Article

Nonparametric Directional Dependence Estimation and Its Application to Cryptocurrency

Hohsuk Noh [1]**, Hyuna Jang** [1] ◌**, Kun Ho Kim** [2] **and Jong-Min Kim** [3],* ◌

1   Department of Statistics, Sookmyung Women's University, Seoul 04310, Republic of Korea
2   John Molson School of Business, Concordia University, Montreal, QC H3H 0A1, Canada
3   Statistics Discipline, University of Minnesota-Morris, Morris, MN 56267, USA
*   Correspondence: jongmink@morris.umn.edu

Abstract: This paper proposes a nonparametric directional dependence by using the local polynomial regression technique. With data generated from a bivariate copula having a nonmonotone regression structure, we show that our nonparametric directional dependence is superior to the copula directional dependence method in terms of the root-mean-square error. To validate the directional dependence with real data, we use the log returns of daily prices of Bitcoin, Ethereum, Ripple, and Stellar. We conclude that our nonparametric directional dependence, by using the local polynomial regression technique with asymmetric-threshold GARCH models for marginal distributions, detects the directional dependence better than the copula directional dependence method by an asymmetric GARCH model.

Keywords: nonparametric estimation; directional dependence; copula; cryptocurrency

MSC: 62-07

1. Introduction

Since the COVID-19 pandemic, chaos and uncertainty in the financial market has been rising rapidly, resulting in investors having difficulty with the risk management of their financial portfolios. The authors in [1] studied the relationship between cryptocurrency price and the price of stock and gold using copulas. The copula function is useful as it relaxes the assumptions of normality, linearity, and independence of the residuals [2]. In particular, the application of the copula method in economics and finance is explained well in [3–12]

Gaussian copula marginal regression (GCMR) was proposed by [13], and a beta regression model to analyze bounded time series was proposed by [14]. With the GCMR method, copula directional dependence methods by asymmetric generalized autoregressive conditional heteroskedasticity (GARCH) and stochastic volatility models were developed by [15,16]. For the applications of the copula directional dependence model to cryptocurrency, the copula directional dependence using neural network models was applied to four major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Stella (XLM), and Ripple (XRP) [17]. However, the copula directional dependence uses the logit function, so it has difficulty detecting directional dependence when the data have an extremely nonlinear pattern. The purpose of this research was to propose a versatile and flexible directional dependence for all kinds of data patterns by using an optimized nonparametric method.

The novelty of this paper is that we are the first to apply a nonparametric method to directional dependence. The contribution of this paper is that our proposed nonparametric directional dependence method can detect nonlinear pattern directional dependence better than the current copula directional dependence or traditional linear regression directional dependence.

The remainder of this review paper is organized as follows: Section 2 describes the copula directional dependence. In Section 3, we explain our nonparametric directional dependence and present a numerical study. To verify the proposed directional dependence, a real data illustration with four cryptocurrencies is given in Section 4. Finally, conclusions are presented in Section 5.

2. Copula and Directional Dependence

A d-dimensional copula C is a multivariate joint distribution function associated with uniformly distributed marginals. In this article, we focus on a bivariate copula ($n = 2$). Let $F_{X,Y}(x, y)$ be the joint distribution function of a random vector (X, Y) with marginal distributions F_X and F_Y, respectively. According to Sklar's theorem [2], the distribution of $F_{X,Y}$ can be represented as $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$ for some copula C. Actually, the copula $C(u, v)$ is a joint distribution function of $U = F_X(X)$ and $V = F_Y(Y)$. Let $r_{V|U}(u)$ denote the conditional expectation of V given $U = u$. Using the regression function $r_{V|U}(u)$, ref. [18] defined a directional dependence from X to Y:

$$\rho^2_{X \to Y} = \frac{Var(r_{V|U}(U))}{Var(V)} = 12\, Var(r_{V|U}(U)). \tag{1}$$

In a similar way, we can define a directional dependence from Y to X:

$$\rho^2_{Y \to X} = \frac{Var(r_{U|V}(V))}{Var(U)} = 12\, Var(r_{U|V}(V)). \tag{2}$$

Note that $\rho^2_{X \to Y}$ is the nonparametric R-squared between U and V, which was proposed in [19]. Since the influence of the marginals has been eliminated in U and V, we can interpret $\rho^2_{X \to Y}$ as the margin-adjusted nonparametric R-squared. The two directional dependence measures $\rho^2_{X \to Y}$ and $\rho^2_{Y \to X}$ can be used to identify the stronger direction, i.e., a larger value indicates a stronger direction of influence.

To estimate the directional dependence, ref. [15] assumed a specific parametric form of the copula regression function. Concerning the random variables $U = F_X(X)$ and $V = F_Y(Y)$, they assumed that V_i given $U_i = u_i$ followed a beta distribution $Beta(\mu_i, \kappa_i)$ with the mean parameter $0 < \mu_i < 1$ and the precision parameter $\kappa_i > 0$. Specifically, the density function of $V_i|U_i = u_i$ was written as

$$f(v_i; \mu_i, \kappa_i) = \frac{\Gamma(\kappa_i)}{\Gamma(\mu_i \kappa_i)\Gamma((1 - \mu_i)\kappa_i)} v_i^{\mu_i \kappa_i - 1}(1 - v_i)^{(1 - \mu_i)\kappa_i - 1}, \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function. They assumed that the mean parameter $\mu_i = E(V_i|U_i = u_i)$ was linked with the covariate u_i through a logit function

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_0 + \beta_1 u_i \tag{4}$$

and the precision parameter κ_i was given as $\kappa_i = 1 + \exp(\beta_0 + \beta_1 u_i)$. They proposed to estimate the parameters (β_0, β_1) using the maximum likelihood method of GCMR [13,20]. Then, the directional dependence $\rho^2_{X \to Y}$ could be estimated as the sample variance of $\hat{\mu}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 u_i)/(1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 u_i))$ divided by the sample variance of u_i. $\rho^2_{Y \to X}$ could be estimated in a similar way. When transforming the given data (x_i, y_i) into uniform data (u_i, v_i), the empirical distribution function was used. The approach of [15] assumed that the regression functions defined between the transformed variables (U and V) had the form of a logistic curve, which is monotonic and has rotational symmetry. This assumption distorted the estimation of copula directional dependence, especially when the regression function was not a monotonic function. In Section 3, we show through numerical studies that their method has this issue in the estimation of the copula directional dependence in certain cases. The authors of [15] assumed that the uniform variable V had a condi-

tional distribution $V|U = u$, which followed a beta distribution $Beta(\mu_u, \kappa_u)$ with the mean parameter $\mu_u = E(V|U = u) = \exp(\beta_0 + \beta_1 u)/(1 + \exp(\beta_0 + \beta_1 u))$ and the precision parameter $\kappa_u = 1 + \exp(\beta_0 + \beta_1 u)$. In our numerical study, instead of assuming a specific form of the regression function, we simply estimate $r_{V|U}(u) = E(V|U = u)$ using the local polynomial regression technique [21]. Then, we estimate $\rho^2_{X \to Y}$ as the sample variance of $\hat{r}_{V|U}(u_i)$ divided by the sample variance of u_i.

3. Nonparametric Directional Dependence and Numerical Study

In our simulation, we assumed that the data were uniform, thus negating the need to convert it to a uniform distribution. We considered two scenarios. The first one was the case where the regression function between the transformed variables was monotonic, so the approach of [15] was expected to work well. The second one was the case where the regression function between the transformed variables was nonmonotonic.

In the first scenario, we generated data from an asymmetric copula distribution, which was constructed by combining two symmetric copulas [22] as

$$C_{\alpha,\beta}(u,v) = A_{\theta_1}(u^\alpha, v^\beta) A_{\theta_2}(u^{\bar{\alpha}}, v^{\bar{\beta}}), \tag{5}$$

where $0 < \alpha, \beta < 1, \bar{\alpha} = 1 - \alpha, \bar{\beta} = 1 - \beta$, and $A_\theta(u, v)$ is a Plackett copula with parameter $\theta > 0$ (Plackett, 1965). For the simulation, we set $\theta_1 = 5000, \theta_2 = 5, \alpha = 0.9$, and $\beta = 0.2$. Figure 1 shows the two regression functions between the transformed variables with their best approximations using logistic curves (in terms of L_2-distance). Note that both functions are monotonic but the regression function $E(U|V = V)$ was better approximated than the function $E(V|U = u)$. The directional dependence measures were $\rho^2_{U \to V} = 0.1759$ and $\rho^2_{V \to U} = 0.1242$.
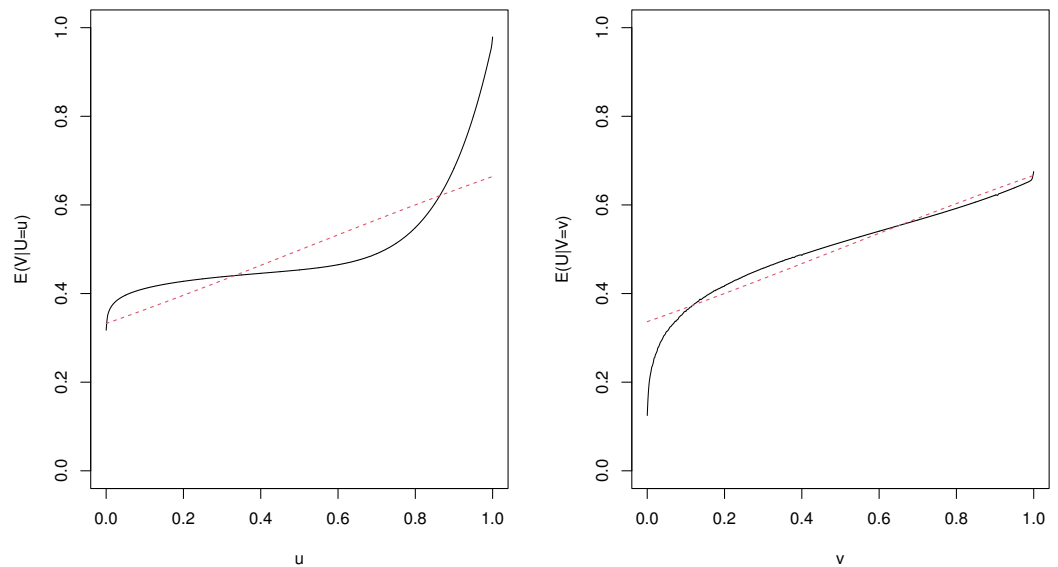

Figure 1. Regression functions between the transformed variables with their best approximations (dotted line) under the first scenario.

In the second scenario, we generated uniform data from a bivariate copula with a nonmonotone regression structure referring to the work of [23]. Ref. [23] devised a method to modify a bivariate copula family so that it had a nonmonotone regression structure using an appropriate measure-preserving function $f : [0, 1] \to [0, 1]$ that characterized the dependence structure. We chose $f(x) = 3xI(0 \le x \le 1/3) + (-(3/2)x + 3/2)I(1/3 \le x \le 1)$, where $I(\cdot)$ is an indicator function and modified a normal copula with parameter $\rho = 0.5$. Note that the nonmonotonicity of the measure-preserving function $f(x)$ produces

a nonmonotone regression function. The detailed steps of the modification are described as follows (Section 3 of [23]):

(1) Choose a measure-preserving function $f : [0,1] \to [0,1]$ that captures the desired dependence structure.

(2) Choose a parametric family of copulas $\{C_\alpha : \alpha \text{ is a copula parameter}\}$ (the copulas C_α admit densities ϕ_α).

(3) The resulting copula density function is given as $p_\alpha(u_1, u_2) = \phi_\alpha(f(x_1), x_2)$ and the corresponding distribution function is given as $D_\alpha(u_1, u_2) = \int_0^{u_1} \int_0^{u_2} p_\alpha(s_1, s_2) ds_1 ds_2$.

Figure 2 shows the two regression functions between the transformed variables with their best approximations using logistic curves (in terms of L_2-distance). Note that the function $E(U|V = v)$ is monotonic but the function $E(V|U = u)$ is not monotonic. Hence, the function $E(V|U = u)$ was not approximated by a logistic curve. Even the best approximation showed significant differences from the original function. The directional dependence measures were $\rho^2_{U \to V} = 0.2394$ and $\rho^2_{V \to U} = 0.0267$. Because of the monotonicity constraint, ref. [15]'s method was expected to have difficulties in estimating the directional dependence from U to V ($\rho^2_{U \to V}$).
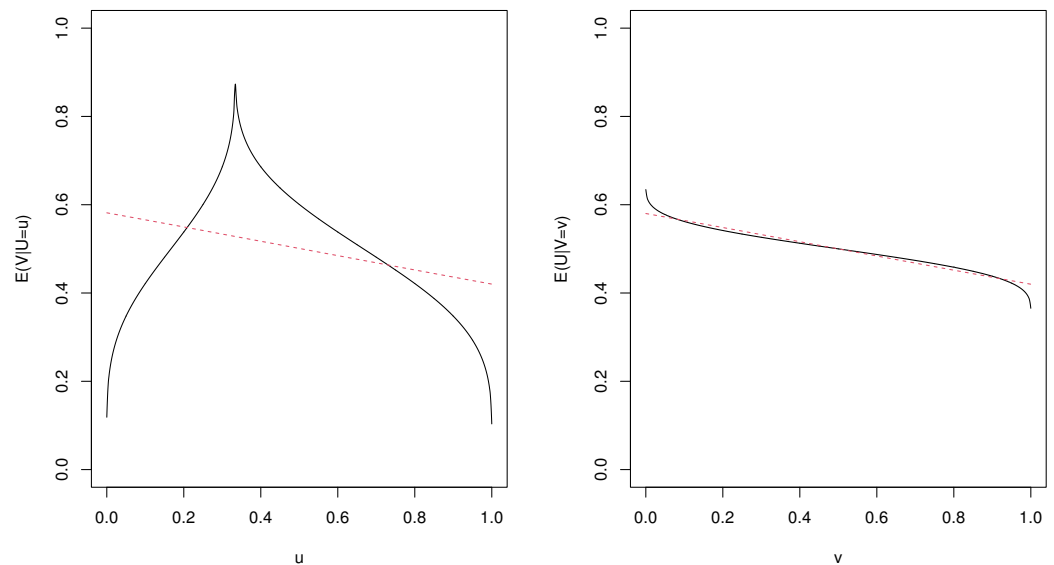

Figure 2. Regression functions between the transformed variables with their best approximations (dotted line) under the second scenario.

For the implementation of the copula directional dependence method by [15] (referred to as copula DD method hereafter), we used the R package *gcmr*, which also supports beta regressions. Through communication with the authors of [15], we learned that in their simulations they set κ as a constant. Therefore, we set the precision parameter κ to be constant as in their simulations. To implement the alternate estimation method using the local polynomial regression, we used the normal kernel and set the degree to be one (local linear estimator). For the bandwidth selection, we computed the plug-in-bandwidth selector implemented in the R package *locpol*.

Tables 1 and 2 show the simulation results. Under the first scenario, the two methods seemed to work well. The root-mean-square errors decreased as the sample size increased. However, under the second scenario, the copula method did not work when estimating the directional dependence from U to V. The reason was that the regression function could not be approximated using logistic curves as shown in the left panel of Figure 2. The proposed estimator did not suffer from such problems as expected.

Table 1. Root-mean-square error of the two estimators under the first scenario.

Sample Size	$u \rightarrow v$		$v \rightarrow u$	
	Nonparametric DD	Copula DD	Nonparametric DD	Copula DD
$n = 100$	0.077	0.083	0.076	0.056
$n = 200$	0.050	0.076	0.050	0.042
$n = 400$	0.035	0.072	0.033	0.031
$n = 800$	0.025	0.071	0.023	0.025
$n = 1600$	0.018	0.069	0.016	0.021

Table 2. Root-mean-square error of the two estimators under the second scenario.

Sample Size	$u \rightarrow v$		$v \rightarrow u$	
	Nonparametric DD	Copula DD	Nonparametric DD	Copula DD
$n = 100$	0.072	0.208	0.076	0.034
$n = 200$	0.052	0.212	0.040	0.022
$n = 400$	0.038	0.213	0.025	0.017
$n = 800$	0.028	0.214	0.015	0.011
$n = 1600$	0.021	0.214	0.010	0.008

4. Real Data Illustration

We wanted to show that the nonparametric directional dependence method performed better than the copula directional dependence method for financial market data. We applied both copula and nonparametric methods for the directional dependence to cryptocurrency data. Cryptocurrency is a digital currency that does not rely on any central authority to uphold or maintain it. Due to advances in computer technology, such as blockchain and cloud computing, the number of cryptocurrencies has rapidly increased. The rise of cryptocurrencies has led to many investors including cryptocurrencies in their financial portfolios.

Based on a high market cap and cryptocurrencies' initial offering history, we chose Bitcoin (BTC was created in 2009 and has a $431, 130, 583, 788$ market cap), Ethereum (ETH was created in 2015 and has a $191, 357, 725, 770$ market cap), Ripple (XRP was created in 2012 and has a $19, 151, 338, 397$ market cap), and Stellar (XLM was created in 2015 and has a $2, 227, 918, 509$ market cap). We collected BTC, ETH, XRP, and XLM daily prices from 1 July 2016 to 23 June 2021 from the following website (https://coinmarketcap.com/, accessed on 24 June 2021). We focused on the log return of the four cryptocurrency prices (in percentages) instead of the original prices. To eliminate the serial dependence of a given time series data, we fit the asymmetric threshold-GARCH (1,1) model to the log return data, generated standard residuals, and then transformed them to the uniformly distributed data using the empirical distribution function.

Using the four uniform datasets, we computed the directional dependence among the log returns using both copula and nonparametric methods. Figure 3 compares the violin plot pattern of the log returns of prices of BTC, ETH, XRP, and XLM (LBTC, LETH, LXRP, LXLM). As expected, LBTC had the smallest spread compared to the log returns of the three other cryptocurrencies (LETH, LXRP, LXLM), but LXLM had the largest spread compared to the log returns of the other three cryptocurrencies (LBTC, LETH, LXRP). Table 3 summarizes the descriptive statistics of the log return data of the cryptocurrencies such as the mean, skewness, and kurtosis as well as the maximum, minimum, range, standard deviation, and variance. Table 3 shows the average log return of ETH was the highest among the four cryptocurrencies. In Table 3, it is recognized that the standard deviation of LBTC was smaller than the other remaining cryptocurrencies, which means

that LBTC had a smaller risk than the other cryptocurrencies in terms of investment. Moreover, the values of the kurtosis for the four cryptocurrencies were greater than three, meaning heavy tails compared to a normal distribution. LBTC and LETH were left-skewed while LXRP and LXLM were right-skewed. This meant that the log returns of Bitcoin and Ethereum were more likely to decrease soon whereas the log returns of Ripple and Stellar were more likely to increase. Table 4 shows the directional dependence (DD) computed from the nonparametric and copula methods. We noticed the DD difference between the two methods. Using the nonparametric DD method, we found that the DD from LBTC to LETH was slightly higher than the DD from LETH to LBTC, but with the copula DD method, we found that the DD from LBTC to LETH was slightly lower than the DD from LETH to LBTC. In terms of market cap, BTC was two times bigger than ETH. In terms of the long-term period viewpoint, it makes sense that the log returns of other altcoins and stablecoins are more influenced by LBTC rather than the other directional case. In light of these facts, the nonparametric DD method performed better than the copula DD method proposed by [15].
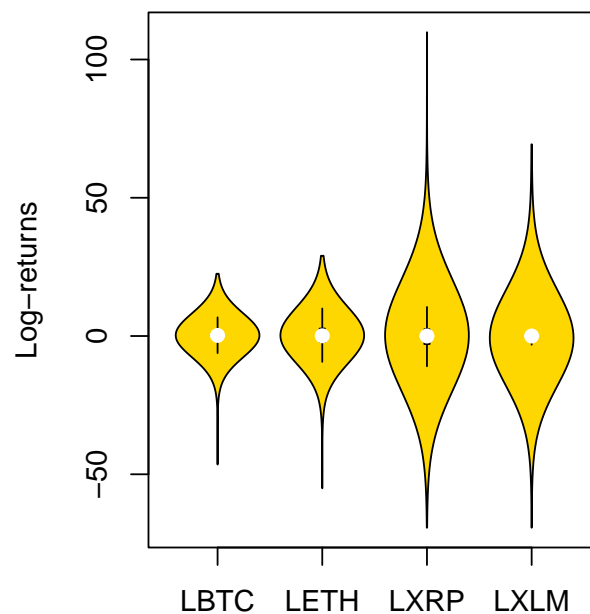

Figure 3. Log return violin plots of LBTC, LETH, LXRP, and LXLM.

Table 3. Summary Statistics of Log Returns (LBTC, LETH, LXRP, and LXLM).

	LBTC	**LETH**	**LXRP**	**LXLM**
Mean	0.213	0.277	0.242	0.260
Standard Deviation	4.133	5.661	7.896	9.118
Sample variance	17.080	32.046	62.342	83.145
Kurtosis	11.885	8.674	34.177	16.897
Skewness	−0.797	−0.418	2.135	1.133
Range	68.985	84.084	179.176	138.629
Minimum	−46.473	−55.071	−69.315	−69.315
Maximum	22.512	29.013	109.861	69.315
Sum	387.253	503.490	439.491	472.758
Count	1817	1817	1817	1817

Table 4. Directional dependence (DD) computed from the nonparametric and copula methods (the direction is from the row to the column).

Nonparametric DD Method				
Coin	**LBTC**	**LETH**	**LXRP**	**LXLM**
LBTC		0.45	0.26	0.18
LETH	0.44		0.36	0.21
LXRP	0.26	0.35		0.26
LXLM	0.18	0.23	0.27	
Copula DD Method				
Coin	**LBTC**	**LETH**	**LXRP**	**LXLM**
LBTC		0.39	0.21	0.14
LETH	0.4		0.29	0.18
LXRP	0.22	0.3		0.23
LXLM	0.14	0.17	0.21	

5. Conclusions

In this paper, we proposed an improved directional dependence by using a nonparametric method to rectify the issues of copula directional dependence used by [15]. Even though the computation time for our proposed method was longer than the current method, our proposed directional dependence did not require any assumptions and improved the accuracy of the proposed estimator in terms of root-mean-square error. Based on only two exemplifications, we also showed the performance of the nonparametric directional dependence method was superior to the copula directional dependence method with both simulated data and real data. In fact, the superiority of the nonparametric directional dependence method was not general in this paper. Thus, we need to consider the detailed general cases of the nonparametric directional dependence method in future studies. Our nonparametric directional dependence method will be useful in many research areas such as bioinformatics, economics, engineering, finance, geology, and neuroscience to see the directional dependence among many variables. In an era of big and complex data, our nonparametric directional dependence can be a good solution to reduce the level of uncertainty, which has been an issue in our society.

Author Contributions: Conceptualization, J.-M.K. and H.N.; methodology, J.-M.K. and H.N.; software, H.J. and H.N.; validation, J.-M.K., K.H.K. and H.N.; formal analysis, H.J., J.-M.K. and H.N.; investigation, J.-M.K., K.H.K. and H.N.; resources, H.N.; data curation, J.-M.K. and H.N.; writing original draft preparation, J.-M.K. and H.N.; writing review and editing, H.J., J.-M.K. and H.N.; visualization, J.-M.K. and H.N.; supervision, H.N.; project administration, J.-M.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kim, J.-M.; Kim, S.; Kim, S.T. On the Relationship of Cryptocurrency Price with US Stock and Gold Price using Copula Models. *Mathematics* **2020**, *8*, 1859. [CrossRef]
2. Sklar, M. Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **1959**, *8*, 229–231.
3. 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*; Rüschendorf, L., Schweizer, B., Taylor, M.D., Eds.; Institute of Mathematical Statistics: Washington, DC, USA, 1996; pp. 120–141.
4. Bedford, T.; Cooke, R.M. Vines—A new graphical model for dependent random variables. *Ann. Stat.* **2002**, *30*, 1031–1068. [CrossRef]
5. Aas, K.; Berg, D. Models for construction of multivariate dependence: A comparison study. *Eur. Financ.* **2009**, *15*, 639–659. [CrossRef]
6. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**, *44*, 182–198. [CrossRef]
7. Kurowicka, D.; Cooke, R.M. Distribution—Free continuous bayesian belief nets. In Proceedings of the Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice, Santa Fe, NM, USA, 21–25 June 2004.
8. Bauer, A.; Czado, C.; Klein, T. Pair-copula constructions for non-Gaussian DAG models. *Can. J. Stat.* **2012**, *40*, 86–109. [CrossRef]
9. Brechmann, E.; Czado, C.; Aas, K. Truncated and simplified regular vines in high dimensions with application to financial data. *Can. J. Stat.* **2012**, *40*, 68–85. [CrossRef]
10. Hobæk Haff, I.; Aas, K.; Frigessi, A. On the simplified pair-copula construction—Simply useful or too simplistic? *J. Multivar. Anal.* **2010**, *101*, 1296–1310. [CrossRef]
11. Panagiotelis, A.; Czado, C.; Joe, H. Pair Copula Constructions for Multivariate Discrete Data. *J. Am. Stat. Assoc.* **2012**, *107*, 1063–1072. [CrossRef]
12. Smith, M.; Min, A.; Almeida, C.; Czado, C. Modelling longitudinal data using a pair-copula decomposition of serial dependence. *J. Am. Stat.* **2010**, *105*, 1467–1479. [CrossRef]
13. Masarotto, G.; Varin, C. Gaussian copula marginal regression. *Electron. J. Stat.* **2012**, *6*, 1517–1549. [CrossRef]
14. Guolo, A.; Varin, C. Beta regression for time series analysis of bounded data, with application to Canada Google flu trends. *Ann. Appl. Stat.* **2014**, *8*, 74–88. [CrossRef]
15. Kim, J.-M.; Hwang, S.Y. Directional dependence via gaussian copula beta regression model with asymmetric garch marginals. *Commun. Stat. Simul. Comput.* **2017**, *46*, 7639–7653. [CrossRef]
16. Kim, J.-M.; Hwang, S.Y. The Copula Directional Dependence by Stochastic Volatility Models. *Commun. Stat. Simul. Comput.* **2019**, *48*, 1153–1175. [CrossRef]
17. Hyun, S.; Lee, J.; Kim, J.M.; Jun, C. What Coins lead in the Cryptocurrency Market? Using Copula and Neural Networks Model. *J. Risk Fin. Manag.* **2019**, *12*, 132. [CrossRef]
18. Sungur, E.A. Some observations on copula regression functions. *Commun. Stat. Theory Methods* **2005**, *34*, 1967–1978 [CrossRef]
19. Doksum, K.; Samarov, A. Nonparametric Estimation of Global Functionals and a Measure of the Explanatory Power of Covariates in Regression. *Ann. Stat.* **1995**, *23*, 1443–1473. [CrossRef]
20. Ferrari, S.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [CrossRef]
21. Fan, J.; Gijbels, I. *Local Polynomial Modelling and Its Applications*; Chapman & Hall: London, UK, 1996.
22. Durante, F. Construction of non-exchangeable bivariate distribution functions. *Stat. Pap.* **2009**, *50*, 383–391. [CrossRef]
23. Scarsini, M.; Venetoulias, A. Bivariate distributions with nonmonotone dependence structure. *J. Am. Stat. Assoc.* **1993**, *88*, 338–344.