

Article

# Copula Dynamic Conditional Correlation and Functional Principal Component Analysis of COVID-19 Mortality in the United States

Jong-Min Kim 

Statistics Discipline, Division of Science and Mathematics, University of Minnesota-Morris,  
Morris, MN 56267, USA; jongmink@morris.umn.edu

**Abstract:** This paper shows a visual analysis and the dependence relationships of COVID-19 mortality data in 50 states plus Washington, D.C., from January 2020 to 1 September 2022. Since the mortality data are severely skewed and highly dispersed, a traditional linear model is not suitable for the data. As such, we use a Gaussian copula marginal regression (GCMR) model and vine copula-based quantile regression to analyze the COVID-19 mortality data. For a visual analysis of the COVID-19 mortality data, a functional principal component analysis (FPCA), graphical model, and copula dynamic conditional correlation (copula-DCC) are applied. The visual from the graphical model shows five COVID-19 mortality equivalence groups in the US, and the results of the FPCA visualize the COVID-19 daily mortality time trends for 50 states plus Washington, D.C. The GCMR model investigates the COVID-19 daily mortality relationship between four major states and the rest of the states in the US. The copula-DCC models investigate the time-trend dependence relationship between the COVID-19 daily mortality data of four major states.

**Keywords:** COVID-19; mortality; functional PCA; Gaussian copula regression; graphical model; vine copula-based quantile regression; copula-DCC

**MSC:** 62P10



**Citation:** Kim, J.-M. Copula Dynamic Conditional Correlation and Functional Principal Component Analysis of COVID-19 Mortality in the United States. *Axioms* **2022**, *11*, 619. <https://doi.org/10.3390/axioms11110619>

Academic Editor: Radko Mesiar

Received: 5 October 2022

Accepted: 4 November 2022

Published: 7 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The COVID-19 outbreak began in the city of Wuhan in China's Hubei province and quickly became a global pandemic. The pandemic paralyzed the public health systems of many countries throughout the world, resulting in the death of millions. The COVID-19 pandemic negatively affected the global economy, which has now been negatively impacted further by the 2022 Russia–Ukraine war and the skyrocketing prices of commodities and oil. Throughout this pandemic period, the negative events that have affected the global economy have occurred sequentially and continuously like a tsunami.

To better protect world citizens, we need to think about how to prepare for negative events such as these. Analyzing disasters such as the COVID-19 pandemic can provide insights into how we can be better prepared for future disasters. In this paper, we focus on US COVID-19 mortality data analyses to examine how the US COVID-19 mortality rate spread geographically in different directions and formed clusters, as well as the relationships between major states and their neighboring states in terms of COVID-19 mortalities. Through this study, we attempt to help to reduce the number of mortalities in future pandemics or endemics by learning how COVID-19 affected US states' mortalities so we can revise US public health measures accordingly. To illustrate the visual and data analyses, we apply functional data analysis (FDA) [1,2] and copula dependence methods [3,4] to US COVID-19 mortality time-course data.

Before applying these methods to the US COVID-19 mortality data, we review the COVID-19-related FDA and copula research papers studied so far. [5] studied the canonical

correlation between confirmed and mortality cases in US COVID-19 data and used functional principal component analysis (FPCA) to examine the types of variations in the data. Forecasting based on dynamic FPCA with the cumulative confirmed cases in the US was also explored in [5]. The time-series data of the COVID-19 confirmed and mortality cases during the lockdown in Wuhan were analyzed using FPCA and the functional canonical correlation analysis methods in [6]. FDA was used by [7] to model daily hospitalized, deceased, intensive care unit (ICU), and return-home patient numbers throughout the COVID-19 outbreak for the number of vaccinations, mortalities, infected people, recovered people, and tested people in France. The imputation of missing data of COVID-19 hospitalized and intensive care curves in several Spanish regions using a function-on-function regression model to estimate the missing values of the functional responses associated with the hospitalized and intensive care curves was considered in [8]. By looking at a literature review of FDA modeling for COVID-19 data, we can ensure that FPCA is an effective clustering visualization analysis for time-course data as it provides a more informative way of examining the sample covariance structure than PCA. Another useful statistical method we consider for US COVID-19 mortality data analysis is the copula dependence method, as the copula does not require the independence, linearity, and normality of the residuals (see references [3,4,9–11]). In particular, we employ both the Gaussian copula marginal regression (GCMR) model by [12] and vine copulas (proposed by [13] and explained in more detail by [14–16]). The GCMR can deal with heteroscedasticity and the non-normal distribution of the residuals by including a dispersion parameter to model and adjust for non-constant variance. Vine copulas are a graphical model that represents a  $d$ -dimensional multivariate density in a hierarchical manner [17,18]. The main determinants of the COVID-19 spread in Italy were investigated by [19] through the use of a D-vine copula-based quantile regression with a spatial autoregressive component for considering spatial dependence. Vine copulas were used as they enhance model flexibility and can account for nonlinear relationships and tail dependencies. Vine copulas also provide the model selection procedure with a rank of the covariates based on their explanatory power with respect to the outcome. The semi/non-parametric estimators of the health concentration (HC) curve that quantify inequalities in COVID-19 infections and mortalities and help to identify the social classes that are most at risk of infection and death from the virus in terms of the copula function, as well as the copula-based estimators of the health Gini coefficient, were derived by [20]. Vine copulas applied to COVID-19 data were exploited for the dependencies between the different sources of information as they combine structured datasets retrieved from official sources and a big unstructured dataset of information collected from social media from [21].

The remainder of this paper is organized as follows. Section 2 provides a description of the daily and cumulative mortality data. Graphical visualization by FPCA is introduced in Section 3. Section 4 describes the copula methods (GCMR model, vine copula-based quantile regression, and copula dynamic conditional correlation) used to analyze the data and the discussion is presented in Section 5.

## 2. Data Description

We downloaded US COVID-19 daily cumulative mortality data from the USA Facts website, which can be found here: [https://static.usafacts.org/public/data/covid-19/covid\\_deaths\\_usafacts.csv](https://static.usafacts.org/public/data/covid-19/covid_deaths_usafacts.csv), accessed on 2 September 2022. The time period of the data collected from the website was 22 January 2020 to 1 September 2022. We converted the US COVID-19 daily cumulative mortality data to the daily COVID-19 mortality data in 50 states and Washington, D.C. After converting the cumulative COVID-19 mortality data to the daily COVID-19 mortality data in 50 states and Washington, D.C., we found that the original data needed to be manually corrected due to some of the daily cumulative COVID-19 mortality data for some states being inconsistent and recorded in such a way that the previous day's cumulative number of COVID-19 mortalities was higher than the current day's number of mortalities.

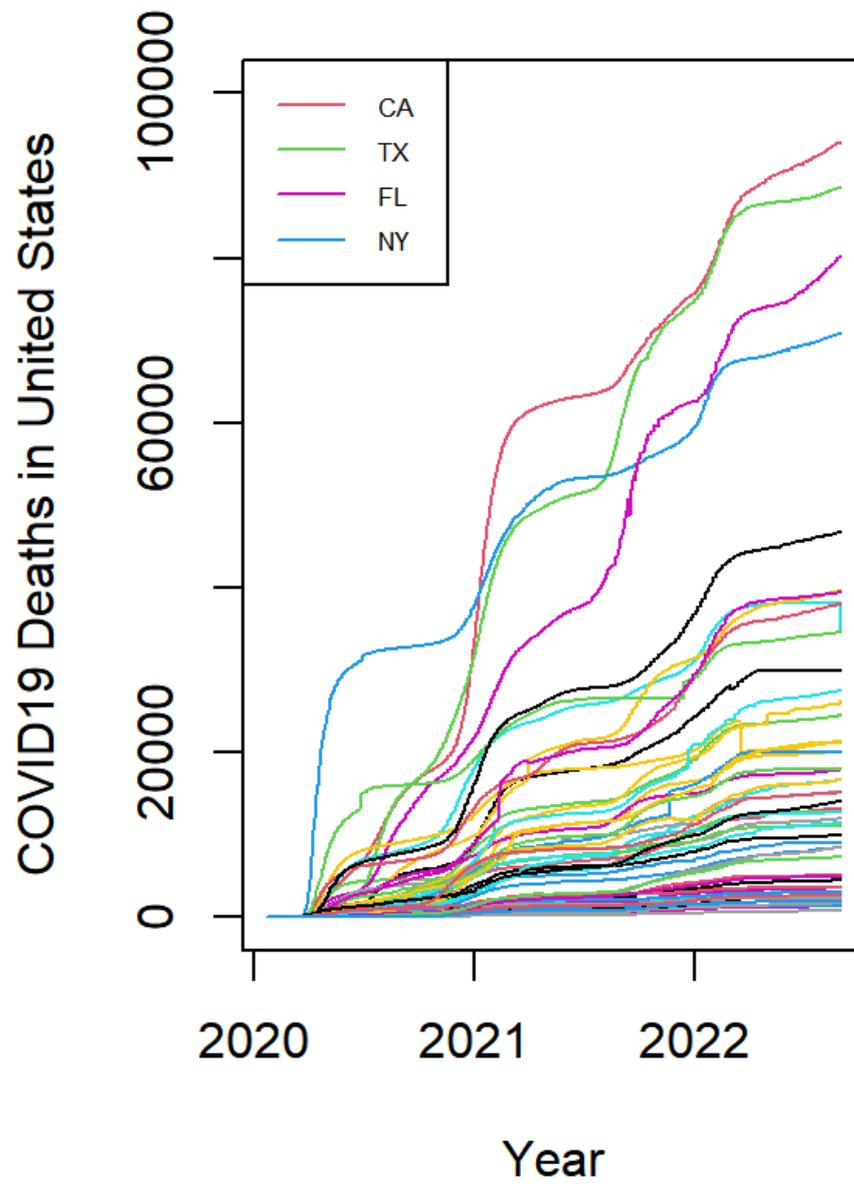
Table 1 shows the summary statistics for the daily COVID-19 mortality data in 50 states and Washington D.C. and the 2022 US state populations. In Table 1, we can see that for our dataset, the state with the highest total number of COVID-19 mortalities was California (CA) with 93,924, followed by Texas (TX) with 88,578, Florida (FL) with 80,027, and New York (NY) with 70,877. All 50 states and Washington, D.C. had a positive skewness distribution of daily COVID-19 mortalities. This means that more people could die in the near future because of COVID-19. All 50 states and Washington, D.C. also had high kurtosis. This means that the daily number of COVID-19 mortalities had high variation clustering similar to a highly volatile financial market pattern. Southern states such as AL (0.40%), AZ (0.41%), MS (0.43%), OK (0.42%), and WV (0.41%) had higher COVID-19 mortality rates based on 2022 state populations than other states in the US. In Figure 1, we visualize the total number of COVID-19 mortalities in each state in the US. In Figure 1, it can be seen that the number of COVID-19 mortalities in New York rapidly increased at the beginning of the pandemic, but California, Texas, and Florida eventually surpassed New York in terms of COVID-19 mortalities and now lead the US in mortalities.

**Table 1.** Summary Statistics for US Daily COVID-19 Mortalities and 2022 US State Populations.

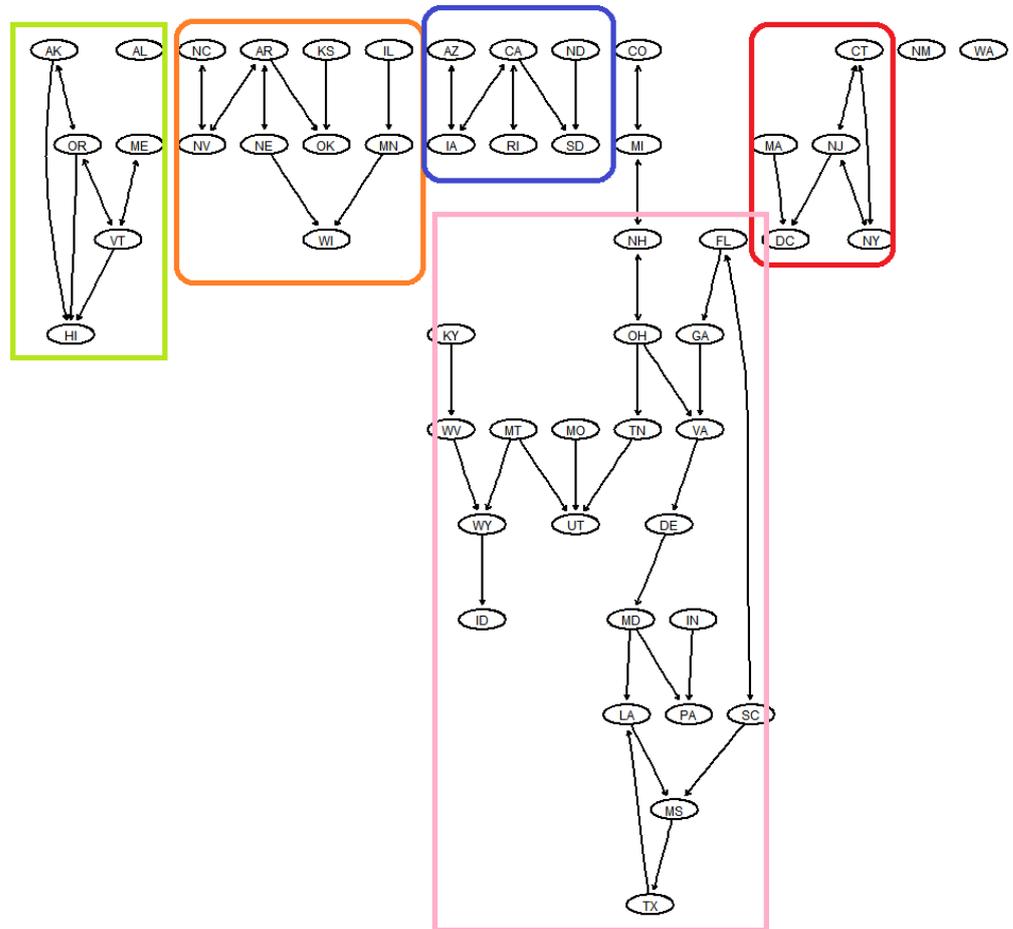
	AK	AL	AR	AZ	CA	CO	CT	DC	DE	FL	GA	HI	IA	ID	IL	IN	KS
Mean	1.3	21.2	12.5	31.3	98.6	13.8	11.6	1.5	3.2	84	41.7	1.7	10.4	5.4	40	25.7	9.4
Median	0	7	6	6	54	6	1	0	0	44	25	0	0	1	22	13	0
SD	5.2	39	19.5	58.1	142.2	25.3	23.3	2.9	7.6	152.2	92.8	4.1	29.1	9.6	53.1	58.7	27.4
Kurtosis	87	21.4	70.6	14.1	6	36.2	13.5	56.5	104.6	29.2	516.8	44.4	108	9.7	6.5	471.8	44.6
Skewness	8.5	3.9	5.8	3.3	2.4	4.8	3.2	5.4	7.9	4.7	19.7	5	8.1	2.8	2.3	18.6	5.5
Minimum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum	68	389	322	498	704	309	204	45	133	1552	2499	59	518	65	401	1546	371
Total	1281	20,160	11,923	29,852	93,924	13,148	11,034	1382	3042	80,027	39,772	1644	9940	5115	38,161	24,454	8958
Count	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953
Population	738,023	5,073,187	3,030,646	7,303,398	39,995,077	5,922,618	3,612,314	644,743	1,008,350	22,085,563	10,916,760	1,474,265	3,219,171	1,893,410	12,808,884	6,845,874	2,954,832
Mortality Rate	0.17%	0.40%	0.39%	0.41%	0.23%	0.22%	0.31%	0.21%	0.30%	0.36%	0.36%	0.11%	0.31%	0.27%	0.30%	0.36%	0.30%
	KY	LA	MA	MD	ME	MI	MN	MO	MS	MT	NC	ND	NE	NH	NJ	NM	NV
Mean	17.5	18.8	22.1	15.9	2.6	39.9	13.4	21	13.4	3.7	27.6	2.3	4.7	2.8	36.3	8.9	12
Median	6	10	10	9	1	7	7	3	6	1	11	0	0	1	9	5	3
SD	33.7	27.3	37	29	6	71.9	18.8	104	20	6.9	56.4	6.5	15.6	5	108.6	10.7	21.6
Kurtosis	44.6	40.5	26.3	159.5	48.6	10.7	6.2	603.5	9.6	14.2	186.9	220.7	433.7	24.1	213	7.4	82.2
Skewness	5.4	4.6	3.9	10.1	5.6	2.9	2.3	22.3	2.6	3.3	10.5	11.6	17.8	3.9	12.7	2.1	6.4
Minimum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum	448	362	459	549	83	566	140	2881	177	55	1172	140	399	59	2037	99	365
Total	16,679	17,877	21,035	15,199	2512	38,038	12,806	19,993	12,794	3504	26,335	2232	4455	2662	34,567	8465	11,400
Count	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953
Population	4,539,130	4,682,633	7,126,375	6,257,958	1,369,159	10,116,069	5,787,008	6,188,111	2,960,075	1,103,187	10,620,168	800,394	1,988,536	1,389,741	9,388,414	2,129,190	3,185,426
Mortality Rate	0.37%	0.38%	0.30%	0.24%	0.18%	0.38%	0.22%	0.32%	0.43%	0.32%	0.25%	0.28%	0.22%	0.19%	0.37%	0.40%	0.36%
	NY	OH	OK	OR	PA	RI	SC	SD	TN	TX	UT	VA	VT	WA	WI	WV	WY
Mean	74.4	41.4	17.5	8.8	49	3.8	18.8	3.1	28.8	92.9	5.2	22.5	0.7	14.7	15.8	7.7	2
Median	23	0	0	3	22	0	7	0	8	45	2	11	0	6	6	2	0
SD	149.7	125	66.9	14.1	69.9	12.1	30.5	8.4	90.6	103.5	7.6	38.4	1.5	24.1	25	14.5	7.7
Kurtosis	23.9	185.4	26	13.6	7.1	53.1	12.1	19.2	338.9	2.5	7.6	34.1	17.5	19.8	11	34.3	39.2
Skewness	4.4	10.9	4.9	3	2.3	6.6	3	4	15.6	1.4	2.3	4.9	3.3	3.5	2.9	4.6	5.7
Minimum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maximum	1460	2559	548	134	547	137	241	78	2174	798	65	404	16	256	206	170	81
Total	70,877	39,490	16,720	8415	46,716	3645	17,869	2993	27,487	88,578	4981	21,439	707	14,039	15,084	7291	1881
Count	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953	953
Population	20,365,879	11,852,036	4,000,953	4,318,492	13,062,764	1,106,341	5,217,037	901,165	7,023,788	29,945,493	3,373,162	8,757,467	646,545	7,901,429	5,935,064	1,781,860	579,495
Mortality Rate	0.35%	0.33%	0.42%	0.19%	0.36%	0.33%	0.34%	0.33%	0.39%	0.30%	0.15%	0.24%	0.11%	0.18%	0.25%	0.41%	0.32%

Figure 2 shows the estimated CPDAG (completed partially directed acyclic graph) for the COVID-19 daily cumulative mortality data (22 January 2020 to 1 September 2022) for 50 states and Washington, D.C. The CPDAG uniquely represents a Markov equivalence class and contains undirected and directed edges. We estimated the equivalence class of a directed acyclic graph (DAG) from the observational data using the PC algorithm (named after its inventors Peter Spirtes and Clark Glymour) found in the data using the pcalg R package [22]. We defined the independence test (partial correlations) by using the gaussCItest command in pcalg and then defined the sufficient statistics based on the correlations of our data (51 variables and n = 953 observations). We estimated the CPDAG with alpha = 0.008. Using the Rgraphviz R package, we created Figure 2, which shows the

approximately five equivalence classes of the COVID-19 daily cumulative mortality data for 50 states and Washington, D.C. The major equivalence group included Florida, Georgia, Virginia, Colorado, Michigan, New Hampshire, Ohio, Tennessee, Missouri, Montana, Kentucky, West Virginia, Wyoming, Idaho, Delaware, Maryland, Indiana, Pennsylvania, Louisiana, Mississippi, and Texas. The second equivalence group included North Carolina, Nevada, Arkansas, Nebraska, Kansas, Oklahoma, Illinois, Minnesota, and Wisconsin. The third equivalence group included Arizona, Iowa, California, Rhode Island, South Dakota, and North Dakota. The next equivalence group included Massachusetts, Washington, D.C., New Jersey, New York, and Connecticut. The states in this group are located in the northeast and New York City (NYC) was an early epicenter of the COVID-19 pandemic in the United States and approximately 203,000 cases of laboratory-confirmed COVID-19 were reported in NYC during the first 3 months of the pandemic. The crude fatality rate among confirmed cases was 9.2% overall and 32.1% among hospitalized patients according to the CDC (Centers for Disease Control and Prevention)'s Morbidity and Mortality Weekly Report <https://www.cdc.gov/mmwr/volumes/69/wr/mm6946a2.htm>, accessed on 12 September 2022. The graphical relationship between the northeast states belonging to this equivalence group and the daily COVID-19 mortality data was investigated using the vine copula method in Section 4. The last equivalence group included Alaska, Oregon, Vermont, Maine, and Hawaii. In this group, we found that although Alaska and Hawaii are isolated from the mainland US, they belonged to the same equivalence group as the mainland states Oregon, Vermont, and Maine. Even though Washington is very close to Oregon, it did not belong to any equivalence groups, as seen in Figure 2. We also found that New Mexico and Alabama did not belong to any equivalence groups, as seen in Figure 2. With the cumulative mortality data from 22 January 2020 to 1 September 2022 from 50 states plus Washington, D.C., the graphical model produced by the CPDAG seen in Figure 2 shows that the effect of geographical distance mainly influenced the grouping of the equivalence classes. However, there were some states that were not a close distance to most of the other states in the same equivalence class. Similar COVID-19 mortality data could be one reason for this. For example, AK (0.17%), HI (0.11%), ME (0.18%), OR (0.19%), and VT (0.11%) in the first equivalence group on the left in Figure 2 had similar mortality rates ranging from 0.11% to 0.19%; CT (0.31%), DC (0.21%), MA (0.30%), NJ (0.37%), and NY (0.35%) in the first equivalence group on the right in Figure 2 had similar mortality rates ranging from 0.30% to 0.37%, except for DC (0.21%). AZ (0.41%), CA (0.23%), IA (0.31%), ND (0.28%), SD (0.33%), and RI (0.33%) in the equivalence group located in the middle of Figure 2 had similar mortality rates ranging from 0.28% to 0.33%, except for the two neighboring states AZ (0.41%) and CA (0.21%). From these findings, we can conclude that the geographical location and mortality rate may be the main factors for creating the equivalence classes in the graphical model by the estimated CPDAG. However, there may be some other reasons such as each state government's budget for health and hospitals. Using the daily COVID-19 mortality data, NY, the first US state with a COVID-19 outbreak in 2020, has a unique daily COVID-19 mortality pattern by employing functional PCA, as seen in the following section.



**Figure 1.** COVID-19 daily cumulative mortality data for 50 states and Washington, D.C. (22 January 2020 to 1 September 2022).



**Figure 2.** Estimated CPDAG with COVID-19 daily cumulative mortality data from 50 states and Washington, D.C. (22 January 2020 to 1 September 2022).

### 3. Graphical Visualization by FPCA

The basic concept of FDA is to represent a function by a linear combination of basis elements. FDA and its applications are explained by [23,24]. The basic concept of FPCA decomposes density variations into a set of orthogonal principal component functions that maximize the variance along each component [2]. FPCA is defined in a separable Hilbert space of square-integrable random functions. Diverse basis functions, such as B-spline vectors, the Fourier series, or an empirical basis are used in FPCA. FPCA provides a more informative way of examining the sample covariance structure than the PCA proposed by [25]. We used the Fourier series as the basis function for FPCA in this research because the Fourier series is used for periodic or near periodic seasonal data and COVID-19 is a contagious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is affected by the weather. As with the preliminary data analysis, we employed FPCA to determine the factors (i.e., principal components) explaining the total variation in the daily COVID-19 mortality data in 50 states and Washington, D.C. Table 2 shows the percentage variations (PV) and cumulative percentage variations (CPV) in the daily COVID-19 mortality data in 50 states and Washington, D.C. The PV of the first functional principal component was 0.8194, the PV of the second functional principal component was 0.1125, the PV of the third functional principal component was 0.0512, the PV of the fourth functional principal component was 0.0117, and the PV of the fifth functional principal component was 0.0052 so the CPV of the five functional principal components was 1.000.

**Table 2.** Total Cumulative Percentage Variations in the Daily COVID-19 Mortality Data of Five Functional Principal Components for Daily COVID-19 Mortality Data.

	FPC1	FPC2	FPC3	FPC4	FPC5
PV	0.8194	0.1125	0.0512	0.0117	0.0052
CPV	0.8194	0.9320	0.9831	0.9948	1.0000

Let  $y_i(t)$  be the number of daily COVID-19 mortalities in 50 states and Washington, D.C. ( $i = 1, 2, \dots, 51$ ) in discrete time,  $t = 1, 2, \dots, 953$ .  $y_i(t)$  can be stated as  $y_i(t) = x_i(t) + e_i(t)$ , with  $x_i(t)$  denoting their underlying smooth functions and  $e_i(t)$  indicating the unobserved error components. The functional form of  $x_i(t)$  is given by the sum of the weighted basis functions,  $\phi_k(t)$ , across the set of times  $T$ .

$$x_i(t) = \sum_{k=1}^K c_{ik}\phi_k(t),$$

where  $K$  is the number of basis functions. In this study, a Fourier basis is used to represent smooth functions as the basis function due to its flexibility and computational advantages. Here, our goal is to obtain a smooth function that fits well into the observed return series,  $y_i(t)$ . We consider the following smoothing criterion:

$$SSE(y|c) = \sum_{i=1}^n \sum_{t=1}^T \left[ y_i(t) - \sum_{k=1}^K c_{ik}\phi_k(t) \right]^2 = (y - \Phi c)'(y - \Phi c),$$

where  $\Phi$  is a  $K \times T$  matrix, with  $\Phi_k = \phi_k(t)$ . We have  $K = 5$  and  $T = 953$  in this study. Therefore, the Fourier series of the functional forms is  $\phi_1(t) = 1, \phi_2(t) = \sin(\omega t), \phi_3(t) = \cos(\omega t), \phi_4(t) = \sin(2\omega t)$ , and  $\phi_5(t) = \cos(2\omega t)$ , where the parameter  $\omega = \frac{2\pi}{T}$ . For further details, see [26]. We estimate the vector of coefficients  $c$  by minimizing the smoothing criterion. In particular, we utilize the generalized cross-validation measure GCV developed by [27]:

$$GCV(\lambda) = \left( \frac{n}{n - df(\lambda)} \right) \left( \frac{SSE}{n - df(\lambda)} \right),$$

where  $df(\lambda)$  is a measure of the effective degree of freedom of the fit defined by smoothing parameter  $\lambda$ , and the best value for  $\lambda$  is the one that minimizes the criterion. In particular, we obtain the smoothing parameter of  $\lambda = 10^{10.2}$  by using our sample data. Given the estimates  $\hat{c}$ , we are able to obtain the smoothed return series  $\hat{y} = \Phi \hat{c}$ .

After having  $\hat{y}_i(t)$ , the next step is to seek a set of orthogonal functions,  $\psi_j(t)$  such that

$$\begin{aligned} \langle \psi_j(t), \psi_k(t) \rangle &= \int \psi_j(t)\psi_k(t)dt = 0, \text{ for all } j \neq k, \text{ and} \\ \|\psi_j(t)\|^2 &= \langle \psi_j(t), \psi_k(t) \rangle = 1 \text{ for all } j. \end{aligned}$$

For example,  $\psi_1(t)$  can be achieved by maximizing the following objective function:

$$\sum_i (\langle \hat{y}_i(t), \psi_1(t) \rangle)^2 = \sum_i \left( \int \hat{y}_i(t)\psi_1(t)dt \right)^2,$$

subject to the constraint  $\|\psi_1(t)\|^2 = 1$ . Note that the function  $\psi_1(t)$  is the first principal component.

Figures 3 and 4 show a two-dimensional (2D) plot with two main functional principal components and a 3D plot with three main functional principal components from the FPCA. The 3D plot clearly shows that Texas and Florida had similar time course patterns for daily COVID-19 mortalities, whereas California and New York had their own time course patterns for daily COVID-19 mortalities. We can make an inference that the location and population size of a state are related to the number of COVID-19 mortalities in the US because California is located on the west coast, New York is an east-coast state, Texas

and Florida are located in the south, and all of these states have high population sizes. However, the rest of the states other than California, Florida, New York, and Texas had similar time course patterns for daily COVID-19 mortalities. This is an interesting finding in this research.

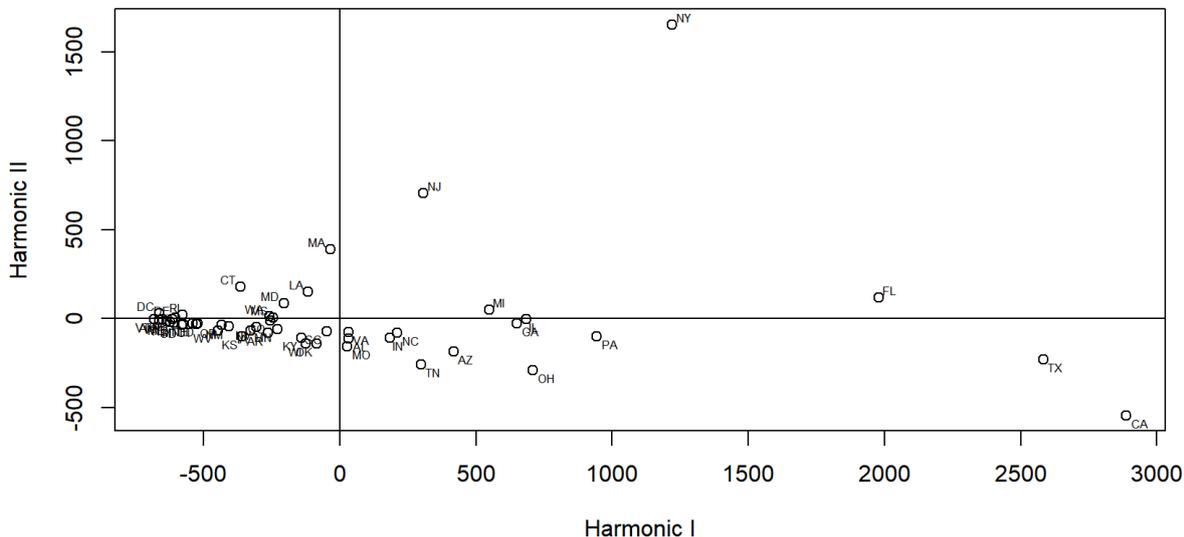


Figure 3. 2D Plot of FPCA for Daily COVID-19 Mortality Data.

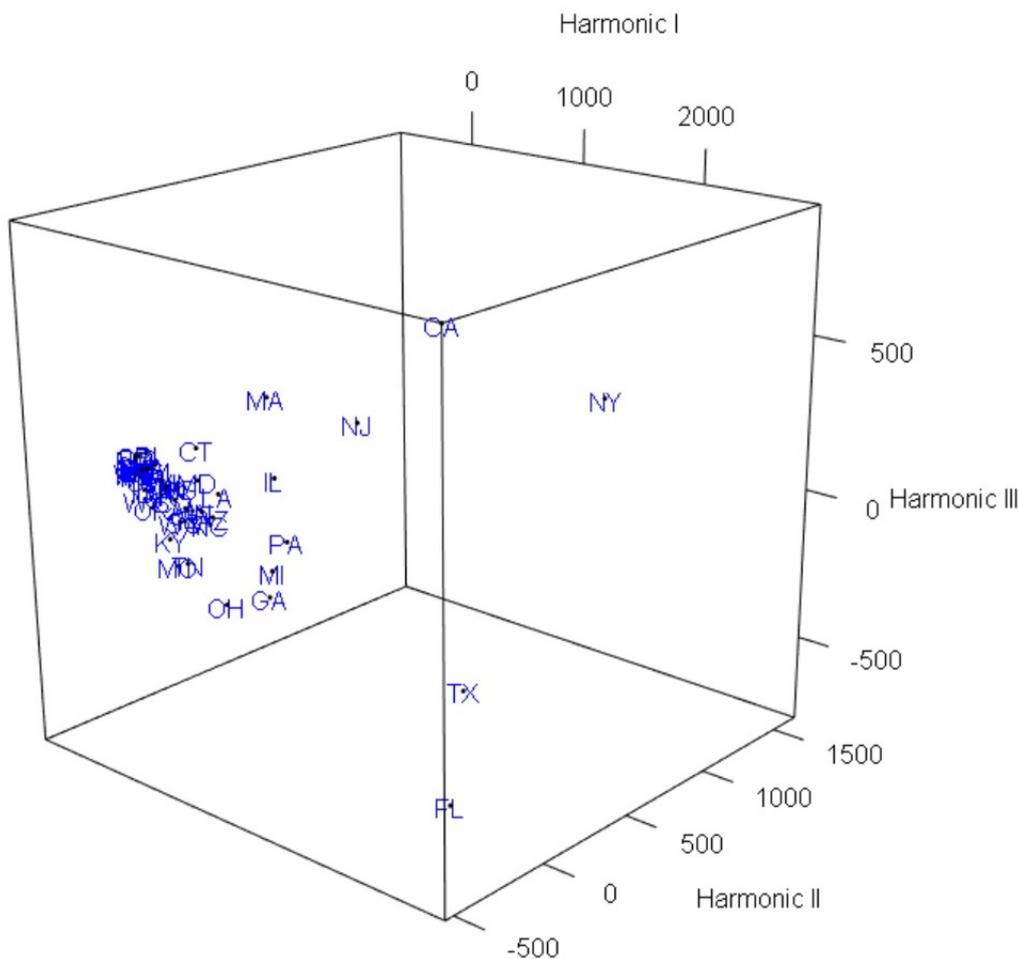
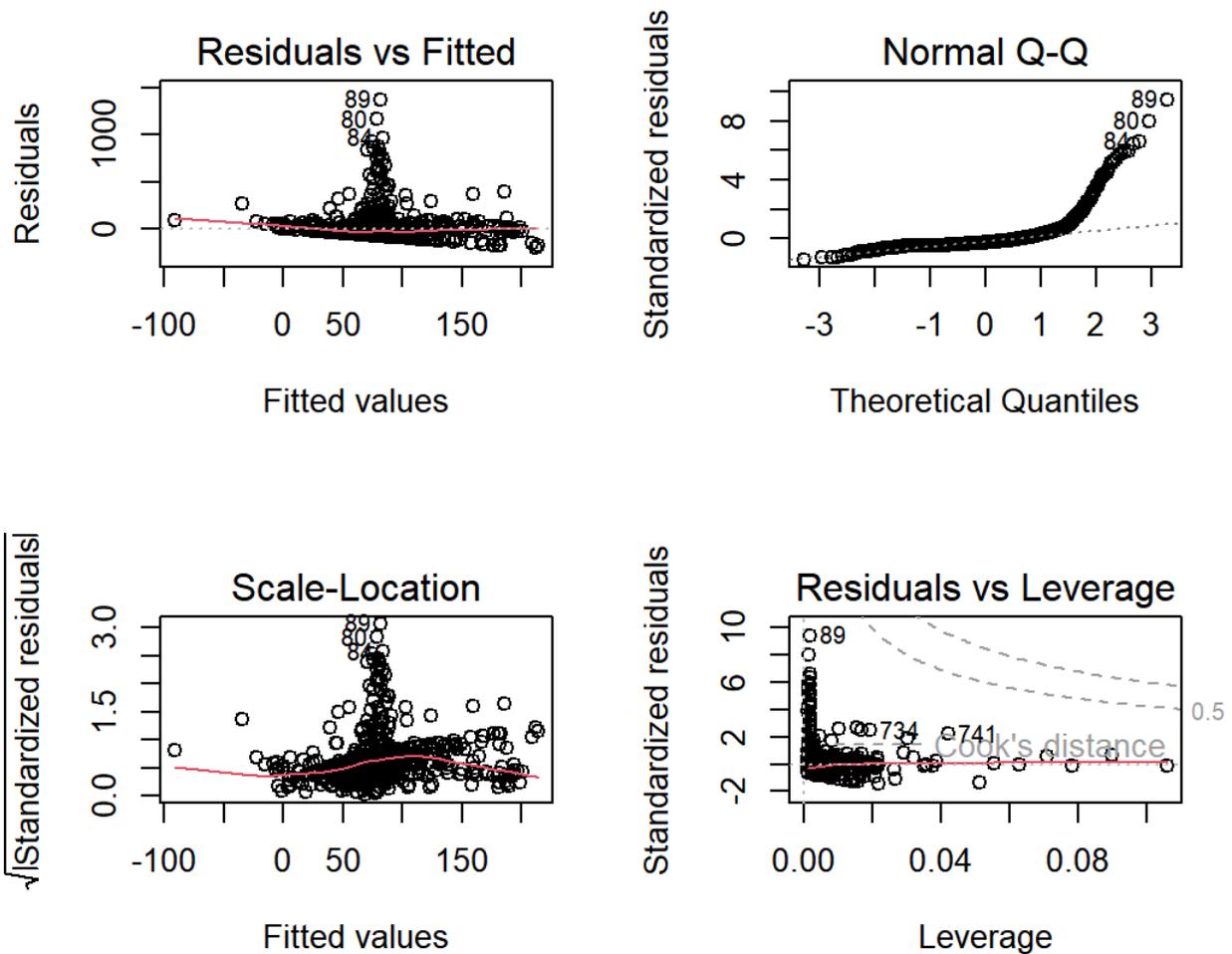


Figure 4. 3D Plot of FPCA for Daily COVID-19 Mortality Data.

#### 4. Copula Methods

The traditional linear regression model cannot be used for COVID-19 data analysis because of the violations of linear regression assumptions such as the normality of residuals and the homogeneity of the residuals' variances. To verify the assumption violation for the linear regression with the COVID-19 daily mortality data, we performed a linear regression for NY with the CA, TX, and FL COVID-19 daily mortality data (22 January 2020 to 1 September 2022). In Figure 5, we can see that the residual errors of the linear regression for NY with the CA, TX, and FL COVID-19 daily mortality data did not follow a normal distribution. The homogeneity of the variance assumptions can be checked by examining the scale-location plot. It can be seen that the variances in the residual points fluctuated with the values of the fitted outcome variables, suggesting non-constant variances in the residual errors. We also computed the Breusch–Pagan score of the hypothesis of the constant error variance against the alternative that the error variance changes with the level of the response. The Chi-square test statistic was 28.78 and the  $p$ -value was 0.000. The test confirmed that the linear regression had non-constant variances in the residual errors. There existed outliers and high leverage points in the linear regression, as shown in Figure 5. From this linear regression with the COVID-19 mortality data, we can say that the linear regression assumptions were violated. To rectify the difficulties, we need to use the copula method on the COVID-19 mortality data.



**Figure 5.** Plots of linear regression for NY with CA, TX, FL cumulative daily COVID-19 mortality data (22 January 2020 to 1 September 2022).

#### 4.1. Graphical Visualization Using Copula

A copula is a multivariate distribution function defined by the unit  $[0, 1]^2$  with uniformly distributed marginals and describes the dependence mechanism between two random variables by eliminating the influence of the marginals or any monotone transformations of the marginals [3,4,11]. A bivariate distribution function,  $F(y_1, y_2)$ , can be represented as a function of its marginal distribution of  $Y_1$  and  $Y_2$ ,  $F(y_1)$  and  $F(y_2)$ , by using a two-dimensional copula  $C(\cdot, \cdot)$ . More specifically, the copula may be written as

$$F(y_1, y_2) = C(F(y_1), F(y_2)) = C(u, v),$$

where  $u$  and  $v$  are the continuous empirical marginal distribution functions  $F(y_1)$  and  $F(y_2)$ , respectively. Note that  $u$  and  $v$  have a uniform distribution  $U(0, 1)$ .

Our study employs the Gaussian copula regression method to investigate the relationship between the state with the highest mortality and the rest of the states in the US. Let  $F(\cdot|x_i)$  be the marginal cumulative distribution for  $x_i$ , then, the joint cumulative distribution function in the Gaussian copula regression can be expressed as

$$\Pr(Y_1 \leq y_1, \dots, Y_n \leq y_n) = \Phi_n(\epsilon_1, \dots, \epsilon_n; \mathbf{P}),$$

where  $\epsilon_i$  indicates a stochastic error that follows a multivariate standard normal distribution with a correlation matrix  $\mathbf{P}$ . See [12] for more details.

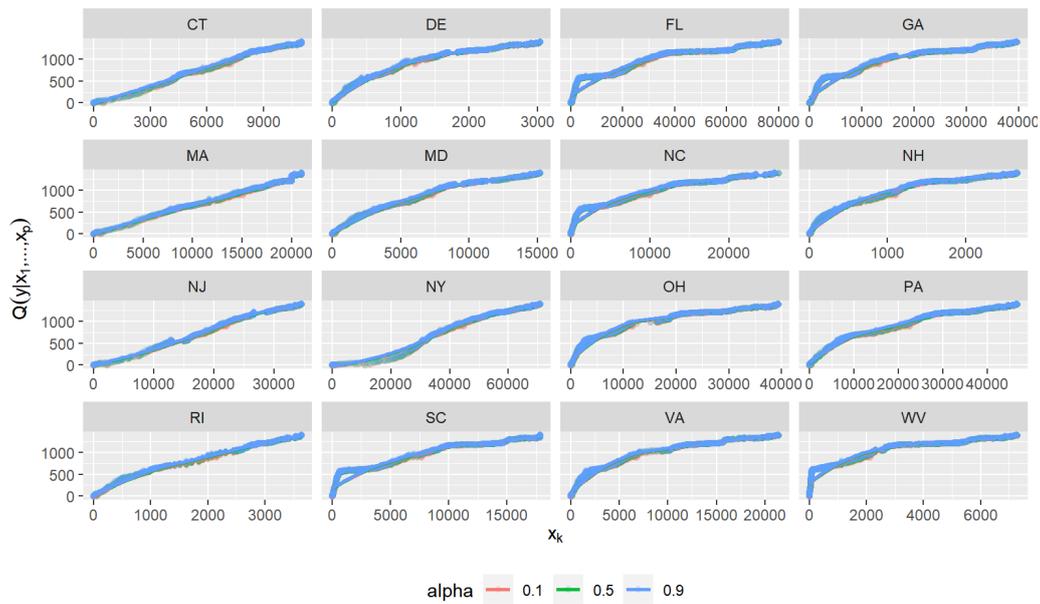
Vine copulas were proposed by [13] to explain a multivariate dependence structure using copula due to the difficulty of expressing a multivariate joint distribution by copula [14–16]. Vine copulas are a graphical model that represent a  $d$ -dimensional multivariate density based on a pair-copula method given by [16] as follows:

$$f(\mathbf{y}; \boldsymbol{\phi}) = \prod_{k=1}^d f_k(y_k) \times \prod_{i=1}^{d-1} \prod_{j=1}^{d-i} c_{j, j+i|(j+1):(j+i-1)} \left( F(y_j | y_{j+1}, \dots, y_{j+i-1}), F(y_{j+i} | y_{j+1}, \dots, y_{j+i-1}); \boldsymbol{\beta}_{j, j+i|(j+1):(j+i-1)} \right),$$

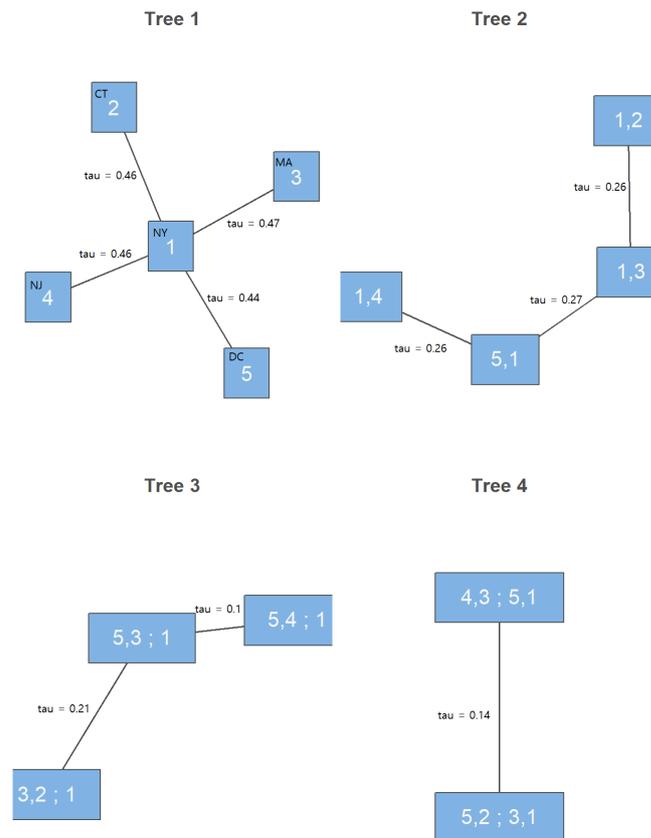
where  $f_k(x_k)$  are the marginal densities,  $c_{j, j+i|(j+1):(j+i-1)}$  are the bivariate copula densities with parameter(s)  $\boldsymbol{\beta}_{j, j+i|(j+1):(j+i-1)}$ , and  $\boldsymbol{\phi}$  is the set of all parameters in the D-vine density.

Figure 6 shows the marginal effects of a D-vine quantile regression model (10%, 50%, 90%) for the target variable DC with east coast states (CT, DE, FL, GA, MA, MD, NC, NH, NJ, NY, OH, PA, RI, SC, VA, WV) on the COVID-19 daily cumulative mortality data from 22 January 2020 to 1 September 2022. By using D-vine-based quantile regression with a selected copula out of all copula parametric and nonparametric families [18], we showed the linear and increasing marginal effects of each east coast state on Washington, D.C., as seen in Figure 6.

Figure 7 shows the R-vine copula [28]-based hierarchical tree dependence structure of five states' (NY, CT, MA, NJ, DC) daily COVID-19 mortalities from 22 January 2020 to 1 September 2022 using the RVineStructureSelect command in the VineCopula R package. New York is located in the center among five east coast states and had relatively high Kendall's tau correlations with CT (0.46), MA (0.47), NJ (0.46), and DC (0.44) in the level-one tree. This reminds us that New York City (NYC) was the early epicenter of the COVID-19 pandemic in the United States and affected neighboring states' COVID-19 mortality numbers. In the level-two tree, NY had a 0.26 Kendall's tau correlation with NJ and DC, a 0.27 Kendall's tau correlation with DC and MA, and a 0.26 Kendall's tau correlation with MA and CT. In the level-three tree, NY and MA had a 0.21 Kendall's tau correlation with DC and CT and a 0.1 Kendall's tau correlation with MA and NJ. In the level-four tree, CT and NJ had 0.14 Kendall's tau correlation with NY, MA, and DC. We can see the conditional dependence relationships among five east coast states in Figure 7.



**Figure 6.** Plots of marginal effects of a D-vine quantile regression model (10%, 50%, 90%) for DC on the COVID-19 daily cumulative mortality data for CT, DE, FL, GA, MA, MD, NC, NH, NJ, NY, OH, PA, RI, SC, VA, WV (22 January 2020 to 1 September 2022).



**Figure 7.** R-vine copula model-based hierarchical tree structure plots of daily COVID-19 mortality data (22 January 2020 to 1 September 2022).

#### 4.2. Gaussian Copula Marginal Regression

Since a traditional multiple linear regression is not appropriate for a non-normal and dispersed COVID-19 mortality data analysis, we applied the daily COVID-19 mortality data to the Gaussian copula marginal regression (GCMR) model using the *gcmr* R package. For our data analysis, GCMR models enable one to specify the correlation matrix of the errors. For this study, the correlation matrices of the autoregressive moving average (ARMA)(0,0), ARMA(0, 1), ARMA(1, 0), and ARMA(1, 1) were considered. To select the best GCMR model for the correlation matrix, four different GCMR models were compared with the Akaike information criterion (AIC). Before applying the GCMR models to the COVID-19 daily mortality data, we performed a stationary test using the augmented Dickey–Fuller statistic. Table 3 shows that the mortality data from the four big states (CA, TX, FL, NY) were stationary. We applied the GCMR models with a correlation matrix of ARMA(p,q), where  $p = 0, 1$  and  $q = 0, 1$ . Table 4 shows that the GCMR model that best fit the daily COVID-19 mortality data of CA was the GCMR model with a correlation matrix of ARMA(0,0), whereas the best fitting model for the daily COVID-19 mortality data of TX, FL, and NY was the GCMR model with a correlation matrix of ARMA(1,1). For the mortality data of CA, we found the following positively statistically significant states (AK 0.105, FL 0.249, GA 0.090, ID 0.070, NE 0.057, NY 0.072, TX 0.430, VT 0.071, WI 0.201) and the following negatively statistically significant states (HI −0.099, KY −0.091, MO −0.067, MT −0.083, OR −0.115, TN −0.108). Even though NY is far away from CA, there was a positive statistical effect on the mortality data of CA of 0.072. TX had the largest statistical effect on the mortality data of CA of 0.43. For the mortality data of TX, we found the following positively statistically significant states (AK 0.053, CA 0.273, CO 0.176, DC 0.050, FL 0.092, HI 0.051, MA 0.074, OH 0.077, WA 0.054) and the following negatively statistically significant states (AR −0.067, GA −0.067, IA −0.050, ID −0.049, MN −0.085). CA had the largest statistical effect on the mortality data of TX of 0.273. For the mortality data of FL, we found the following positively statistically significant states (CA 0.391, DC 0.069, HI 0.106, LA 0.078, NC 0.130, OH 0.225, TX 0.188) and the following negatively statistically significant states (AZ −0.075, NY −0.089). It is interesting to see that NY had a negative statistically significant effect on the mortality data of FL. CA had the largest statistical effect on the mortality data of FL of 0.391. For the mortality data of NY, we found the following positively statistically significant states (AR 0.169, CA 0.063, DC 0.069, HI 0.062, IN 0.094, MA 0.116, NM 0.130, UT 0.094, WV 0.139, WY 0.073) and the following negatively statistically significant states (GA −0.095, MT −0.086, PA −0.079). PA had a negative statistically significant effect on the mortality data of NY despite PA being a neighboring state of NY.

**Table 3.** Stationary Test Using Augmented Dickey–Fuller Statistic.

	CA	TX	FL	NY
Dickey–Fuller Statistic	−18.02	−12.118	−21.365	−15.552
<i>p</i> -value	0.01	0.01	0.01	0.01
Stationary	Yes	Yes	Yes	Yes

Table 4. GCMR results.

CA	Estimate	Std. Error	z value	p-Value	TX	Estimate	Std. Error	z Value	p-Value	FL	Estimate	Std. Error	z Value	p-Value	NY	Estimate	Std. Error	z Value	p-Value
Intercept	0.007	0.036	0.210	0.834	Intercept	0.334	0.077	4.355	0.000	Intercept	-0.019	0.065	-0.292	0.770	Intercept	0.030	0.055	0.542	0.588
AK	0.105	0.028	3.813	0.000	AK	0.053	0.021	2.505	0.012	AK	0.000	0.032	0.014	0.989	AK	-0.032	0.024	-1.359	0.174
AL	-0.026	0.030	-0.865	0.387	AL	0.024	0.023	1.031	0.303	AL	0.009	0.034	0.262	0.793	AL	-0.036	0.026	-1.424	0.154
AR	-0.012	0.035	-0.355	0.723	AR	-0.067	0.029	-2.338	0.019	AR	-0.071	0.042	-1.683	0.092	AR	0.169	0.032	5.316	0.000
AZ	0.054	0.032	1.686	0.092	AZ	0.022	0.025	0.866	0.386	AZ	-0.075	0.037	-2.013	0.044	AZ	0.041	0.028	1.452	0.146
CO	-0.008	0.024	-0.317	0.752	CA	0.273	0.025	11.032	0.000	CA	0.391	0.037	10.621	0.000	CA	0.063	0.029	2.161	0.031
CT	-0.034	0.034	-0.999	0.318	CO	0.176	0.019	9.321	0.000	CO	0.016	0.029	0.549	0.583	CO	-0.023	0.022	-1.041	0.298
DC	0.001	0.029	0.022	0.982	CT	-0.004	0.025	-0.158	0.874	CT	0.026	0.038	0.686	0.493	CT	0.011	0.028	0.377	0.706
DE	0.063	0.027	2.383	0.017	DC	0.050	0.022	2.265	0.024	DC	0.069	0.033	2.088	0.037	DC	0.069	0.025	2.764	0.006
FL	0.249	0.026	9.718	0.000	DE	-0.017	0.020	-0.879	0.380	DE	-0.045	0.029	-1.534	0.125	DE	0.043	0.022	1.951	0.051
GA	0.090	0.034	2.655	0.008	FL	0.092	0.022	4.260	0.000	GA	-0.012	0.039	-0.318	0.751	FL	-0.047	0.024	-1.939	0.053
HI	-0.099	0.031	-3.213	0.001	GA	-0.067	0.026	-2.566	0.010	HI	0.106	0.035	3.023	0.003	GA	-0.095	0.029	-3.271	0.001
IA	0.030	0.028	1.042	0.297	HI	0.051	0.024	2.174	0.030	IA	0.018	0.033	0.544	0.586	HI	0.062	0.026	2.353	0.019
ID	0.070	0.030	2.314	0.021	IA	-0.050	0.022	-2.290	0.022	ID	-0.019	0.034	-0.551	0.581	IA	-0.004	0.024	-0.184	0.854
IL	-0.015	0.041	-0.361	0.718	ID	-0.049	0.023	-2.156	0.031	IL	0.052	0.049	1.059	0.289	ID	-0.026	0.026	-1.014	0.310
IN	0.069	0.040	1.698	0.089	IL	-0.022	0.033	-0.679	0.497	IN	0.013	0.047	0.280	0.780	IL	0.011	0.036	0.303	0.762
KS	0.003	0.028	0.113	0.910	IN	-0.006	0.032	-0.199	0.842	KS	0.010	0.031	0.316	0.752	IN	0.094	0.035	2.653	0.008
KY	-0.091	0.033	-2.709	0.007	KS	-0.018	0.020	-0.881	0.378	KY	0.041	0.040	1.014	0.311	KS	0.019	0.023	0.831	0.406
LA	-0.001	0.030	-0.039	0.969	KY	0.020	0.027	0.752	0.452	LA	0.078	0.037	2.105	0.035	KY	0.052	0.030	1.746	0.081
MA	0.014	0.031	0.462	0.644	LA	0.000	0.025	-0.014	0.989	MA	-0.067	0.042	-1.592	0.111	LA	0.038	0.028	1.352	0.176
MD	0.057	0.029	1.952	0.051	MA	0.074	0.029	2.574	0.010	MD	0.062	0.036	1.730	0.084	MA	0.116	0.032	3.596	0.000
ME	-0.029	0.025	-1.137	0.255	MD	0.046	0.024	1.914	0.056	ME	-0.013	0.029	-0.452	0.651	MD	0.015	0.027	0.549	0.583
MI	0.025	0.030	0.816	0.415	MI	-0.028	0.020	-1.417	0.157	MI	0.059	0.034	1.724	0.085	MI	-0.050	0.022	-2.285	0.022
MN	-0.036	0.036	-0.987	0.324	MN	-0.026	0.023	-1.139	0.255	MN	0.081	0.043	1.890	0.059	MN	-0.017	0.025	-0.666	0.506
MO	-0.067	0.030	-2.203	0.028	MO	-0.085	0.029	-2.958	0.003	MO	0.058	0.034	1.709	0.087	MO	0.040	0.032	1.232	0.218
MS	0.025	0.035	0.720	0.472	MS	-0.010	0.023	-0.432	0.666	MS	-0.024	0.040	-0.602	0.547	MS	-0.005	0.026	-0.188	0.851
MT	-0.083	0.032	-2.545	0.011	MT	-0.012	0.027	-0.440	0.660	MT	0.049	0.037	1.319	0.187	MT	-0.024	0.030	-0.811	0.417
NC	-0.002	0.036	-0.054	0.957	MT	-0.003	0.025	-0.111	0.911	NC	0.130	0.041	3.144	0.002	NC	-0.086	0.028	-3.065	0.002
ND	-0.031	0.027	-1.173	0.241	NC	0.011	0.028	0.394	0.694	ND	-0.028	0.031	-0.895	0.371	ND	0.042	0.031	1.342	0.180
NE	0.057	0.027	2.102	0.036	ND	-0.005	0.021	-0.225	0.822	NE	-0.058	0.031	-1.902	0.057	NE	-0.032	0.023	-1.355	0.175
NH	0.016	0.028	0.574	0.566	NE	-0.018	0.021	-0.890	0.374	NH	-0.001	0.032	-0.041	0.967	NH	-0.023	0.023	-1.015	0.310
NJ	0.041	0.030	1.363	0.173	NH	-0.006	0.021	-0.274	0.784	NJ	0.017	0.044	0.393	0.695	NH	0.009	0.024	0.366	0.714
NM	0.052	0.035	1.459	0.145	NJ	0.014	0.031	0.444	0.657	NM	-0.059	0.044	-1.355	0.175	NJ	0.061	0.035	1.770	0.077
NV	0.014	0.033	0.428	0.668	NM	0.045	0.030	1.520	0.128	NV	-0.002	0.036	-0.062	0.950	NM	0.130	0.033	3.962	0.000
NY	0.072	0.035	2.065	0.039	NV	-0.036	0.024	-1.511	0.131	NY	-0.089	0.043	-2.089	0.037	NV	-0.034	0.027	-1.275	0.202
OH	-0.038	0.029	-1.279	0.201	NY	-0.038	0.029	-1.305	0.192	OH	0.225	0.033	6.799	0.000	OH	0.001	0.025	0.057	0.954
OK	0.006	0.036	0.165	0.869	OH	0.077	0.023	3.435	0.001	OK	-0.045	0.042	-1.075	0.282	OK	0.014	0.031	0.455	0.649
OR	-0.115	0.033	-3.454	0.001	OR	-0.055	0.028	-1.996	0.046	OR	0.034	0.038	0.890	0.374	OR	0.025	0.028	0.882	0.378
PA	-0.019	0.031	-0.603	0.546	PA	-0.016	0.025	-0.650	0.515	PA	-0.035	0.037	-0.959	0.338	PA	-0.079	0.027	-2.913	0.004
RI	0.007	0.027	0.256	0.798	RI	-0.021	0.025	-0.872	0.383	RI	-0.015	0.032	-0.469	0.639	RI	0.027	0.024	1.131	0.258
SC	0.047	0.034	1.384	0.166	RI	-0.015	0.021	-0.715	0.475	SC	0.059	0.038	1.543	0.123	SC	0.005	0.029	0.184	0.854
SD	-0.022	0.030	-0.740	0.459	SC	-0.031	0.026	-1.221	0.222	SD	0.017	0.034	0.490	0.624	SD	0.007	0.025	0.260	0.795
TN	-0.108	0.029	-3.684	0.000	SD	-0.043	0.022	-1.922	0.055	TN	-0.033	0.035	-0.925	0.355	TN	-0.005	0.026	-0.194	0.846
TX	0.430	0.030	14.238	0.000	TN	0.016	0.023	0.667	0.505	TX	0.188	0.046	4.107	0.000	TX	-0.053	0.036	-1.494	0.135
UT	-0.023	0.032	-0.726	0.468	UT	-0.011	0.024	-0.463	0.644	UT	-0.018	0.036	-0.504	0.615	UT	0.094	0.027	3.513	0.000
VA	0.038	0.029	1.284	0.199	VA	0.002	0.024	0.068	0.946	VA	0.045	0.035	1.292	0.196	VA	0.092	0.026	3.493	0.000
VT	0.071	0.027	2.691	0.007	VT	0.005	0.021	0.230	0.818	VT	-0.042	0.031	-1.342	0.179	VT	-0.003	0.023	-0.133	0.894
WA	-0.028	0.027	-1.051	0.293	WA	0.054	0.020	2.676	0.007	WA	0.047	0.030	1.554	0.120	WA	0.042	0.023	1.839	0.066
WI	0.201	0.027	7.466	0.000	WI	0.030	0.022	1.386	0.166	WI	-0.003	0.033	-0.105	0.917	WI	0.010	0.024	0.406	0.685
WV	-0.005	0.031	-0.165	0.869	WV	-0.019	0.025	-0.784	0.433	WV	-0.062	0.037	-1.697	0.090	WV	0.139	0.027	5.072	0.000
WY	-0.031	0.035	-0.877	0.381	WY	0.000	0.025	0.001	0.999	WY	-0.030	0.038	-0.795	0.427	WY	0.073	0.028	2.582	0.010
sigma	0.164	0.004	43.667	0.000	sigma	0.202	0.019	10.832	0.000	sigma	0.214	0.014	15.139	0.000	sigma	0.182	0.016	11.441	0.000
					AR <sub>1</sub>	0.989	0.003	317.600	0.000	AR <sub>1</sub>	0.983	0.007	137.510	0.000	AR <sub>1</sub>	0.979	0.008	124.010	0.000
					MA <sub>1</sub>	-0.806	0.018	-45.970	0.000	MA <sub>1</sub>	-0.875	0.016	-53.210	0.000	MA <sub>1</sub>	-0.807	0.023	-34.950	0.000

### 4.3. Copula Dynamic Conditional Correlation

The copula dynamic conditional correlation (Copula-DCC) is an extension of the DCC model. The time-varying conditional correlation in the copula framework was developed by [29].

Let  $\mathbf{r}_t = (r_{1t}, \dots, r_{nt})$  be an  $n \times 1$  vector of the daily COVID-19 mortality data and it follows a copula GARCH model with joint distribution given by

$$F(\mathbf{r}_t | \mu_t, \mathbf{h}_t) = C(F_1(r_{1t} | \mu_{1t}, h_{1t}), \dots, F_n(r_{nt} | \mu_{nt}, h_{nt})) \tag{1}$$

where  $F_i$  and  $C$  are the conditional distribution and the copula function, respectively.

It is formulated that the conditional mean  $E[r_{it}|\mathfrak{S}_{t-1}] = \mu_{it}$  is a linear function of its one-lag past the mortality data and follows an ARMA(1,1) process. The conditional variance  $h_{it}$  is assumed to follow a gjr-GARCH(1,1) process. One can thus consider

$$r_{it} = \mu_{it} + \theta_1(r_{it-1} - \mu_{it}) + \theta_2\varepsilon_{it-1}^2 + \varepsilon_{it}, \quad \varepsilon_{it} = \sqrt{h_{it}}z_{it} \quad (2)$$

$$h_{it}^2 = \omega + \alpha_1\varepsilon_{it-1}^2 + \gamma_1 I_{it-1}\varepsilon_{it-1}^2 + \beta_1 h_{it-1} \quad (3)$$

where  $h_{it}^2$  is the conditional variance and  $\omega$  is the intercept. Further,  $\beta_1$  and  $\alpha_1$  are the ARCH and GARCH terms,  $\gamma_1$  denotes the leverage term, the indicator function  $I$  takes values one and zero according to the "bad" news (negative shock,  $\varepsilon_{it-1} < 0$ ) and the "good" news (positive shock,  $\varepsilon_{it-1} \geq 0$ ), respectively. Note that the coefficients  $\alpha_1 + \gamma_1$  and  $\alpha_1$  correspond, respectively, to the "bad" news and "good" news. When  $\gamma_1 > 0$ , the negative shock produces a greater response than the positive shock. Here,  $z_{it}$  are i.i.d. random variables, which follow Johnson's reparametrized SU distribution, viz.,  $z_{it} \sim JSU(\mu, \sigma, \nu, \tau)$  in [30], where the four parameters  $(\mu, \sigma, \nu, \tau)$  are the mean, standard deviation, skew, and shape parameters, respectively. The dependence structure is modeled using elliptical copulas with conditional correlation  $\mathbf{R}_t$  and constant shape parameter  $\tau$ . The conditional density with a Gaussian copula is given by (see, for instance, reference [4])

$$c_t(u_{it}, \dots, u_{nt} | \mathbf{R}_t) = \frac{f_t(F_i^{-1}(u_{it}), \dots, F_i^{-1}(u_{nt}) | \mathbf{R}_t)}{\prod_{i=1}^n f_i(F_i^{-1}(u_{it}))} \quad (4)$$

where  $u_{it} = F_{it}(r_{it} | \mu_{it}, h_{it}, \nu_i, \tau_i)$  is the probability integral transformed values by  $F_{it}$ , which can be obtained using the gjr GARCH process, and  $F_i^{-1}(u_{it} | \tau)$  represents the quantile transformation. The Gaussian copula conditional correlation can be obtained using the function `cgarchspec` command in the R package `rmgarch`.

Table 5 shows that the mean equation error structure followed ARMA(1,1) for the daily COVID-19 mortality data of CA, TX, FL, and NY, and the coefficients on the AR<sub>1</sub> and MA<sub>1</sub> terms were statistically significant in CA, TX, FL, and NY. The estimates of  $\gamma_1$ s in the variance equations in CA and NY were statistically significant at the 1% significance level. This means that the negative shock produced a greater response than the positive shock in terms of the daily COVID-19 mortality data in CA and NY. The estimates of  $\alpha_1$ s in the variance equations in TX and FL were statistically significant at the 1% significance level. This means that the positive shock produced a greater response than the negative shock in terms of the daily COVID-19 mortality data in TX and FL.

Figure 8 shows the two-state plots of the copula-DCC model for CA, TX, FL, and NY. In Figure 8, the estimated time-varying conditional correlations of the daily COVID-19 mortality data for CA and FL had increased over the previous two years and the estimated time-varying conditional correlations of the daily COVID-19 mortality data for CA and TX and TX and FL had decreased rapidly over the 2022 summer period by 1 September 2022. Another interesting finding from Figure 8 is that the time-varying correlation of the daily COVID-19 mortality data for CA and NY over the given period varied between  $-0.10$  and  $0.10$ . The geographical distance between CA (west coast) and NY (east coast) is about 2900 miles, which can make the time-varying correlations of the daily COVID-19 mortality data for CA and NY become smaller.

Table 5. Copula-DCC Models.

CA and TX					TX and FL														
CA	Estimate	Std. Error	z value	p-value	TX	Estimate	Std. Error	z value	p-value	TX	Estimate	Std. Error	z value	p-value	FL	Estimate	Std. Error	z value	p-value
$\mu$	0.10	0.02	6.35	0.00	$\mu$	0.08	0.00	34.66	0.00	$\mu$	0.08	0.00	34.73	0.00	$\mu$	0.10	0.01	9.47	0.00
AR <sub>1</sub>	1.00	0.00	816.06	0.00	AR <sub>1</sub>	1.00	0.01	157.53	0.00	AR <sub>1</sub>	1.00	0.01	157.68	0.00	AR <sub>1</sub>	1.00	0.02	55.00	0.00
MA <sub>1</sub>	-0.64	0.02	-26.30	0.00	MA <sub>1</sub>	-0.50	0.11	-4.67	0.00	MA <sub>1</sub>	-0.50	0.11	-4.67	0.00	MA <sub>1</sub>	-0.78	0.23	-3.40	0.00
$\omega$	0.00	0.00	1.04	0.30	$\omega$	0.00	0.00	0.00	1.00	$\omega$	0.00	0.00	0.00	1.00	$\omega$	0.00	0.00	0.00	1.00
$\alpha_1$	0.13	0.03	4.72	0.00	$\alpha_1$	0.46	0.23	2.01	0.04	$\alpha_1$	0.46	0.23	2.01	0.04	$\alpha_1$	0.45	0.17	2.68	0.01
$\beta_1$	0.83	0.03	28.65	0.00	$\beta_1$	0.52	0.15	3.58	0.00	$\beta_1$	0.52	0.15	3.58	0.00	$\beta_1$	0.47	0.12	3.88	0.00
$\gamma_1$	0.09	0.04	2.24	0.03	$\gamma_1$	0.03	0.19	0.17	0.87	$\gamma_1$	0.03	0.19	0.17	0.87	$\gamma_1$	0.16	0.16	1.02	0.31
skew	1.55	0.54	2.86	0.00	skew	-0.05	0.08	-0.66	0.51	skew	-0.05	0.08	-0.66	0.51	skew	0.20	0.43	0.47	0.64
shape	3.01	0.62	4.86	0.00	shape	1.24	0.17	7.32	0.00	shape	1.24	0.17	7.32	0.00	shape	1.24	0.26	4.82	0.00
[joint]dcca1	0.04	0.01	2.67	0.01	Log-Likelihood	2215.84	AIC	-4.61		[joint]dcca1	0.02	0.02	0.87	0.38	Log-Likelihood	1549.04	AIC	-3.21	
[joint]dccb1	0.94	0.02	43.26	0.00	N	953	BIC	-4.51		[joint]dccb1	0.96	0.06	16.16	0.00	N	953	BIC	-3.11	
CA and FL					TX and NY														
CA	Estimate	Std. Error	z value	p-value	FL	Estimate	Std. Error	z value	p-value	TX	Estimate	Std. Error	z value	p-value	NY	Estimate	Std. Error	z value	p-value
$\mu$	0.10	0.02	6.35	0.00	$\mu$	0.10	0.01	9.47	0.00	$\mu$	0.08	0.00	34.69	0.00	$\mu$	0.12	0.00	679.52	0.00
AR <sub>1</sub>	1.00	0.00	815.72	0.00	AR <sub>1</sub>	1.00	0.02	54.94	0.00	AR <sub>1</sub>	1.00	0.01	157.55	0.00	AR <sub>1</sub>	1.00	0.00	367.85	0.00
MA <sub>1</sub>	-0.64	0.02	-26.34	0.00	MA <sub>1</sub>	-0.78	0.23	-3.40	0.00	MA <sub>1</sub>	-0.50	0.11	-4.67	0.00	MA <sub>1</sub>	-0.73	0.08	-9.13	0.00
$\omega$	0.00	0.00	1.04	0.30	$\omega$	0.00	0.00	0.00	1.00	$\omega$	0.00	0.00	0.00	1.00	$\omega$	0.00	0.00	0.00	1.00
$\alpha_1$	0.13	0.03	4.73	0.00	$\alpha_1$	0.45	0.17	2.68	0.01	$\alpha_1$	0.46	0.23	2.01	0.04	$\alpha_1$	0.35	0.14	2.48	0.01
$\beta_1$	0.83	0.03	28.71	0.00	$\beta_1$	0.47	0.12	3.89	0.00	$\beta_1$	0.52	0.15	3.58	0.00	$\beta_1$	0.49	0.07	6.69	0.00
$\gamma_1$	0.09	0.04	2.24	0.03	$\gamma_1$	0.16	0.16	1.02	0.31	$\gamma_1$	0.03	0.19	0.17	0.87	$\gamma_1$	0.32	0.17	1.94	0.05
skew	1.55	0.54	2.86	0.00	skew	0.20	0.43	0.47	0.64	skew	-0.05	0.08	-0.66	0.51	skew	-0.37	0.19	-1.93	0.05
shape	3.01	0.62	4.86	0.00	shape	1.24	0.26	4.83	0.00	shape	1.24	0.17	7.32	0.00	shape	1.15	0.18	6.55	0.00
[joint]dcca1	0.01	0.00	3.00	0.00	Log-Likelihood	1293.03	AIC	-2.67		[joint]dcca1	0.03	0.01	2.90	0.00	Log-Likelihood	1814.15	AIC	-3.77	
[joint]dccb1	0.99	0.00	253.89	0.00	N	953	BIC	-2.57		[joint]dccb1	0.92	0.03	30.72	0.00	N	953	BIC	-3.66	
CA and NY					FL and NY														
CA	Estimate	Std. Error	z value	p-value	NY	Estimate	Std. Error	z value	p-value	FL	Estimate	Std. Error	z value	p-value	NY	Estimate	Std. Error	z value	p-value
$\mu$	0.10	0.02	6.35	0.00	$\mu$	0.11	0.00	678.78	0.00	$\mu$	0.10	0.01	9.46	0.00	$\mu$	0.12	0.00	679.31	0.00
AR <sub>1</sub>	1.00	0.00	816.34	0.00	AR <sub>1</sub>	1.00	0.00	366.63	0.00	AR <sub>1</sub>	1.00	0.02	54.91	0.00	AR <sub>1</sub>	1.00	0.00	367.74	0.00
MA <sub>1</sub>	-0.64	0.03	-26.19	0.00	MA <sub>1</sub>	-0.73	0.08	-9.11	0.00	MA <sub>1</sub>	-0.78	0.23	-3.40	0.00	MA <sub>1</sub>	-0.73	0.08	-9.13	0.00
$\omega$	0.00	0.00	1.04	0.30	$\omega$	0.00	0.00	0.00	1.00	$\omega$	0.00	0.00	0.00	1.00	$\omega$	0.00	0.00	0.00	1.00
$\alpha_1$	0.13	0.03	4.72	0.00	$\alpha_1$	0.35	0.14	2.48	0.01	$\alpha_1$	0.45	0.17	2.68	0.01	$\alpha_1$	0.35	0.14	2.48	0.01
$\beta_1$	0.83	0.03	28.67	0.00	$\beta_1$	0.49	0.07	6.69	0.00	$\beta_1$	0.47	0.12	3.88	0.00	$\beta_1$	0.49	0.07	6.69	0.00
$\gamma_1$	0.09	0.04	2.24	0.03	$\gamma_1$	0.32	0.17	1.93	0.05	$\gamma_1$	0.16	0.16	1.02	0.31	$\gamma_1$	0.32	0.17	1.93	0.05
skew	1.55	0.54	2.86	0.00	skew	-0.37	0.19	-1.93	0.05	skew	0.20	0.43	0.47	0.64	skew	-0.37	0.19	-1.93	0.05
shape	3.01	0.62	4.85	0.00	shape	1.15	0.18	6.55	0.00	shape	1.24	0.26	4.82	0.00	shape	1.15	0.18	6.55	0.00
[joint]dcca1	0.01	0.01	1.28	0.20	Log-Likelihood	1545.17	AIC	-3.20		[joint]dcca1	0.01	0.01	1.30	0.19	Log-Likelihood	892.30	AIC	-1.83	
[joint]dccb1	0.96	0.04	24.01	0.00	N	953	BIC	-3.10		[joint]dccb1	0.95	0.02	40.34	0.00	N	953	BIC	-1.73	

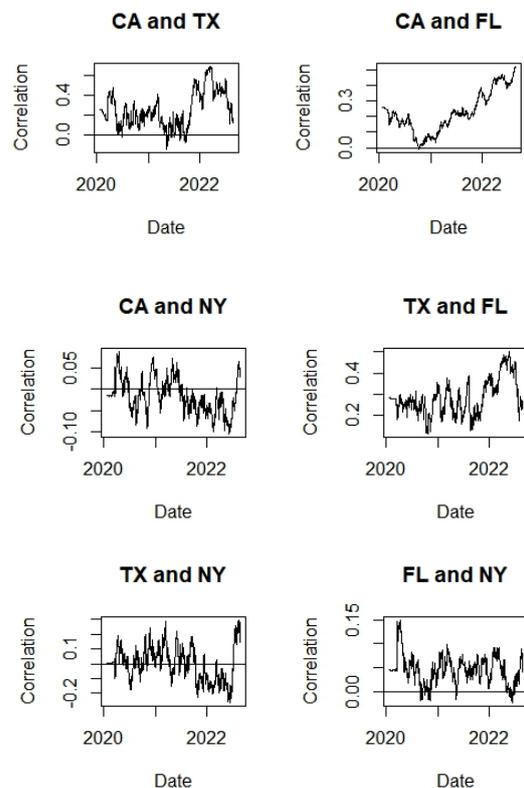


Figure 8. Plots of copula-DCC model for CA, TX, FL, and NY.

## 5. Discussion

We used the graphical models FPCA, GCMR, vine copula-based quantile regression, and copula-DCC for visual and data analysis of the COVID-19 mortality data in the 50 US states plus Washington, D.C., from the beginning of the COVID-19 pandemic to 1 September 2022 because the COVID-19 mortality data have a non-normal distribution and non-constant variance in the errors. Looking at the results of the graphical model, we found five equivalence groups in the US. Looking at the results of the FPCA, we visualized the COVID-19 daily mortality time trends of 50 states plus Washington, D.C. Using the GCMR model, we investigated the COVID-19 daily mortality relationships between four major states and the rest of the states in the US. Using the copula-DCC models, we investigated the time-varying dependence relationships between the COVID-19 daily mortality data of four major states (CA, TX, FL, and NY). Based on the findings of this research, geographical distance can be considered one of the main factors in the exponential increase in the number of pandemic patients from one state to the neighboring states in a short period of time. When a pandemic or an endemic happens in a certain state, the local state government needs to cooperate with federal and neighboring state governments to take immediate public health emergency measures. Time is one of the most important factors in suppressing a pandemic or endemic so we emphasize the need for a timely public health emergency response when facing a pandemic such as COVID-19. Every state in the US needs to regularly check its public health emergency manuals to prepare for future pandemics or natural disasters. In our future study, we will consider visualizing international COVID-19 mortality time-course pattern clustering by functional principal component analysis so that we can help to quickly control future pandemics.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The author thanks the respected editor and the two respected anonymous referees for the constructive and helpful suggestions that led to substantial improvements in the revised version and also thanks Grace Kim (Wayzata High School Student, MN) for her COVID-19 data collection help and discussion of this research topic.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Ramsay, J.; Silverman, B. *Functional Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2005.
2. Kokoszka, P.; Reimherr, M. *Introduction to Functional Data Analysis*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017.
3. Joe, H. *Multivariate Models and Multivariate Dependence Concepts*; CRC Press: Boca Raton, FL, USA, 1997.
4. Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
5. Tang, C.; Wang, T.; Zhang, P. Functional Data Analysis: An Application to COVID-19 Data in the United States. *arXiv* **2020**. [[CrossRef](#)]
6. Lia, X.; Zhang, P.; Feng, Q. Exploring COVID-19 in mainland China during the lockdown of Wuhan via functional data analysis. *Commun. Stat. Methods* **2022**, *29*, 103–125.
7. Oshinubi, K.; Ibrahim, F.; Rachdi, M.; Demongeot, J. Functional data analysis: Application to daily observation of COVID-19 prevalence in France. *AIMS Math.* **2022**, *7*, 5347–5385. [[CrossRef](#)]
8. Acal, C.; Escabias, M.; Aguilera, A.M.; Valderrama, M.J. COVID-19 Data Imputation by Multiple Function-on-Function Principal Component Regression. *Mathematics* **2021**, *9*, 1237. [[CrossRef](#)]
9. Cherubini, U.; Luciano, E.; Vecchiato, W. *Copula Methods in Finance*; John Wiley: Chichester, UK, 2004.
10. Kim, J.-M. A Review of Copula Methods for Measuring Uncertainty in Finance and Economics. *Quant. Bio-Sci.* **2020**, *39*, 81–90.
11. Sklar, A. Fonctions de repartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **1959**, *8*, 229–231.
12. Masarotto, G.; Varin, C. Gaussian copula marginal regression. *Electron. J. Stat.* **2020**, *6*, 1517–1549. [[CrossRef](#)]
13. Joe, H. Families of m-variate distributions with given margins and  $m(m-1)/2$  bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics*; Rüschemdorf, L., Schweizer, B., Taylor, M.D., Eds.; Lecture Notes—Monograph Series; Institute of Mathematical Statistics: Tokyo, Japan, 1996; pp. 120–141. [[CrossRef](#)]
14. Bedford, T.; Cooke, R.M. Vines—A new graphical model for dependent random variables. *Ann. Stat.* **2002**, *30*, 1031–1068. [[CrossRef](#)]

15. Aas, K.; Berg, D. Models for construction of multivariate dependence: A comparison study. *Eur. Financ.* **2009**, *15*, 639–659. [[CrossRef](#)]
16. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**, *44*, 182–198. [[CrossRef](#)]
17. Brechmann, E.C.; Schepsmeier, U. Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine. *J. Stat. Softw.* **2013**, *52*, 1–27. [[CrossRef](#)]
18. Kraus, D.; Czado, C. D-vine copula based quantile regression. *Comput. Stat. Data Anal.* **2017**, *110*, 1–18. [[CrossRef](#)]
19. D’Urso, P.; De Giovanni, L.; Vitale, V. A D-Vine Copula-Based Quantile Regression Model with Spatial Dependence for COVID-19 Infection Rate in Italy. *Spat. Stat.* **2022**, *47*, 100586. [[CrossRef](#)] [[PubMed](#)]
20. Taamouti, A.; Doukali, M.; Bouezmarni, T. Copula-Based Estimation of Health Concentration Curves with an Application to COVID-19. Available online: <https://ssrn.com/abstract=4064991> (accessed on 15 August 2022).
21. Ansell, L.; Valle, L.D. A New Data Integration Framework for Covid-19 Social Media Information. *arXiv* **2021**. [[CrossRef](#)]
22. Kalisch, M.; Hauser, A.; Maathuis, M.H.; Mächler, M. *An Overview of the pcalg Package for R*; CRAN of R Project: Vienna, Austria, 2019. Available online: <https://cran.microsoft.com/snapshot/2019-08-16/web/packages/pcalg/vignettes/vignette2018.pdf> (accessed on 15 August 2022).
23. Wang, J.-L.; Chiou, J.-M.; Müller, H.-G. Functional Data Analysis. *Annu. Rev. Stat. Appl.* **2016**, *3*, 257–295. [[CrossRef](#)]
24. Chen, K.; Zhang, X.; Petersen, A.; Müller, H.-G. Quantifying Infinite-Dimensional Data: Functional Data Analysis in Action. *Stat. Biosci.* **2017**, *9*, 582–604. [[CrossRef](#)]
25. Pearson, K.F.R.S. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [[CrossRef](#)]
26. Ramsay, J.O.; Hooker, G.; Graves, S. *Functional Data Analysis with R and MATLAB*; Springer Science & Business Media: New York, NY, USA, 2009.
27. Craven, P.; Wahba, G. Smoothing Noisy Data with Spline Functions. Estimating the Correct Degree of Smoothing by the Method of Generalized CrossValidation. *Numer. Math.* **1979**, *31*, 377–403. [[CrossRef](#)]
28. Dissmann, J.F.; Brechmann, E.C.; Czado, C.; Kurowicka, D. Selecting and estimating regular vine copulae and application to financial returns. *Comput. Stat. Data Anal.* **2013**, *59*, 52–69. [[CrossRef](#)]
29. Kim, J.-M.; Jung, H. Linear Time Varying Regression with Copula DCC-GARCH Model for Volatility. *Econ. Lett.* **2016**, *145*, 262–265. [[CrossRef](#)]
30. Ghalanos, A. The Rmgarch Models: Background and Properties. (Version 1.3-0). 2019. Available online: [https://cran.microsoft.com/snapshot/2020-05-24/web/packages/rmgarch/vignettes/The\\_rmgarch\\_models.pdf](https://cran.microsoft.com/snapshot/2020-05-24/web/packages/rmgarch/vignettes/The_rmgarch_models.pdf) (accessed on 15 August 2022).