzina Vitale [a],*

[a] Department of Social and Economic Sciences, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy
[b] Department of Political Sciences, LUISS University, Viale Romania, 32, 00197 Rome, Italy

ARTICLE INFO

ABSTRACT

The main determinants of COVID-19 spread in Italy are investigated, in this work, by means of a D-vine copula based quantile regression. The outcome is the COVID-19 cumulative infection rate registered on October 30th 2020, with reference to the 107 Italian provinces, and it is regressed on some covariates of interest accounting for medical, environmental and demographic factors. To deal with the issue of spatial autocorrelation, the D-vine copula based quantile regression also embeds a spatial autoregressive component that controls for the extent of spatial dependence. The use of vine copula enhances model flexibility accounting for non-linear relationships and tail dependencies. Moreover, the model selection procedure leads to parsimonious models providing a rank of covariates based on their explanatory power with respect to the outcome.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

In the late December 2019, the atypical pneumonia originated by SARSCoV-2 virus has exponentially spread out from the city of Wuhan, the capital of the Chinese province of Hubei, identified as the epicentre of the contagion, to all other countries in the world becoming a global health emergency (Li et al., 2020).

---

* Corresponding author.
  E-mail addresses: pierpaolo.durso@uniroma1.it (P. D'Urso), ldegiovanni@luiss.it (L. De Giovanni), vincenzina.vitale@uniroma1.it (V. Vitale).

Italy was the first European country seriously hit by pandemic (Murgante et al., 2020); in particular, the two regions of Lombardia and Veneto, in February 2020, early experimented the strong negative effects of the first COVID-19 epidemic wave on their regional health systems.

On March 9th 2020, the Italian government established a total lockdown extended to the whole territory based on social distancing, use of mask, contact tracing and isolation of positives (Giordano et al., 2020). All these measures reduced the initial exponential growth of infection and mortality rates as well as the hospital admissions; the attention was, above all, on intensive care units in order to prevent beds saturation, being Italy one of the European countries with the lowest availability of acute care beds per person (Farcomeni et al., 2020; Remuzzi and Remuzzi, 2020).

The Italian Civil Protection Department, from the late February 2020, provides, daily, real time data on the Italian outbreak, the most part at the regional level. With reference to the Italian provinces, the only available information is the cumulative number of infected people with no information on the total number of individuals swabbed.

However, the drawback refers to data quality, in general, due to misreporting of the number of positive people and deaths, especially in the first phase. Moreover, since swabs are mainly performed on symptomatic people, the pandemic spread in the Italian population is certainly affected by a negative bias.

For all above considerations, the main determinants of COVID-19 spread in Italy are far from being clearly identified and deeply analysed.

Our proposal has been that of modelling the dependence relationships among the COVID-19 cumulative infection rate and some covariates of interest through a subclass of R-vine copula (Bedford and Cooke, 2001, 2002; Kurowicka and Cooke, 2006), the so-called D-vine copula (Aas et al., 2009). This allowed to deal with non-linear dependencies ensuring a greater model flexibility.

In this work, the combined use of vine copula and quantile regression has been considered: more specifically, a D-vine copula based quantile regression, as proposed by Kraus and Czado (2017), has been applied to model the COVID-19 cumulative infection rate, registered on October 30th 2020 for the 107 Italian provinces, by means of some covariates of interest collected from different data sources and referred to the last available year.

When studying the COVID-19 infection rate, it is important to consider its spatial spread pattern, particularly evident in Italy and, once time again, related to the well-known geographical distinction between North-centre and South.

To take into account the spatial component, the dependent variable, spatially lagged by means of a suitable spatial weight matrix, has been efficiently embedded in the D-vine quantile regression as further covariate, similarly to the Spatial Autoregressive Model (SAR, Anselin (1988)) widely used in spatial Econometrics for the same scope.

In particular, two different Spatial D-vine copula based quantile regression models have been proposed according to two different specifications of the spatial weight matrix.

This theoretical framework has been built with the aim of identifying the main factors that directly, or not, could influence the infection spread, its incidence and severity.

In particular, we argue that the main advantage of using the copula based quantile regression model arises from the consideration that the copula allows great flexibility in dependence modelling overcoming some limits of the classical regression, firstly linearity that, in this context, could result in a very strong and misleading assumption. By means of the spatial autoregressive component, the model accounts for spatial dependence, also providing a general measure of its extent.

Moreover, the vine copula based quantile regression constitutes an interesting methodological novelty in literature (Kraus and Czado, 2017); based on our knowledge, it has been recently applied in few works (Niemierko et al., 2019; Nguyen-Huy et al., 2018; Liu et al., 2020; Martey and Attoh-Okine, 2019) and not yet applied to COVID-19 data.

More generally, for interesting works applying spatial and spatio-temporal model to COVID-19 data, we can refer to Aràndiga et al. (2020), Bertuzzo et al. (2020), Giuliani et al. (2020), Mollalo et al. (2020), Kang et al. (2020), Bartolucci and Farcomeni (2021), Lee et al. (2021), Sahu and Böhning (2021), Vitale et al. (2021). For the application of a robust non linear regression model to the same data, see Girardi et al. (2020).

The paper is organized as follows. The general theory of vine copula is introduced in Section 2 while in Section 3 the theoretical framework of the D-vine copula based quantile regression is

described in detail, focusing on the model selection procedure. The definition of the spatial weight matrices used in the model are described in Section 4 while results are presented in Section 5. Section 6 includes conclusions and further perspectives of research.

2. The vine copula

2.1. Copula model

A *n-dimensional copula* C is a *n-variate* cumulative distribution function (cdf) on the unit hypercube $[0, 1]^n$ with uniformly distributed marginals U on the interval $[0, 1]$. The link between a multivariate distribution and its copula is provided by the Sklar theorem (Sklar, 1959): let F be a *n-dimensional* distribution function of the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ with univariate marginals $F_1 \ldots F_n$, any *n-variate* cdf F can be written as:

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{1}$$

for some appropriate *n-dimensional* copula C that, when $\mathbf{X}$ is absolutely continuous, is unique.

The theorem states that any joint distribution can be retrieved from the marginal distributions $F_1 \ldots F_n$ through the copula. In other words, a copula function enables to separate the marginal distributions from the dependency structure of a given multivariate distribution.

Moreover, the copula from (1) can be expressed as:

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) \tag{2}$$

where $F_i^{-1}(u_i)$ is the inverse function of the *i*th marginal.

In terms of density functions f, if F is absolutely continuous with strictly increasing continuous marginals $F_1 \ldots F_n$, through the chain rule decomposition, we have:

$$f(\mathbf{x}) = c(F_1(x_1) \ldots F_n(x_n)) \prod_{i=1}^{n} f_i(x_i) \tag{3}$$

for some appropriate unique *n-variate* copula density c.

As already said, while there is an exhaustive literature on bivariate copula families, for higher dimensions the choice is very limited.[1] The extension to more than three dimensions is not straightforward; indeed, some multivariate copulae, such as the Gaussian, the Student-t or the Archimedean ones, lack flexibility to accurately model the dependence structure in high dimensions.

The *pair-copula construction* (PCC) has been introduced by Joe (1996) to overcome this limit. It decomposes the multivariate dependence structure, *i.e.* the multivariate copula distribution, into the product of bivariate distributions named *pair-copulae*, each one modelled, independently, by a suitable copula. Therefore, differently than multivariate copulae, it does not assume that all the bivariate dependencies are of the same type.

Since in higher dimensions the number of possible pair-copulae constructions grows up significantly, a graph-theoretic representation called *R-vine*, *i.e. Regular vine*, was proposed by Bedford and Cooke (2001, 2002) and then analysed in detail by Kurowicka and Cooke (2006); inspired by these works, Aas et al. (2009) and Czado (2010) derived the inference techniques for two sub-classes of regular vines, known as *C-* and *D-vines*, widely used in research and the applications.

Therefore, in the following, the R-vine graphical models will be presented focusing, after, on its subclass of *D-vines*, of major interest in this work being used in the copula quantile regression model proposed by Kraus and Czado (2017).

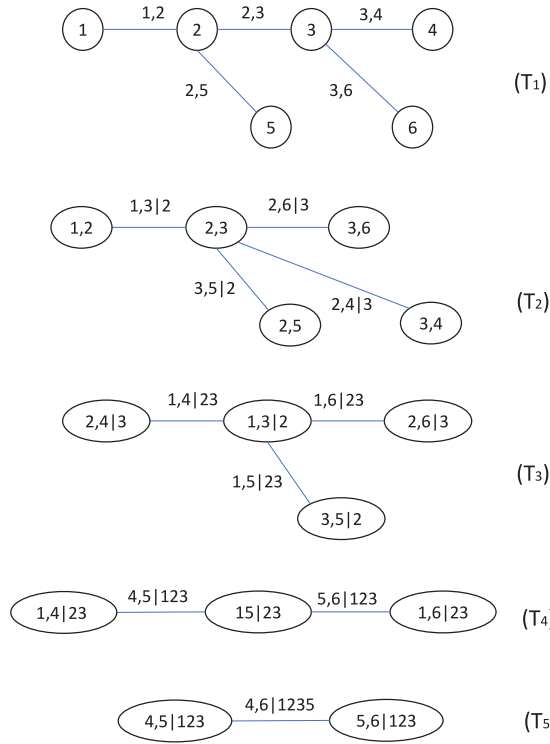[1] For theoretical references on copulae we can refer to Joe (1997).

Fig. 1. A six-dimensional R-vine.

2.2. R-vines and D-vines

More generally, the *R-vine* is a sequence of trees each edge of which corresponds to a pair-copula. According to Kurowicka and Cooke (2006), a R-vine $R = (T_1 \dots T_{n-1})$ on n variables is formally defined as follows:

(a) T_1 is a tree with nodes $N_1 = \{1, \dots, n\}$ and edges E_1.
(b) For $j = 2, \dots, n-1$, the tree T_j has nodes $N_j = E_{j-1}$ and edges set E_j.
(c) If two edges in T_j have to be joined (by an edge) as nodes in T_{j+1}, they must share a common node in T_j (*proximity condition*).

More simply, a *n-dimensional* R-vine is a set of $n-1$ trees such that the first tree comprises n nodes, identifying $n-1$ pairs of variables and also $n-1$ corresponding edges.

Therefore each subsequent tree is derived such that all the edges of tree j turn into nodes of the tree $j+1$ and can be joined by an edge in T_{j+1} only if the proximity condition holds. An example of six-dimensional R-vine is displayed in Fig. 1. Since for a R-vine many possible tree sequences can be specified, two sub-classes of R-vines, the Canonical vines (C-vines) and the Drawables vines (D-vines), are more often applied being their tree sequences easier to identify.

For the scope of this work, we will focus only on the latter. Properly, a D-vine has a path structure since no node in any tree T_j is connected to more than two edges. It is completely defined by the ordering of the sequence of the first tree only; all other subsequent trees are univocally identified according to the given sequence of the first one.

An example of D-vine on six dimensions is shown Fig. 2. Its nested tree structure consists of five trees T_j for $j = 1, \dots, 5$. Each tree T_j is composed of $7 - j$ nodes and $6 - j$ edges and, differently than a R-vine specification, each node in any tree T_j is connected to almost two edges.
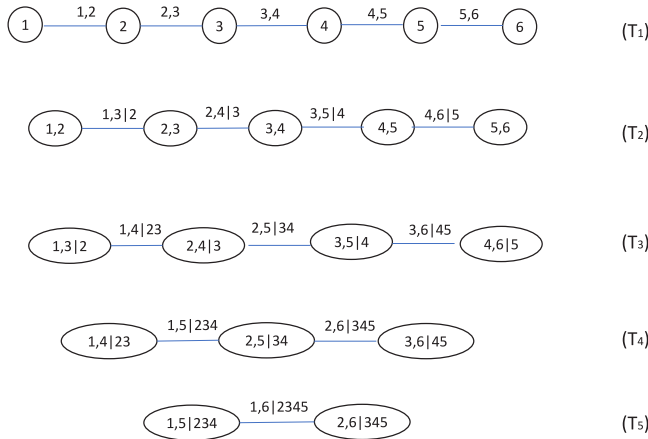
Fig. 2. A six-dimensional D-vine.

According to the proximity condition, the nodes in T_{j+1}, joined by an edge, are only those corresponding to the edges in T_j sharing a common node. For example, the edges $\{1, 2\}$ and $\{3, 4\}$ of the first tree cannot be joined by an edge when they become nodes in the second tree since they do not share a common node in T_1.

The tree structure allows the specification of all pair copulae of the *pair copula construction* since a bivariate copula density is associated with each edge and, hence, its label is the subscript of this copula: i.e. the edge $\{1, 4|2, 3\}$ in Fig. 2 defines the copula density $c_{14|23}(\cdot)$.

In particular, the copulae defined in the first tree are *unconditional* copulae while the others are all *conditional*; the copulae of the second tree have only one node as conditioning set, increasing by one for each tree.

Following Aas et al. (2009), with reference to a D-vine, the density $f(x_1, \ldots, x_n)$ can be written as:

$$f(x_1, \ldots, x_n) = \prod_{k=1}^{n} f_k(x_k) \times \tag{4}$$

$$\prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\ldots,i+j-1}\{(F(x_i|x_{i+1}, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})\}$$

where f_k is the density of the corresponding F_k, for $k = 1 \ldots n$ and $c_{i,i+j|i+1,\ldots,(i+j-1)}$ is the bivariate (conditional) copula density.

The density in (4) corresponds to a simplified PCC (Hobæk Haff et al., 2010) for which the assumption is that all pair-copulae depend on the conditioning vector $x_{i+1}, \ldots, x_{i+j-1}$ only through their arguments, not directly.

According to (4), any *n-dimensional* absolutely continuous pdf can be written as the product of the n marginal densities and the $n(n-1)/2$ pair-copulae corresponding to the edges of the $n-1$ trees; each bivariate copula has, as its arguments, the marginal CDFs at the first level, the conditional CDFs for all T_j, $j = 2, \ldots, n-1$.

Let denote the generic CDF with $F(x|\mathbf{v})$, Joe (1996) derived the recursive formula defined in (5) to compute it using only the pair-copulae specified from lower trees, the so called *h-function*:

$$h(x|\mathbf{v}) := F(x|\mathbf{v}) = \frac{\partial C_{xv_j|\mathbf{v}_{-j}}(F_{x|\mathbf{v}_{-j}}, F_{v_j|\mathbf{v}_{-j}})}{\partial F_{v_j|\mathbf{v}_{-j}}}, \forall j \tag{5}$$

where $\mathbf{V}$ is the random vector excluding X while $\mathbf{V}_{-j}$ is the same vector excluding the jth variable.

Suppose to consider the simplest case, a three-dimensional D-vine with order 1, 2, 3, the pdf is equal to:

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3) \cdot \\ \cdot c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{23}\{F_2(x_2), F_3(x_3)\} \cdot \\ \cdot c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \tag{6}$$

where, according to (5):

$$F(x_1|x_2) = \frac{\partial C_{12}(F(x_1), F(x_2))}{\partial F(x_2)} \tag{7}$$

$$F(x_3|x_2) = \frac{\partial C_{32}(F(x_3), F(x_2))}{\partial F(x_2)} \tag{8}$$

It is worth remembering that the importance of all the above results arises from the fact that all bivariate copulae can belong to different families and their parameters can be specified independently from each other.

The theoretical definition of the D-vine allows to introduce, in the next section, the D-vine copula based quantile regression model, recently proposed in literature by Kraus and Czado (2017).

3. The D-vine copula based quantile regression

In the field of quantile regression, the most famous and applied method is the linear quantile regression proposed by Koenker and Bassett (1978); properly, an extension of the well-known linear regression model based on the ordinary least squares extended, later, to account for non-parametric effects by Koenker (2011) and Fenske et al. (2011) (for a literature review on quantile regression models, see Kraus and Czado (2017) and reference therein.). In the work of Bernard and Czado (2015), it has been pointed out that linear quantile regression suffers from some limitations such as model mis-specification, crossing of quantiles, multicollinearity and from all the other typical drawbacks of linear models.

In this context, the use of vine copula enhances flexibility overcoming all previous issues affecting linear models; as shown later, this novel class of models inherits all the advantages related to the fact that the dependence relationship between the response variable and its covariates could be modelled through a vine copula.

The use of vine copula in the framework of quantile regression is essentially a novel proposal due to Kraus and Czado (2017) and Schallhorn et al. (2017), where the vine structure is restricted to the class of D-vine. Some extensions to the use of the general class of R-vine copula are in Cooke et al. (2015), where the response and the explanatory variables are both continuous, and in Chang and Joe (2019), where the variables are mixed. The latter approach finds the locally optimal regular vine structure among all predictors and then adds the response to each selected tree in the vine structure as a leaf.

We chose to apply the D-vine based approach (Kraus and Czado, 2017) since the R-vine copula regression proposed by Chang and Joe (2019) suffers from the drawback of a high computational cost for high-dimensional data; moreover, as argued by Tepegjozova et al. (2021), the procedure may not be that maximizes the conditional response likelihood while the approach proposed by Kraus and Czado (2017) is specifically based on optimizing the conditional log-likelihood and selecting covariates using a procedure similar to the forward selection in multiple regression that could be easier realized with a D-vine.

Indeed, as described in detail later, the methodology proposed by Kraus and Czado (2017) is not only able to account for non linear dependencies but also to select parsimonious models since the algorithm used to identify the D-vine copula model implicitly implements a variable selection procedure too. In a more formal way, according to Kraus and Czado (2017), the main target of a quantile regression, based on a D-vine copula specification, is the quantiles' estimate q_α, where $\alpha \in (0, 1)$ is some quantile level of the response variable Y given its covariate vector $\mathbf{X} = (X_1, \ldots, X_d)'$.

In particular, the conditional quantile function is defined as:

$$q_\alpha(x_1, \ldots, x_d) := F^{-1}_{Y|X_1,\ldots,X_d}(\alpha|x_1, \ldots, x_d) \tag{9}$$

Let V and U_j denote the uniformly distributed probability integral transforms (PIT) of Y and the jth component of $\mathbf{X}$, i.e. $F_Y(Y)$ and $F_j(X_j)$, respectively. By Sklar theorem, it follows that:

$$F_{Y|X_1,\ldots,X_d}(y|x_1, \ldots, x_d) = C_{V|U_1,\ldots,U_d}(v|u_1, \ldots, u_d) \tag{10}$$

$$F^{-1}_{Y|X_1,\ldots,X_d}(\alpha|x_1, \ldots, x_d) = F^{-1}_Y\left(C^{-1}_{V|U_1,\ldots,U_d}(\alpha|u_1, \ldots, u_d)\right) \tag{11}$$

Therefore, to get $\widehat{q}_\alpha(x_1, \ldots, x_d)$, suitable estimators of the univariate marginals F_Y and $F_j \forall j = 1, \ldots, d$ as well as of the copula $C_{V,U_1,\ldots,U_d}$ have to be computed; accordingly, a kernel estimator is used for the marginals while a simplified D-vine copula is used for C.

To this purpose, the D-vine structure is estimated such that the ordered sequence of nodes of the first tree has V as its first node followed by one among the possible permutations of the covariates $U_1, \ldots, U_d$.

By using the recursive formula (5), the copula $C^{-1}_{V|U_1,\ldots,U_d}(\alpha|u_1, \ldots, u_d)$ could be computed in terms of nested inverse h-functions.

As argued by Kraus and Czado (2017), being the inverse copula monotonically increasing in α, it does not allow the crossing of quantiles functions. Moreover, the use of multivariate copula enhances flexibility embedding non-linear and asymmetric dependencies, heavy tails and tail dependencies.

The fitting procedure as well as the open question about which permutation of the covariates has to be chosen are both analysed, in detail, in the next paragraph.

3.1. The variables selection procedure

The algorithm used to fit a D-vine based quantile regression model is implemented in the R package `vinereg` (Nagler, 2018).

As already specified, the first step consists in fitting the marginals non-parametrically by means of kernel estimators in order transform the n observed values to the corresponding so called pseudo copula data v_i and $\mathbf{u}_i$ for $i = 1, \ldots, n$, respectively.

In the second step, the pseudo copula data are used in the estimation of the D-vine copula. In order to increase the model explanatory power, the $d + 1$ dimensional D-vine is chosen such that, given V as first node of the first tree, the sequence of the other d nodes of the same tree is chosen based on a sequential procedure.

Properly, one covariate at a time is added that improves the conditional log-likelihood[2] (*cll*) of the estimated D-vine copula.

Suppose that, at the lth step, the unselected covariates are U_k and U_m. Two D-vine copula models are fitted according to the sequence $V - U_1 - \cdots U_{l-1} - U_k$ and $V - U_1 - \cdots U_{l-1} - U_m$, respectively. Also denote the associated *clls* with cll_k and cll_m and that associated to the D-Vine of previous step with cll_{l-1}.

If $cll_k > cll_m > cll_{l-1}$, the lth selected covariate will be U_k, i.e. that associated with the highest increase of the *cll* (U_m if $cll_m > cll_k > cll_{l-1}$). If none of them increases cll_{l-1}, the algorithm stops selecting $l - 1$ covariates.

Therefore, an implicit variable selection procedure is defined allowing to choose the most influential covariates, ranking them according to their explanatory power with respect to the response variable, leading to parsimonious models.

---

[2] The algorithm allows to choose the best model also according to the penalized conditional loglikelihoods AIC and BIC. See Kraus and Czado (2017) for the mathematical derivation of the *cll* in the model.

4. Spatial dependence

A variable is said to be spatially autocorrelated if it shows a systematic spatial pattern, that is, the variable values at a location are highly influenced by those at the adjacent locations and, more generally, the spatial autocorrelation declines with increasing distance.

Spatial autocorrelation could be positive or negative: the former occurs when a high value of a variable at a location is associated with high values at neighbouring locations, the latter when a high value at a location is associated with low values at the neighbouring ones.

In this application, in order to take into account the strong spatial dependence among units, a spatial lagged dependent variable has been added to the model, as in the Spatial Autoregressive Models (SAR) widely used in spatial Econometrics (Anselin, 1988). Hence, the dependent variable, say **Y** of dimension n, is usually lagged by means of a spatial weight matrix, say $\mathbf{W}_{n \times n}$, that is a square matrix whose element w_{ij} is 0 if zones i and j are not adjacent while is equal to some positive value if they are neighbours:

$$\mathbf{W} = \begin{cases} w_{ij} > 0 & \textit{if zones i and j are adjacent} \quad \textit{with } i \neq j \\ 0 & \textit{otherwise} \end{cases}$$

Here, we adopted the adjacency definition between spatial polygons known as "Queen's Case adjacency" according to which two units are close if they share a side or an edge. In Fig. 3, the links between the polygons of the Italian provinces are reported: a red edge has been drawn between the centroids of the two zones that are adjacent according the Queen criterion.

In this study, two different definitions of weights matrices have been taken into account, named $\mathbf{W_1}$ and $\mathbf{W_2}$:

$$\mathbf{W_1} = \begin{cases} w_{ij} = \dfrac{1}{s_i} & \textit{if zones i and j are adjacent} \quad \textit{with } i \neq j \\ 0 & \textit{otherwise} \end{cases}$$

where s_i is the number of neighbours of polygon i,

$$\mathbf{W_2} = \begin{cases} w_{ij} = 1 & \textit{if zones i and j are adjacent} \quad \textit{with } i \neq j \\ 0 & \textit{otherwise.} \end{cases}$$
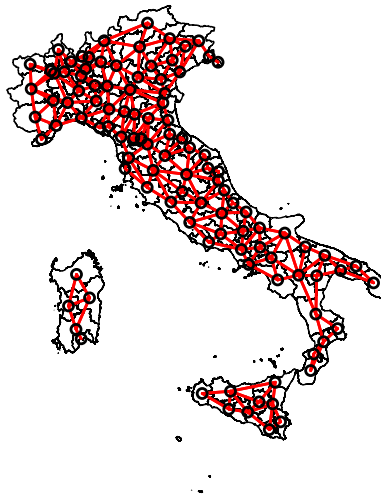

Fig. 3. The links between polygons according to the Queen criterion.

In the former, each element of *i*th row is divided by the number of its neighbours so that the effect of any individual neighbour decreases as the number of neighbours increases. This produces what called the row normalization of **W**. In the latter definition, the simplest one, the matrix **W** corresponds to the indicator matrix.

We argue that the role of the spatial autoregressive component in the model is twofold: at the same time, it allows to evaluate the effects of spatial autocorrelation as well as the impact of all other covariates controlled for the spatial dependence.

5. Application

5.1. Data description

In this study, the analysis focuses on the application of the D-vine copula-based quantile regression, with spatial autoregressive component, to model the COVID-19 cumulative infection rate per 10000 inhabitants registered on October 30th 2020, with reference to the 107 Italian provinces as shown in the map of Fig. 4.

In Table 1, all involved covariates, with their labels, the reference year and their sources, are reported.[3] In particular, the disposable income and the employment rate have the role of control variables, while the total-age dependency ratio, the old-age dependency ratio and the healthy expectancy at birth account for the population structure knowing that COVID-19 infection is more insidious for the elderly population.

The age-adjusted mortality rates per gender, as well as the drugs consumption for diabetes, hypertension, asthma and BPCO, have been included in the model as proxies of the spread and severity of the main diseases in the territories.

The variables related to the climate and air pollution have been taken into account in order to investigate the possible relationship between infection and environment conditions.

---

[3] For the new province named "South Sardinia", born in 2017, when the datum was not available, it has been imputed with the mean of its neighbours.
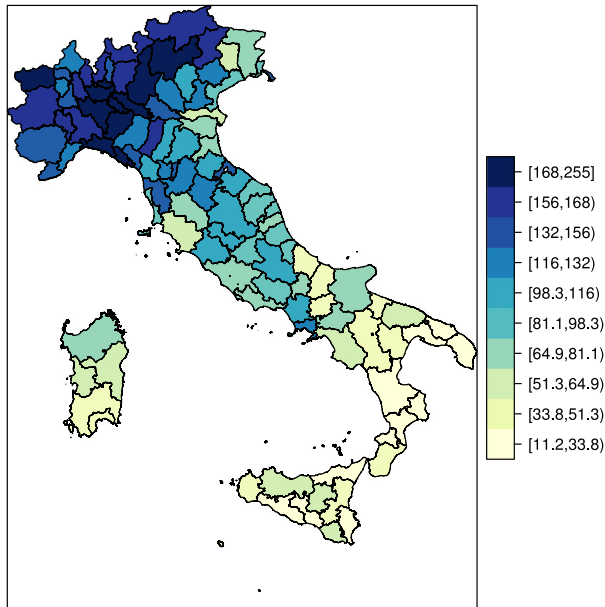

Fig. 4. The COVID-19 cumulative infection rate per 10000 inhabitants registered on October 30th, 2020.

Table 1

The variables used in the D-vine copula-based quantile regression model.

Label	Variable	Reference year	Source
COVID_Infection_rate	Total number of infections per 10000 inhabitants	2020	Civil Protection Department
Life_expectancy_at_birth	Life expectancy at birth	2018	ISTAT
Income	Per capita disposable income	2016	ISTAT
Employment_rate	Employment rate (20-64 years old)	2018	ISTAT
Total_age_dependency	Total-age dependency ratio	2020	ISTAT
Old_age_Dependency	Old-age dependency ratio	2020	ISTAT
MortalityInfections_M	Age-adjusted mortality rate from infectious diseases for Male per 10000 inhabitants	2017	Health For All Italia 2020
MortalityInfections_F	Age-adjusted mortality rate from infectious diseases for Female per 10000 inhabitants	2017	Health For All Italia 2020
MortalityCancer_M	Age-adjusted mortality rate from cancer for Male per 10000 inhabitants	2017	Health For All Italia 2020
MortalityCancer_F	Age-adjusted mortality rate from cancer for Female per 10000 inhabitants	2017	Health For All Italia 2020
MortalityPneumonia_Flu_M	Age-adjusted mortality rate from pneumonia and influenza for Male per 10000 inhabitants	2017	Health For All Italia 2020
MortalityPneumonia_Flu_F	Age-adjusted mortality rate from pneumonia and influenza for Female per 10000 inhabitants	2017	Health For All Italia 2020
NO2	Nitrogen dioxide annual mean (average of mean values of stations belonging to the Province)	2018	ISPRA
PM10	PM10 annual mean (average of mean values of stations belonging to the same Province)	2018	ISPRA
Climate	Climate index	2019	Sole 24 Ore's life-quality index
General_practitioners	General practitioners per 10000 inhabitants	2019	Sole 24 Ore's life-quality index
Diabetes	Diabetes drug consumption (per capita minimum units)	2019	Sole 24 Ore's life-quality index
Hypertension	Hypertension drug consumption (per capita minimum units)	2019	Sole 24 Ore's life-quality index
Asthma	Asthma and Chronic obstructive pulmonary diseases drug consumption (per capita minimum units)	2019	Sole 24 Ore's life-quality index
Lagged_COVID_Infection1	Total number of infections per 10000 inhabitants weighted by the contiguity matrix $\mathbf{W_1}$)	2020	Authors' elaboration
Lagged_COVID_Infection2	Total number of infections per 10000 inhabitants weighted by the contiguity matrix $\mathbf{W_2}$)	2020	Authors' elaboration

The rate of general practitioners has been included being a topic of a strong debate in Italy; the COVID-19 outbreak has highlighted several disparities in the regional health systems showing, in some cases, a lack of an adequate primary care network, although it has been known the key role of general practitioners in meeting the needs of patients in their communities (Mugnai and Bilato, 2020; Cicchetti et al., 2021; Mauro and Giancotti, 2021).

5.2. The estimated models

According to the model selection procedure described in Section 3.1, in the first step, each marginal has been fitted using a local polynomial kernel density estimator, implemented in the R package kde1d, also suitable to handle data with a bounded support[4]. In the second step, the D-vine copula has been fitted to the transformed data, involving an implicit variable selection procedure that ranks the covariates according to their explanatory power with respect to the response variable, from the most influential to the less one with respect to the outcome. The best model has been chosen by minimizing the penalized conditional loglikelihood (cAIC).

Two D-vine copula quantile regression models have been estimated depending on the spatially lagged dependent variable used in the model to control for the extent of spatial dependence. The former, henceforth called *DVQR_1* model, includes the variable named Lagged_COVID_Infection1 in Table 1 based on the contiguity matrix $\mathbf{W_1}$, *i.e.* the mean of the infection rate per 10000 inhabitants of the neighbouring provinces. The latter, henceforth called *DVQR_2* model, includes the variable named Lagged_COVID_Infection2 in Table 1 based on the contiguity matrix $\mathbf{W_2}$, *i.e.* the sum of infection rate per 10000 inhabitants of the neighbouring provinces.

Based on the estimation procedure described in Section 3.1, the *DVQR_1* model includes the selected covariates reported in Table 2 while the *DVQR_2* model includes those reported in Table 3. The associated cAIC value is 987.544 and 1000.29 for the *DVQR_1* and *DVQR_2* model, respectively.

Table 2
The outcome and the selected covariates for the *DVQR_1* model.

Vine Node label	Variables	p_value
1	COVID_Infection_rate	
7	Lagged_COVID_Infection1	0.000
6	Hypertension	0.012
5	MortalityPneumonia_Flu_M	0.029
2	Life_expectancy_at_birth	0.028
8	Old_age_Dependency	0.050
3	Income	0.035
4	MortalityInfections_F	0.041

Table 3
The outcome and the selected covariates for the *DVQR_2* model.

Vine Node label	Variables	p_value
1	COVID_Infection_rate	
11	Lagged_COVID_Infection2	0.000
8	MortalityPneumonia_Flu_M	0.000
10	General_pratictioners	0.002
6	MortalityCancer_M	0.009
7	MortalityCancer_F	0.011
9	Hypertension	0.042
5	MortalityInfections_F	0.030
3	Income	0.002
2	Life_expectancy_at_birth	0.022
4	MortalityInfections_M	0.044

The covariates in Tables 2 and 3 are ranked according to their own explanatory power with respect to the outcome[5]; the order is retrieved by the sequence of the first tree of the estimated D-vine, shown in Figs. 5 and 6 for the *DVQR_1* and *DVQR_2* model, respectively.

---

[4] The choice of the polynomial degree as well as the bandwidth multiplier have been individually set for each variable.

[5] The first column in the tables represents the identification number of the variable in the D-vine representation.
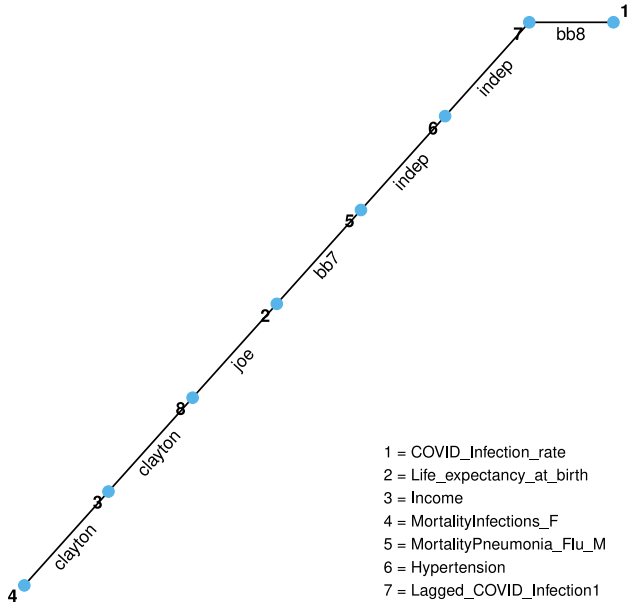
Tree 1

Fig. 5. The first tree of D-vine for the *DVQR_1* model (the sequence has to be read from the right to the left).
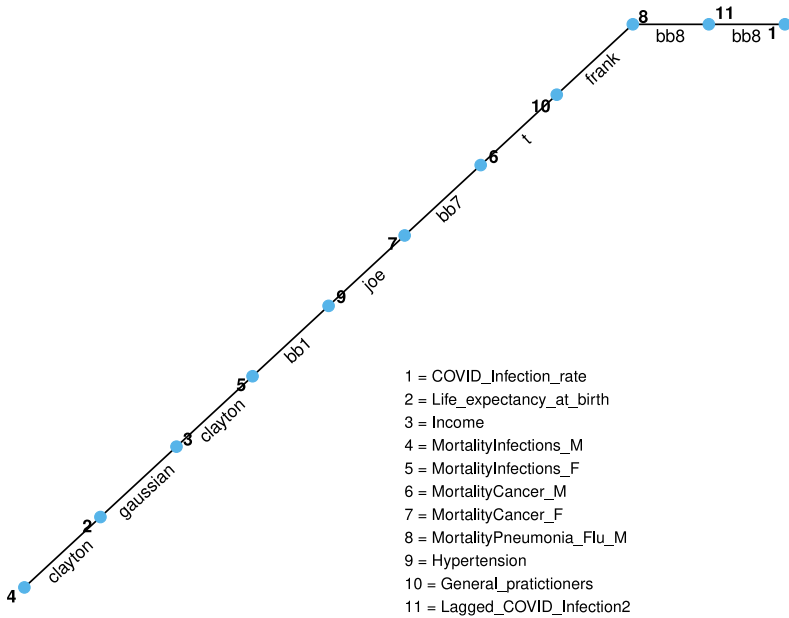
Tree 1

Fig. 6. The first tree of D-vine for the *DVQR_2* model (the sequence has to be read from the right to the left).

For the sake of completeness, all other trees are shown in Fig. A. 1 of Appendix A, for the *DVQR_1* model, and in Fig. B. 1 of Appendix B, for the *DVQR_2* model. Moreover, in the same appendices, Tables A. 1 and B. 1 show the selected copula and its parameters associated with all edges of the D-vine of the *DVQR_1* model and *DVQR_2* model, respectively: the last column, in particular, shows the corresponding theoretical Kendall Tau.

By looking at the first tree, for both models, the first influential covariate is represented by the lagged dependent infection rate, as expected, proving that the spread of COVID-19 outbreak is affected by spatial dependence, *i.e.* its values are influenced by those of its neighbours.

We argue that the strength of spatial dependence could be assessed, here, by means of the Kendall Tau coefficient associated to the estimated marginal bivariate copula density between the infection rate and itself spatially lagged (the copula associated to the edge 1 of Tree 1 of the D-vine). Like the Moran's I index, widely used in spatial statistics to the same scope, the Tau coefficient properly provides a suitable measure of the degree of spatial dependence.

In particular, the Tau coefficient associated to the marginal bivariate copula density between the infection rate and itself spatially lagged, depicted in Fig. 7, is 0.678 and 0.611 for the *DVQR_1* and *DVQR_2* model respectively, suggesting a positive spatial autocorrelation for the spread of pandemic in Italy. In particular, we argue that, in both plots, there is a higher positive correlation between the lower percentiles of both variables.

For the sake of completeness, in Appendices A and B, Figs. A. 2 and B. 2 show the conditional copula densities between the outcome and each selected covariate for the *DVQR_1* and *DVQR_2* model, respectively.

As far as the other covariates are concerned, some other important considerations can be addressed when looking at their marginal effects, as shown in the next section.
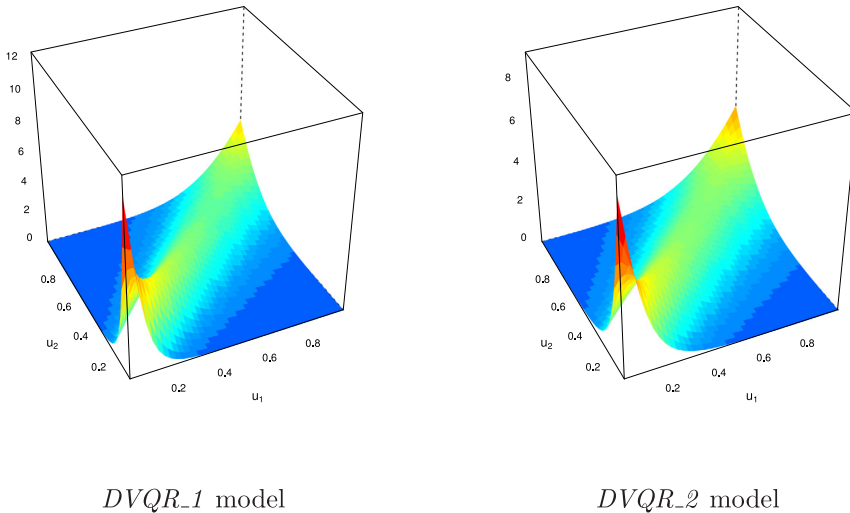

$DVQR\_1$ model $DVQR\_2$ model

Fig. 7. The estimated marginal bivariate copula density with uniform margins between the COVID-19 Infection rate and itself, spatially lagged, for both models.

5.3. Marginal effects

The plots of Fig. 8 are the marginal effects of the covariates belonging to *DVQR_1* model to which a smoothed line is added for the quantile levels 0.1, 0.5, 0.9. As already said, the first influential covariate is the spatially lagged dependent variable; the other main factors contributing to explain the COVID-19 infection rate are, in the order, the hypertension drug consumption, the age-adjusted mortality rate for Pneumonia and Flu in the males, the two covariates related to the population age
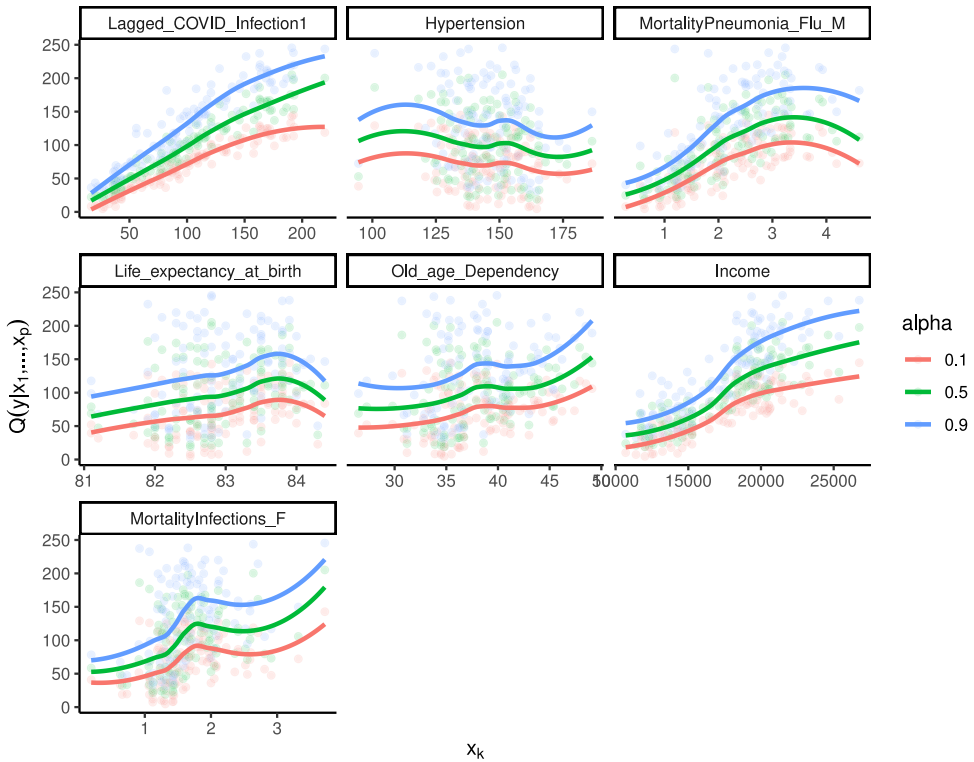
Fig. 8. The marginal effects of the selected covariates in *DVQR_1* model.

structure followed by the disposable income levels and the age-adjusted mortality rate for infections in the females.

At a first insight, the marginal effects clearly give evidence of non-linear relationships between the outcome and its covariates, as expected. Moreover, it is worth noting that the only negative relationship is that between the COVID-19 infection rate and hypertension drug consumption. The role of this covariate is twofold: it is a proxy of the incidence of the chronic disease in the area but also of the level of diffusion of treatments. Therefore, a territory with a higher consumption of this kind of drug could be also an area in which the prevention and the therapy are more widespread and incisive or, more simply, in which people take more care of their health. More generally, we believe that this unexpected negative relationship could be more likely due to the fact that the provinces of Lombardy, seriously hit by pandemic, show lower consumptions of hypertension drugs (AIFA, 2020).

Both positive relationships with the life expectancy at birth and with the old-age dependency ratio support the evidence that COVID-19 virus is particularly insidious in those territories characterized by an elderly population, for which it is known that the co-occurrence of chronic diseases is also a very frequent problem.

As far as the positive relationship with the income is concerned, it reflects the geography of the COVID-19 spread in Italy that has particularly hit the Northern regions for which the income levels and the employment rate are markedly higher than those of southern ones.

The relationship with the mortality rate for Pneumonia and Flu in the males, the third influential covariate, deserves particular attention being related to two important features of COVID-19 pandemic, that is the Pneumonia is also the most common serious complication of the new coronavirus and the mortality rate caused by COVID-19 in the male is higher than in the females.

Therefore, the model seems to highlight that COVID-19 virus has been spread in those territories already characterized by a high incidence and severity of Pneumonia, in particular in the male population. This is confirmed by the geographical spread of the 2017 age-adjusted mortality rate due to Pneumonia in the males, strongly concentrated in the North of Italy as shown in Fig. 9. It is worth noting that the only exception is represented by the province of Lodi, that was also the epicentre of contagion in Lombardy in the late February. As the same way, the relationship with the Mortality rate for infections in the females suggests that this new coronavirus has been spread mainly in those territories already marked by a higher prevalence of severe infections.

As far as the marginal effects of the *DVQR_2* model are concerned (see plots of Fig. 10) we notice that, in this model, the mortality rate for Pneumonia and Flu in the males becomes the second most influential covariate and, more interesting, it is also the role of the subsequent three covariates leading to the following observations.

There is a marked reduction of the infection rate as the number of general practitioners per 10000 inhabitants increases; this yields important insights into the debate recently developed in Italy around the important role of the general practitioners in providing the first treatments to COVID-19 patients contributing to prevent hospital bed saturation. An in-depth analysis led to the following considerations. From the map in Fig. 11, the provinces of Lombardy are those both seriously hit by pandemic and with lower rates of general practitioners per 10000 inhabitants. As pointed out by Cicchetti et al. (2021), general practitioners have a fundamental role in the system, acting as a filter, being able to decide on hospitalization or home care for positives thus reducing pressure on hospital services. The different approach adopted to face Covid-19 emergency revealed the corresponding differences in the regional health systems. Some regions were a positive example of integrated-home care such as Emilia Romagna that chose to early care for positives at home, thus reducing infection among health workers; Veneto, the first region hit by pandemic, adopted a territorial or "out-of-hospital" model of management by increasing the number of swabbed people, treating positives early thus reducing hospital admissions. On the contrary, Lombardy adopted the "in-hospital" COVID-19 management of positives (Mugnai and Bilato, 2020; Mauro and Giancotti,
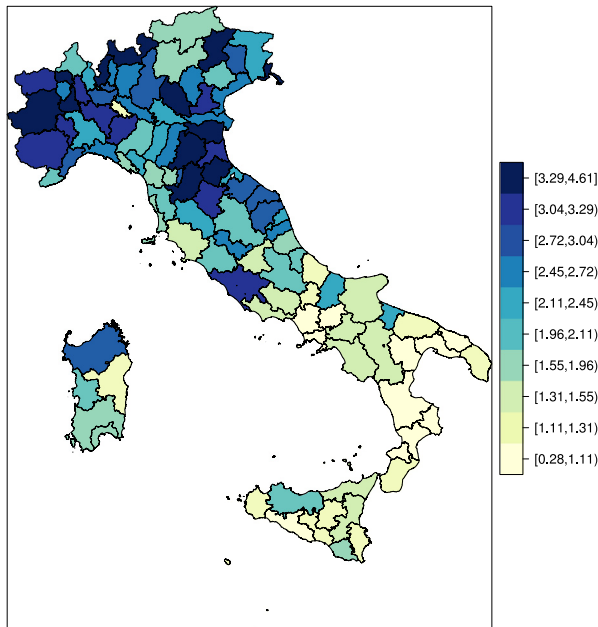

Fig. 9. The map of the Age-adjusted mortality rate from Pneumonia and Flu for Males per 10000 inhabitants, year 2017.
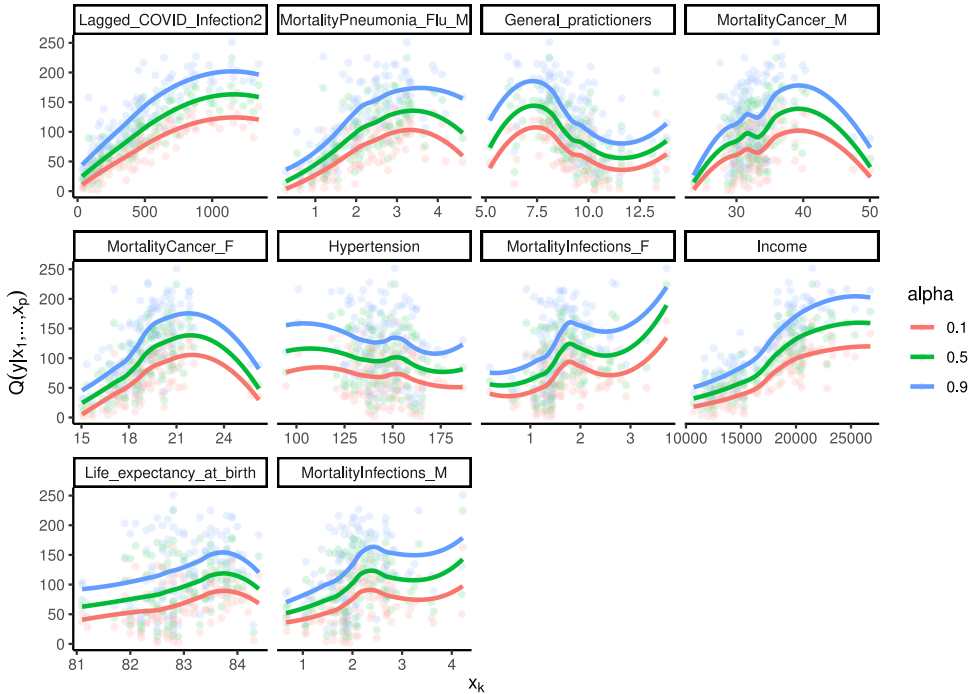
Fig. 10. The marginal effects of the selected covariates in *DVQR_2* model.

2021); indeed, during the first pandemic wave, the ratio between people in hospital versus those who received care at home was 1.14 in Lombardy and 0.61 in Emilia Romagna (Cicchetti et al., 2021).

Summing up, as argued by Cicchetti et al. (2021), the main drawback of some regional health care systems concerned the lack of continuity of care (essential to ensure the sustainability of any health care) due "to a lack of clear, homogeneous and effective approaches for primary care provision due to the still-unclear role of general practitioners" that led to a tendency for a hospital-centred approach. The absence of a primary care forced patients with minor symptoms to crowd the hospitals too. Therefore, a good primary assistance is the only winning strategy avoiding hospitals saturation and preventing the spread of infections.

The positive relationships between the infection rate and the age-adjusted mortality rate for cancer, for both males and females, suggest that the outbreak has been more spread in the same territories for which the incidence of cancer diseases is also significant and severe. It is worth looking, once again, at the geographical spread of the disease in Italy, shown in Fig. 12 for males and females, for which it emerges that the areas of Lombardy (see the province of Bergamo, for example) strongly hit by pandemic are also characterized by very high values of these rates. This does not prove any causal relationship but could suggest an indirect link with the pollution in the same zones, since it is long been known that pollution could increase the risk of contracting respiratory diseases. As far as the other covariates in the model are concerned, we point out that they are the same of the first model, except for the age-adjusted mortality infection rate for male, here selected as the last covariate in the model.

We can resume that the spread of COVID-19 infection seems to follow a spatial pattern with a higher incidence in the same areas, in general, more vulnerable to pulmonary diseases and other infections, for which the old-age of their resident population could represent an additional negative factor contributing to COVID-19 diffusion, above all if combined with a low level of primary medical assistance.
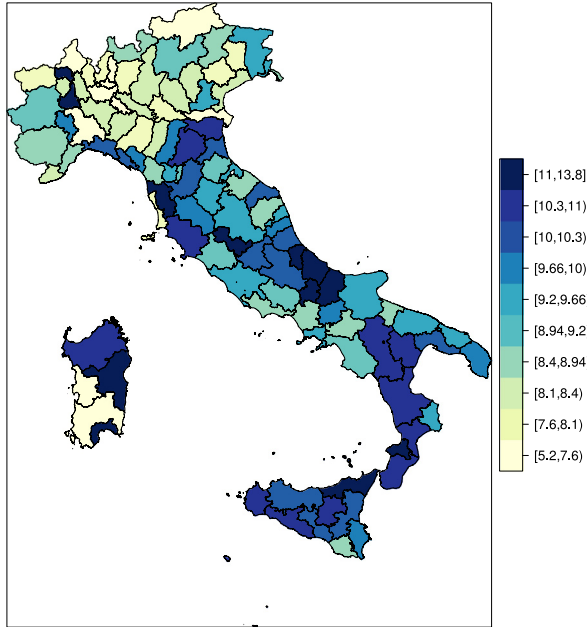
Fig. 11. The map of the number of General Practitioners per 10000 inhabitants, year 2019.

The COVID-19 dynamics in Italy fairly reflects the geographical stratification between North-centre and South of Italy and could be due to the fact that the northern and central regions are much more interested by business relationships promoting social interaction. In general, the question deserves much more attention in order to explore possible causal effects between the COVID-19 incidence rate and some environmental and territorial characteristics.

6. Conclusions

The COVID-19 outbreak in Italy has been massive with 647,674 infections and 38,321 deaths as reported by the Italian Civil Protection Department on October 30th, 2020. As of now, its spread has been characterized by strong territorial differences with the northern regions particularly hit by pandemic in the first wave.

The main results of this work concern the identification of the main factors influencing the spread of infections taking into account its spatial pattern. For this purpose, a D-vine copula based quantile regression has been applied to model the COVID-19 cumulative infection rate by means of some covariates of interest such as the incidence in the territories of the main severe diseases in terms of mortality rates or drug consumption, the population age structure, the levels of air pollution as well as the efficiency of the primary medical care system. The spatial dependence has been embedded in the model by means of the autoregressive component, *i.e.* the spatially lagged dependent variable. The use of pair-copulae and quantile regression allowed to account for non-linear dependencies, overcoming some drawbacks of classical regression models. Moreover, the copula density, modelling the bivariate relationship between the outcome and the spatial autoregressive component, also provided the extent of spatial dependence by means of the associated theoretical Kendall Tau coefficient.

Aware that our results could not be exhaustive of the complexity of the phenomenon under consideration neither to be conclusive about possible "causal" effects, we believe that our study could represent a first interesting insight into identification of the main factors related to the COVID-19 diffusion.
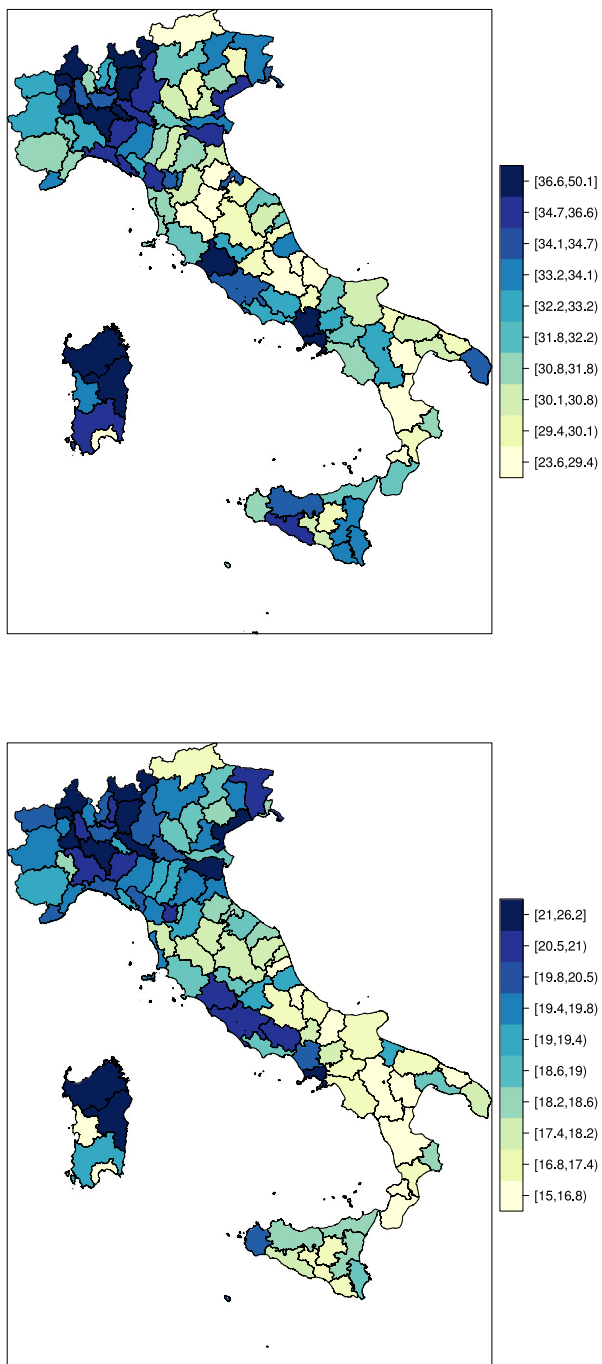
Fig. 12. The map of age-adjusted mortality rate for cancer in the males (top) and the females (bottom) per 10000 inhabitants, year 2017.

Both estimated models account for a positive spatial autocorrelation, a distinctive and predictable trait of pandemic. Moreover, in both proposed models, the influential covariates are strictly related to the main features of the new coronavirus since the COVID-19 infection rate has been more insidious in the same areas already more exposed to Pneumonia, especially in the male population, and vulnerable to infections, also increasing in those territories with a higher value of the old-age dependency rate, *i.e.* the ratio between people aged 65 and over and those aged 15-64. The unexpected negative relationship with the hypertension could be explained by the fact that the provinces of Lombardy, seriously hit by pandemic, show lower consumptions of hypertension drugs (AIFA, 2020), even if this relationship deserves further in-depth analysis.

For one of the two estimated models, also the mortality rate for cancer has been identified as influencing factor and, more interesting, the number of general practitioners per 10000 inhabitants, a proxy of the efficiency of the primary medical care in the territories. The strong negative association with the COVID-19 infection rate corroborates the belief of who, in the Italian public debate on this theme, considers the role of general practitioners essential to prevent the hospitalization of COVID patients and to allow an efficient health monitoring of the territory.

In the future, the further perspectives of research will be focused on studying the mortality risk associated with the COVID-19 virus, with an in-depth analysis on the role of co-morbidities in the hospitalized patients by means of spatial and spatio-temporal models.

Acknowledgments

Appendix A. D-vine for the *DVQR_1* model

Table A. 1
The D vine specification for the *DVQR_1* model.

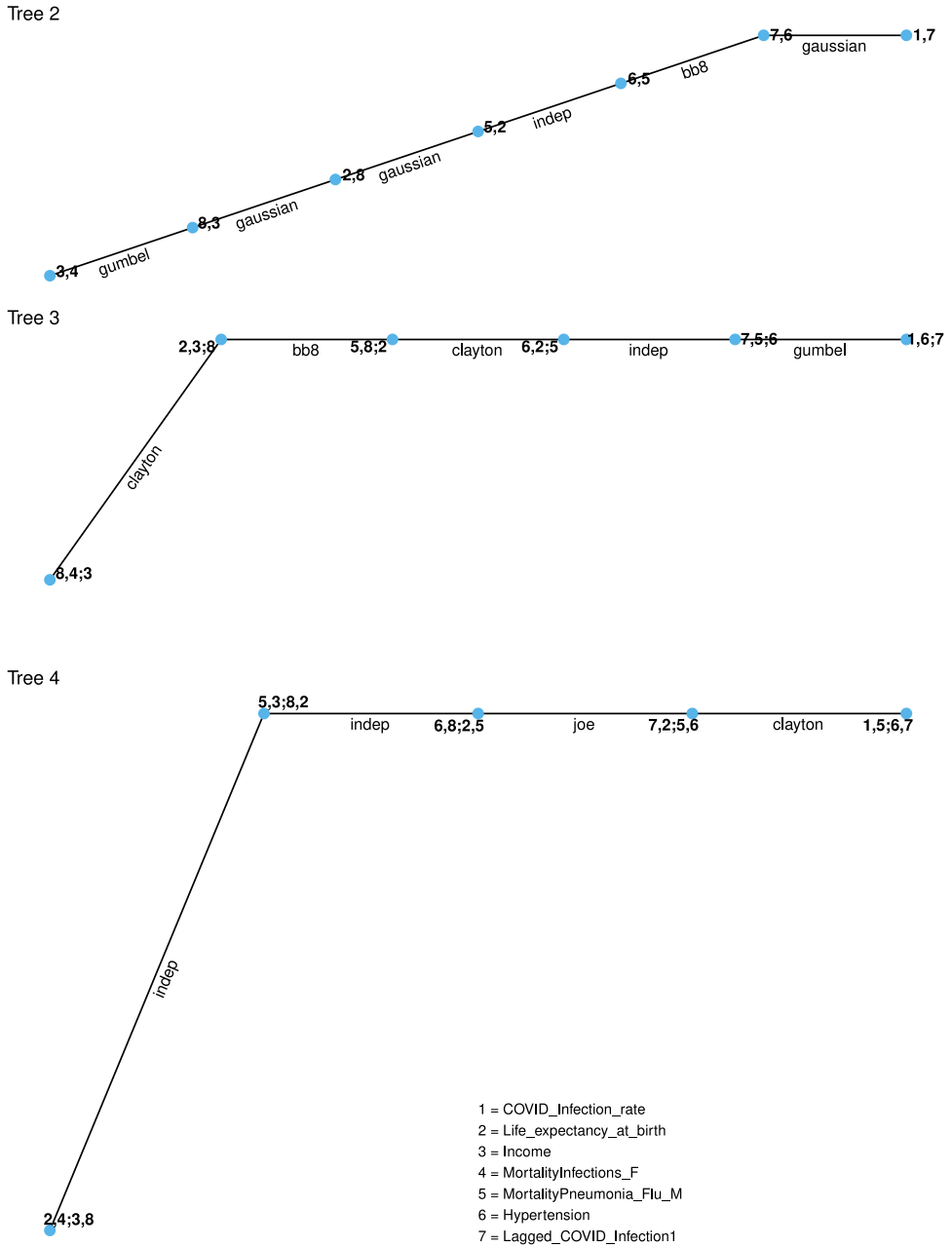|	Tree	Edge	Conditioned	Conditioning	Family	Rotation	Parameter1	Parameter2	df	tau
1	1	1	1, 7		bb8	180	8.000		2	0.678
2	1	2	7, 6		indep	0			0	0.000
3	1	3	6, 5		indep	0			0	0.000
4	1	4	5, 2		bb7	0	1.000	0.578	2	0.224
5	1	5	2, 8		joe	180	1.315		1	0.152
6	1	6	8, 3		clayton	0	0.888		1	0.307
7	1	7	3, 4		clayton	0	0.725		1	0.266
8	2	1	1, 6	7	gaussian	0	−0.218		1	−0.140
9	2	2	7, 5	6	bb8	180	3.343	0.909	2	0.477
10	2	3	6, 2	5	indep	0			0	0.000
11	2	4	5, 8	2	gaussian	0	0.367		1	0.239
12	2	5	2, 3	8	gaussian	0	0.452		1	0.299
13	2	6	8, 4	3	gumbel	180	1.056		1	0.053
14	3	1	1, 5	6, 7	gumbel	180	1.135		1	0.119
15	3	2	7, 2	5, 6	indep	0			0	0.000
16	3	3	6, 8	2, 5	clayton	0	0.731		1	0.268
17	3	4	5, 3	8, 2	bb8	180	3.459	0.858	2	0.448
18	3	5	2, 4	3, 8	clayton	0	0.241		1	0.108
19	4	1	1, 2	5, 6, 7	clayton	90	0.097		1	−0.046
20	4	2	7, 8	2, 5, 6	joe	180	1.193		1	0.100
21	4	3	6, 3	8, 2, 5	indep	0			0	0.000
22	4	4	5, 4	3, 8, 2	indep	0			0	0.000
23	5	1	1, 8	2, 5, 6, 7	frank	0	−1.145		1	−0.126
24	5	2	7, 3	8, 2, 5, 6	frank	0	4.054		1	0.392
25	5	3	6, 4	3, 8, 2, 5	bb8	0	1.198	0.998	2	0.100
26	6	1	1, 3	8, 2, 5, 6, 7	bb8	0	1.355	0.963	2	0.136
27	6	2	7, 4	3, 8, 2, 5, 6	indep	0			0	0.000
28	7	1	1, 4	3, 8, 2, 5, 6, 7	bb8	0	1.299	0.977	2	0.125

See Fig. A. 1, Table A. 1 and Fig. A. 2.
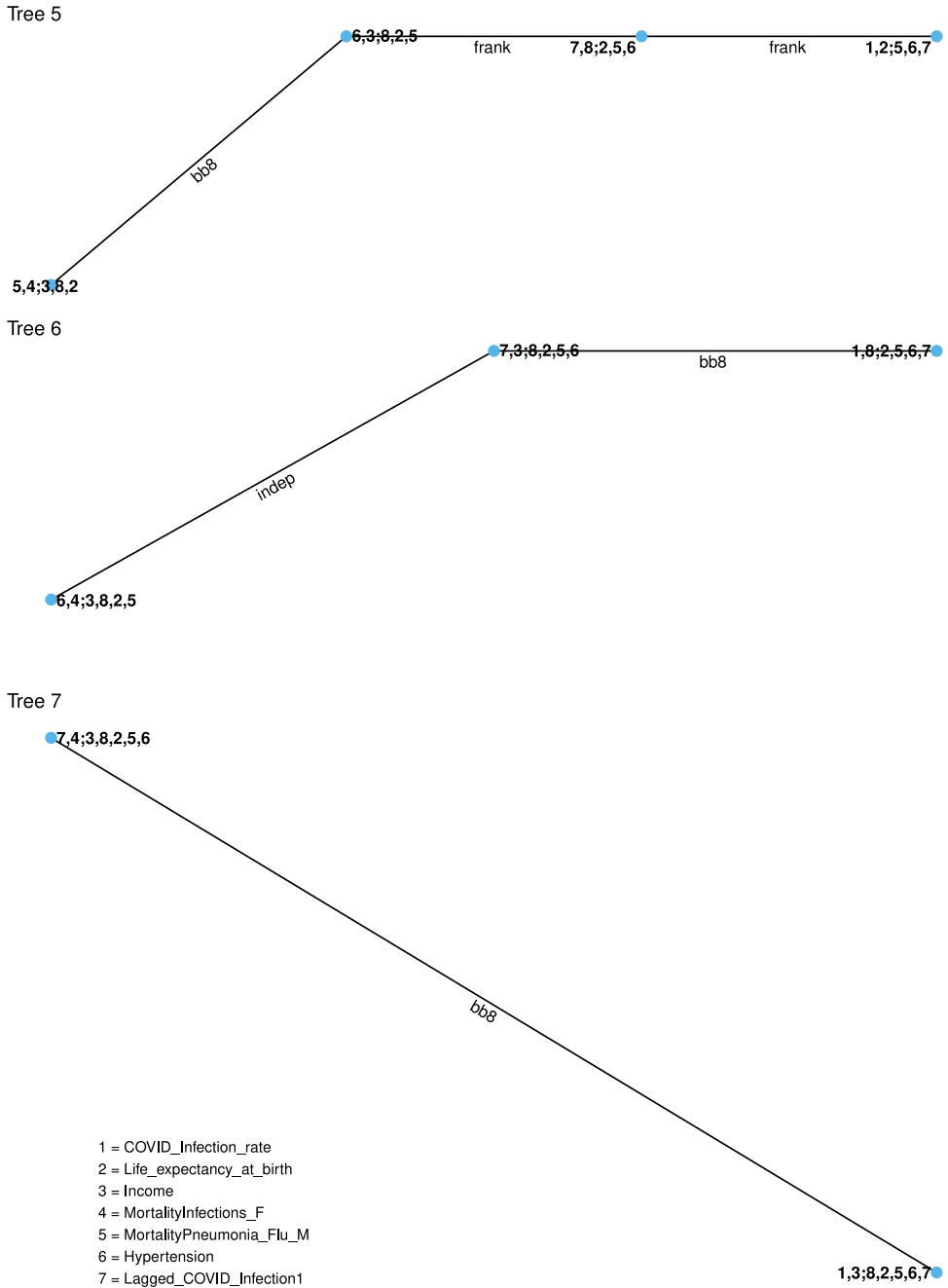
Fig. A. 1. *DVQR_1* model: the D-vine trees.

Tree 5

Tree 6

Tree 7

1 = COVID_Infection_rate
2 = Life_expectancy_at_birth
3 = Income
4 = MortalityInfections_F
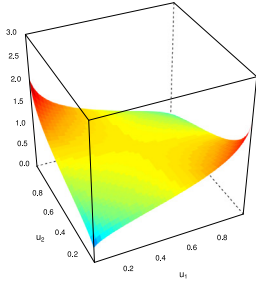5 = MortalityPneumonia_Flu_M
6 = Hypertension
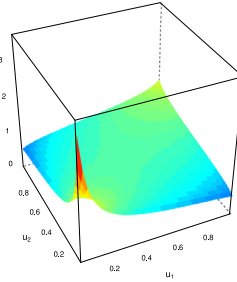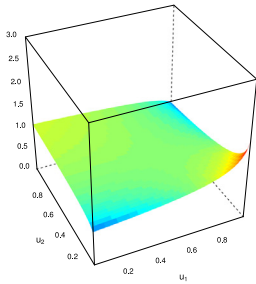7 = Lagged_COVID_Infection1
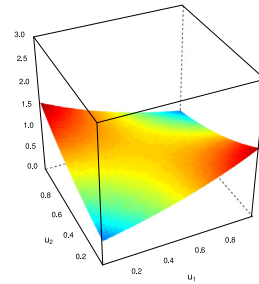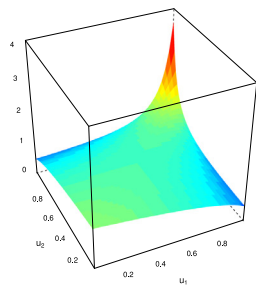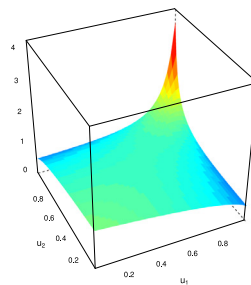
Fig. A. 1. (*continued*).

The conditional copula density of the edge 1 of Tree 2

The conditional copula density of the edge 1 of Tree 3

The conditional copula density of the edge 1 of Tree 4

The conditional copula density of the edge 1 of Tree 5

The conditional copula density of the edge 1 of Tree 6

The conditional copula density of the edge 1 of Tree 7

Fig. A. 2. *DVQR_1* model: the estimated conditional bivariate Copula densities with uniform margins between the outcome and its covariates.

Appendix B. D-vine for the *DVQR_2* model

See Fig. B. 1, Table B. 1 and Fig. B. 2.

Table B. 1

The D vine specification for the *DVQR_2* model.

	Tree	Edge	Conditioned	Conditioning	Family	Rotation	Parameter1	Parameter2	df	tau
1	1	1	1, 11		bb8	180	8.000	0.689	2	0.611
2	1	2	11, 8		bb8	180	3.982	0.747	2	0.414
3	1	3	8, 10		frank	0	−1.911		1	−0.205
4	1	4	10, 6		t	0	−0.255	4.676	2	−0.164
5	1	5	6, 7		bb7	180	1.347	1.669	2	0.495
6	1	6	7, 9		joe	270	1.140		1	−0.075
7	1	7	9, 5		bb1	0	0.000	1.155	2	0.134
8	1	8	5, 3		clayton	0	0.725		1	0.266
9	1	9	3, 2		gaussian	0	0.538		1	0.361
10	1	10	2, 4		clayton	0	0.405		1	0.168
11	2	1	1, 8	11	gumbel	180	1.303		1	0.233
12	2	2	11, 10	8	frank	0	−1.906		1	−0.205
13	2	3	8, 6	10	gaussian	0	0.143		1	0.091
14	2	4	10, 7	6	gumbel	90	1.271		1	−0.213
15	2	5	6, 9	7	indep	0			0	0.000
16	2	6	7, 5	9	gaussian	0	0.267		1	0.172
17	2	7	9, 3	5	indep	0			0	0.000
18	2	8	5, 2	3	indep	0			0	0.000
19	2	9	3, 4	2	gaussian	0	0.320		1	0.207
20	3	1	1, 10	8, 11	frank	0	−1.985		1	−0.212
21	3	2	11, 6	10, 8	frank	0	0.791		1	0.087
22	3	3	8, 7	6, 10	clayton	0	0.608		1	0.233
23	3	4	10, 9	7, 6	bb7	180	1.108	0.199	2	0.139
24	3	5	6, 5	9, 7	clayton	0	0.194		1	0.089
25	3	6	7, 3	5, 9	bb8	180	2.523	0.825	2	0.299
26	3	7	9, 2	3, 5	indep	0			0	0.000
27	3	8	5, 4	2, 3	bb7	0	1.582	0.333	2	0.327
28	4	1	1, 6	10, 8, 11	frank	0	1.506		1	0.164
29	4	2	11, 7	6, 10, 8	joe	180	1.224		1	0.113
30	4	3	8, 9	7, 6, 10	indep	0			0	0.000
31	4	4	10, 5	9, 7, 6	joe	270	1.201		1	−0.103
32	4	5	6, 3	5, 9, 7	frank	0	−1.776		1	−0.191
33	4	6	7, 2	3, 5, 9	gaussian	0	−0.394		1	−0.258
34	4	7	9, 4	2, 3, 5	gaussian	0	−0.173		1	−0.111
35	5	1	1, 7	6, 10, 8, 11	joe	180	1.130		1	0.070
36	5	2	11, 9	7, 6, 10, 8	indep	0			0	0.000
37	5	3	8, 5	9, 7, 6, 10	clayton	0	0.690		1	0.256
38	5	4	10, 3	5, 9, 7, 6	gaussian	0	−0.180		1	−0.115
39	5	5	6, 2	3, 5, 9, 7	frank	0	−2.543		1	−0.266
40	5	6	7, 4	2, 3, 5, 9	gaussian	0	0.224		1	0.144
41	6	1	1, 9	7, 6, 10, 8, 11	gaussian	0	−0.173		1	−0.111
42	6	2	11, 5	9, 7, 6, 10, 8	indep	0			0	0.000
43	6	3	8, 3	5, 9, 7, 6, 10	clayton	0	1.204		1	0.376
44	6	4	10, 2	3, 5, 9, 7, 6	indep	0			0	0.000
45	6	5	6, 4	2, 3, 5, 9, 7	frank	0	−0.995		1	−0.110
46	7	1	1, 5	9, 7, 6, 10, 8, 11	bb8	0	1.237	0.989	2	0.109
47	7	2	11, 3	5, 9, 7, 6, 10, 8	gaussian	0	0.282		1	0.182
48	7	3	8, 2	3, 5, 9, 7, 6, 10	indep	0			0	0.000
49	7	4	10, 4	2, 3, 5, 9, 7, 6	indep	0			0	0.000
50	8	1	1, 3	5, 9, 7, 6, 10, 8, 11	bb8	0	1.593	0.953	2	0.203
51	8	2	11, 2	3, 5, 9, 7, 6, 10, 8	joe	90	1.160		1	−0.084
52	8	3	8, 4	2, 3, 5, 9, 7, 6, 10	gaussian	0	0.152		1	0.097
53	9	1	1, 2	3, 5, 9, 7, 6, 10, 8, 11	clayton	90	0.085		1	−0.041
54	9	2	11, 4	2, 3, 5, 9, 7, 6, 10, 8	bb8	90	1.703	0.853	2	−0.169
55	10	1	1, 4	2, 3, 5, 9, 7, 6, 10, 8, 11	joe	90	1.209		1	−0.107
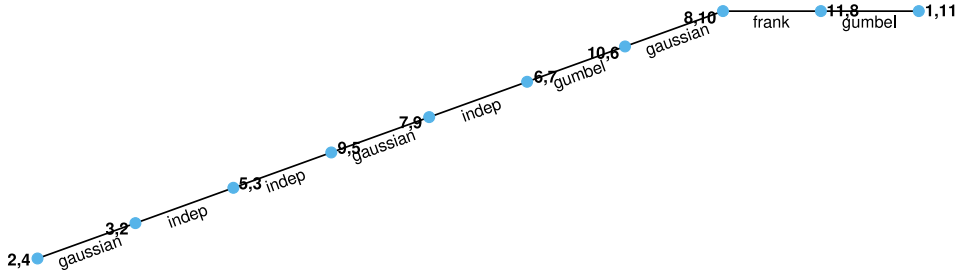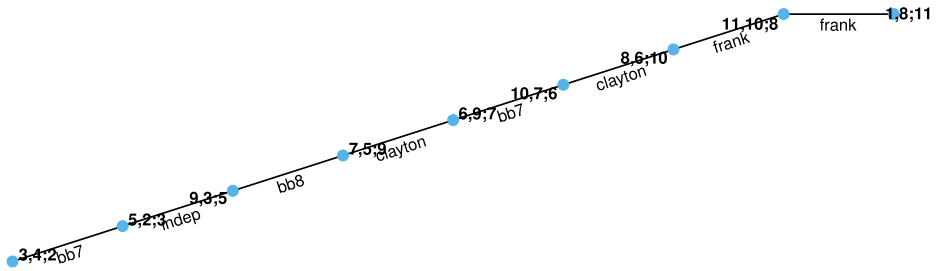
Fig. B. 1. *DVQR_2* model: the D-vine trees.
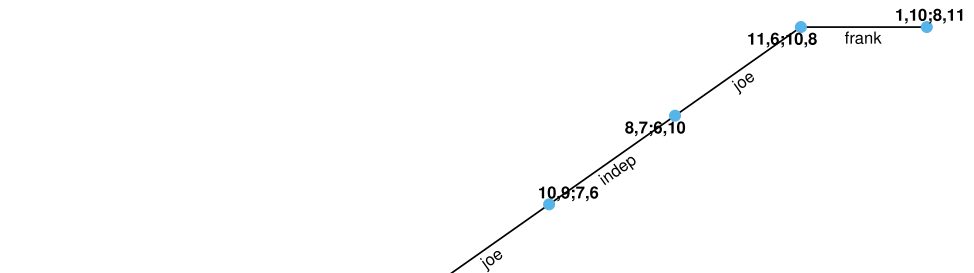
1 = COVID_Infection_rate
2 = Life_expectancy_at_birth
3 = Income
4 = MortalityInfections_M
5 = MortalityInfections_F
6 = MortalityCancer_M
7 = MortalityCancer_F
8 = MortalityPneumonia_Flu_M
9 = Hypertension
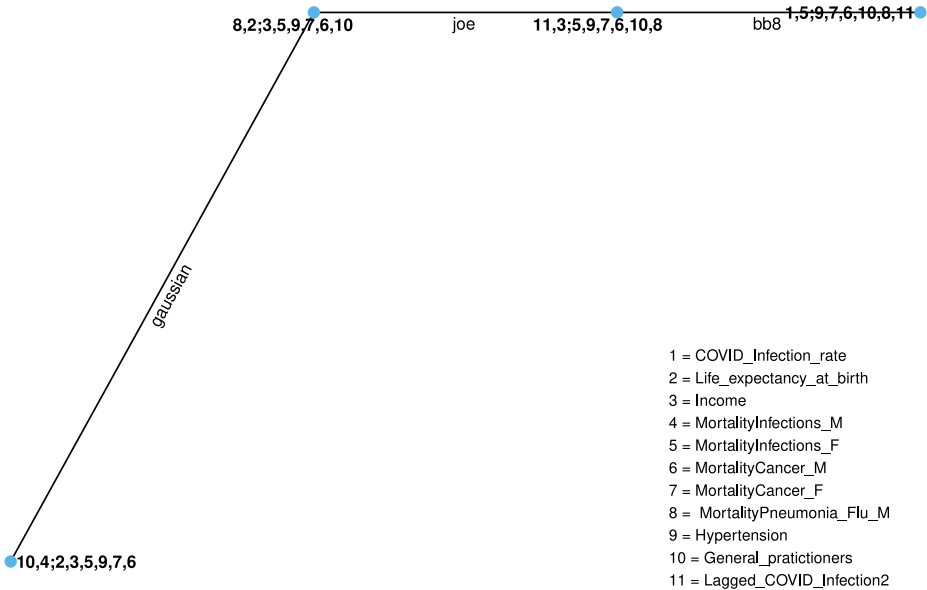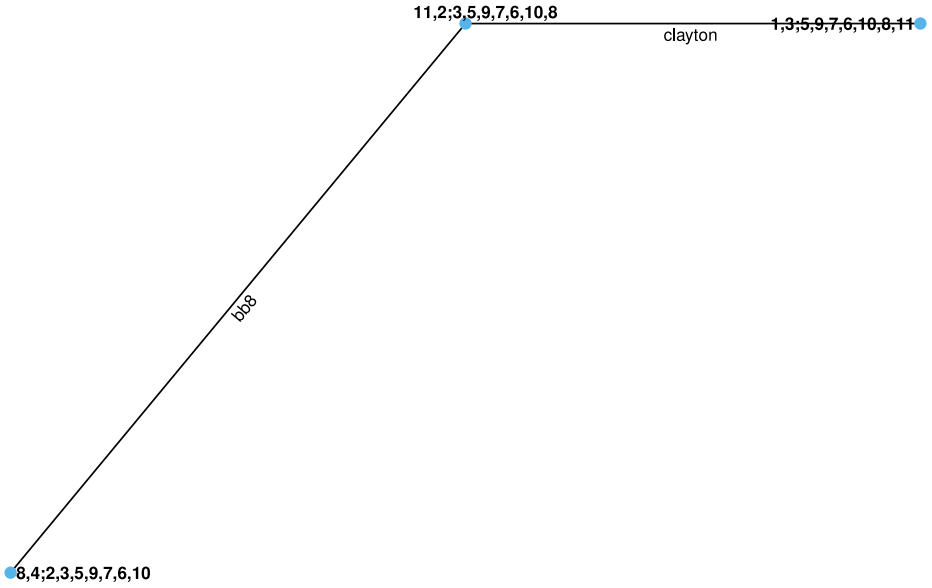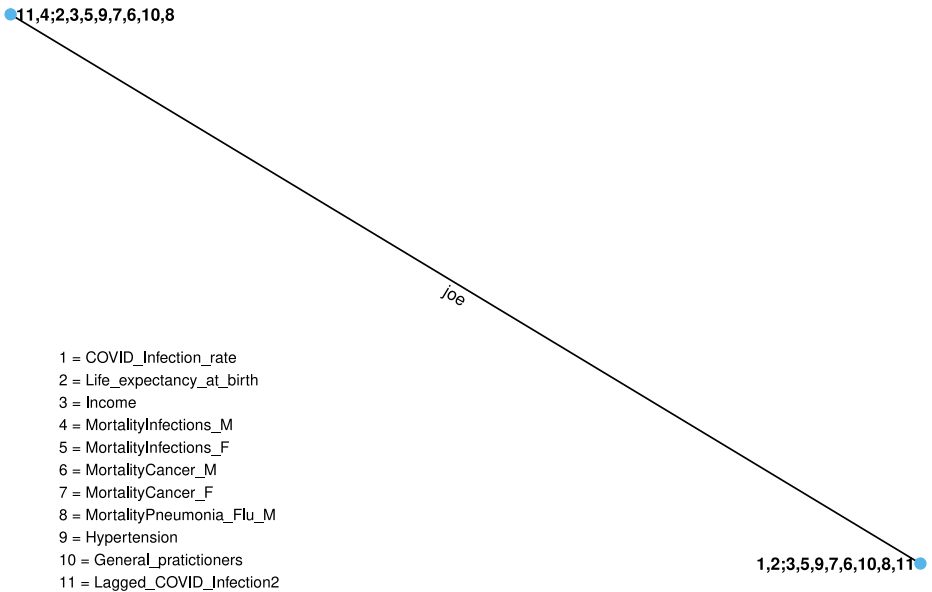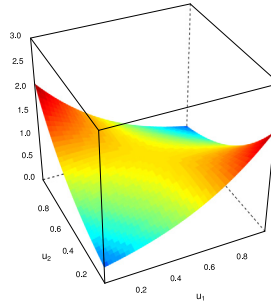10 = General_praticioners
11 = Lagged_COVID_Infection2

Tree 5

Tree 6

1 = COVID_Infection_rate
2 = Life_expectancy_at_birth
3 = Income
4 = MortalityInfections_M
5 = MortalityInfections_F
6 = MortalityCancer_M
7 = MortalityCancer_F
8 = MortalityPneumonia_Flu_M
9 = Hypertension
10 = General_pratictioners
11 = Lagged_COVID_Infection2

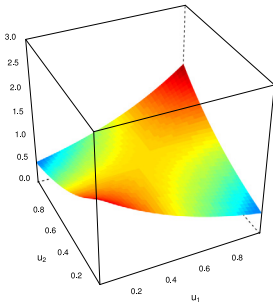**Fig. B. 1.** (*continued*).

Tree 7

Tree 8

1 = COVID_Infection_rate
2 = Life_expectancy_at_birth
3 = Income
4 = MortalityInfections_M
5 = MortalityInfections_F
6 = MortalityCancer_M
7 = MortalityCancer_F
8 = MortalityPneumonia_Flu_M
9 = Hypertension
10 = General_pratictioners
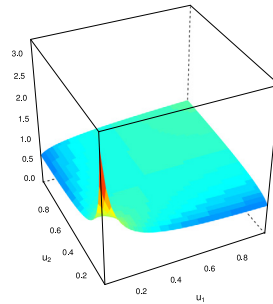11 = Lagged_COVID_Infection2

Fig. B. 1. (*continued*).

Tree 9

11,2;3,5,9,7,6,10,8 — clayton — **1,3;5,9,7,6,10,8,11**

bb8

8,4;2,3,5,9,7,6,10

Tree 10

11,4;2,3,5,9,7,6,10,8

joe

1,2;3,5,9,7,6,10,8,11

1 = COVID_Infection_rate
2 = Life_expectancy_at_birth
3 = Income
4 = MortalityInfections_M
5 = MortalityInfections_F
6 = MortalityCancer_M
7 = MortalityCancer_F
8 = MortalityPneumonia_Flu_M
9 = Hypertension
10 = General_pratictioners
11 = Lagged_COVID_Infection2

Fig. B. 1. (*continued*).
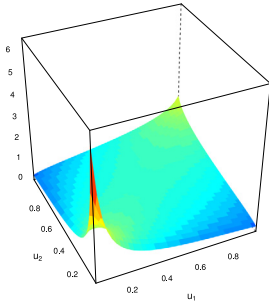
The conditional copula density of the edge 1 of Tree 2

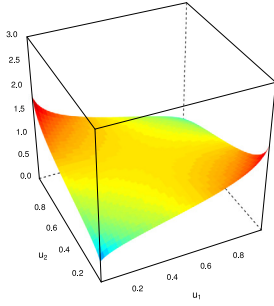The conditional copula density of the edge 1 of Tree 3

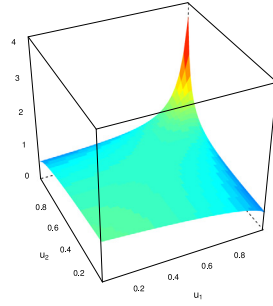The conditional copula density of the edge 1 of Tree 4
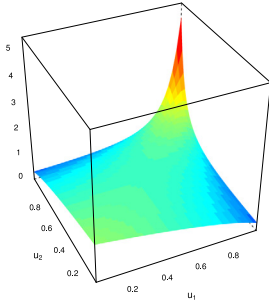

The conditional copula density of the edge 1 of Tree 5

Fig. B. 2. *DVQR_2* model: the estimated conditional bivariate Copula densities with uniform margins between the outcome and its covariates.
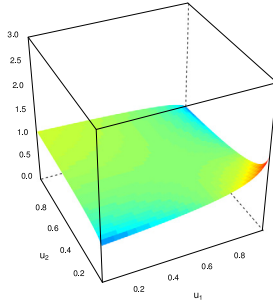
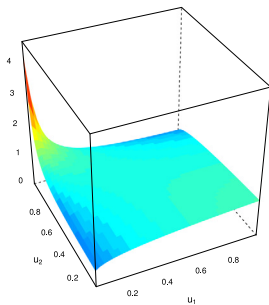The conditional copula density of the edge 1 of Tree 6

The conditional copula density of the edge 1 of Tree 7

The conditional copula density of the edge 1 of Tree 8

The conditional copula density of the edge 1 of Tree 9

The conditional copula density of the edge 1 of Tree 10

Fig. B. 2. (*continued*).

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insur. Math. Econ. 44 (2), 182–198.

AIFA, 2020. The Medicines Utilisation Monitoring Centre. National Report on Medicines use in Italy. Year 2019. Technical report, ISBN: 9788894494365.

Anselin, L., 1988. Spatial Econometrics: Methods And Models. Kluwer Academic Publishers, Dordrecht; Boston.

Aràndiga, F., Baeza, A., Cordero-Carrión, I., Donat, R., Martí, M.C., Mulet, P., Yáñez, D.F., 2020. A spatial-temporal model for the evolution of the COVID-19 pandemic in Spain including mobility. Mathematics 8 (10:1677), http://dx.doi.org/10.3390/math8101677.

Bartolucci, F., Farcomeni, A., 2021. A spatio-temporal model based on discrete latent variables for the analysis of COVID-19 incidence. Spatial Stat. 100504.

Bedford, T., Cooke, R.M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. Intell. 32 (1), 245–268.

Bedford, T., Cooke, R.M., 2002. Vines: A new graphical model for dependent random variables.. Ann. Stat. 30 (4), 1031–1068..

Bernard, C., Czado, C., 2015. Conditional quantiles and tail dependence. J. Multivariate Anal. 138, 104–126.

Bertuzzo, E., Mari, L., Pasetto, D., Miccoli, S., Casagrandi, R., Gatto, M., Rinaldo, R., 2020. The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures. Nature Commun. 11 (4264), http://dx.doi.org/10.1038/s41467-020-18050-2.

Chang, B., Joe, H., 2019. Prediction based on conditional distributions of vine copulas. Comput. Statist. Data Anal. 139, 45–63.

Cicchetti, A., Giorgio, L., Di Paolo, M.G., Daugbjerg, S., Laurita, R., Morandi, F., Tattoli, A., Villani, L., 2021. Sustainability and Resilience in the Italian Health System. Technical report, The Partnership for Health System Sustainability and Resilience (PHSSR).

Cooke, R.M., Joe, H., Chang, B., 2015. Vine regression, in: Resources For The Future Discussion Paper, pp. 15–52.

Czado, C., 2010. Pair-copula constructions of multivariate copulas. In: Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (Eds.), Copula Theory And Its Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 93–109.

Farcomeni, A., Maruotti, A., Divino, F., Jona-Lasinio, G., Lovison, G., 2020. An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions. Biom. J. http://dx.doi.org/10.1002/bimj.202000189.

Fenske, N., Kneib, T., Hothorn, T., 2011. Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. J. Am. Stat. Assoc. 106 (494), 494–510.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., Colaneri, M., 2020. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. 26 (5064), 855–860. http://dx.doi.org/10.1038/s41591-020-0883-7.

Girardi, P., Greco, L., Mameli, V., Musio, M., Racugno, W., Ruli, E., Ventura, L., 2020. Robust inference for non-linear regression models from the tsallis score: Application to coronavirus disease 2019 contagion in Italy. Stat 9 (1), e309. http://dx.doi.org/10.1002/sta4.309.

Giuliani, D., Dickson, M.M., Espa, G., Santi, F., 2020. Modelling and predicting the spatio-temporal spread of COVID-19 in Italy. BMC Infect. Dis. 20 (700), http://dx.doi.org/10.1186/s12879-020-05415-7.

Hobæk Haff, I., Aas, K., Frigessi, A., 2010. On the simplified pair-copula construction — Simply useful or too simplistic? J. Multivariate Anal. 101 (5), 1296–1310.

Joe, H., 1996. Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. In: Lecture Notes-Monograph Series. JSTOR, pp. 120–141.

Joe, H., 1997. Multivariate Models And Multivariate Dependence Concepts. CRC Press.

Kang, D., Choi, H., Kim, J.-H., Choi, J., 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. Int. J. Infect. Dis. 94, 96–102. http://dx.doi.org/10.1016/j.ijid.2020.03.076.

Koenker, R., 2011. Additive models for quantile regression: Model selection and confidence bandaids. Braz. J. Probab. Stat. 25 (3), 239–262.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica: J. Econom. Soc. 46 (1), 33–50.

Kraus, D., Czado, C., 2017. D-vine copula based quantile regression. Comput. Statist. Data Anal. 110, 1–18.

Kurowicka, D., Cooke, R., 2006. Uncertainty Analysis With High Dimensional Dependence Modelling. John Wiley & Sons, Ltd..

Lee, D., Robertson, C., Marques, D., 2021. Quantifying the small-area spatio-temporal dynamics of the Covid-19 pandemic in Scotland during a period with limited testing capacity. Spatial Stat. 100508.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus–Infected pneumonia. N. Engl. J. Med. 382 (13), 1199–1207, DOI: 10.1056/NEJMoa2001316.

Liu, Y., Zhou, Y., Chen, Y., Wang, D., Wang, Y., Zhu, Y., 2020. Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. Renew. Energy 146, 1101–1112. http://dx.doi.org/10.1016/j.renene.2019.07.053.

Martey, E.N., Attoh-Okine, N., 2019. Analysis of train derailment severity using vine copula quantile regression modeling. Transp. Res. C (ISSN: 0968-090X) 105, 485–503. http://dx.doi.org/10.1016/j.trc.2019.06.015.

Mauro, M., Giancotti, M., 2021. Italian responses to the COVID-19 emergency: Overthrowing 30 years of health reforms? Health Policy (ISSN: 0168-8510) 125 (4), 548–552. http://dx.doi.org/10.1016/j.healthpol.2020.12.015.

Mollalo, A., Vahedi, B., Rivera, K.M., 2020. Gis-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci. Total Environ. 728, 138884. http://dx.doi.org/10.1016/j.scitotenv.2020.138884.

Mugnai, G., Bilato, C., 2020. Covid-19 in Italy: Lesson from the veneto region. Eur. J. Intern. Med. 77, 161–162. http://dx.doi.org/10.1002/sta4.309.

Murgante, B., Borruso, G., Balletto, G., Castiglia, P., Dettori, M., 2020. Why Italy first? Health, geographical and planning aspects of the COVID-19 outbreak. Sustainability 12 (12:5064), http://dx.doi.org/10.3390/su12125064.

Nagler, T., 2018. Vinereg: D-vine quantile regression. R package version 0.5.0.

Nguyen-Huy, T., Deo, R.C., Mushtaq, S., An-Vo, D.-A., Khan, S., 2018. Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach. Eur. J. Agron. 98, 65–81. http://dx.doi.org/10.1016/j.eja.2018.05.006.

Niemierko, R., Töppel, J., Tränkler, T., 2019. A D-vine copula quantile regression approach for the prediction of residential heating energy consumption based on historical data. Appl. Energy 233–234, 691–708. http://dx.doi.org/10.1016/j.apenergy.2018.10.025.

Remuzzi, A., Remuzzi, G., 2020. COVID-19 And Italy: what next? Lancet 395 (10261), 1225–1228.

Sahu, S.K., Böhning, D., 2021. BayesIan spatio-temporal joint disease mapping of Covid-19 cases and deaths in local authorities of England. Spatial Stat. 100519. http://dx.doi.org/10.1016/j.spasta.2021.100519.

Schallhorn, N., Kraus, D., Nagler, T., Czado, C., 2017. D-vine quantile regression with discrete variables. arXiv preprint arXiv:1705.08310.

Sklar, A., 1959. Fonctions de repartition a n dimensions et leurs marges.. Publications De L'Institut De Statistique De L'Universite De Paris 8, 229–231.

Tepegjozova, M., Zhou, J., Claeskens, G., Czado, C., 2021. Nonparametric C-and D-vine based quantile regression. arXiv preprint arXiv:2102.04873.

Vitale, V., D'Urso, P., De Giovanni, L., 2021. Spatio-temporal object-oriented Bayesian network modelling of the COVID-19 Italian outbreak data. Spatial Stat. 100529.