

# ***Analysing the Impact of Digital Economy Development on the Income Gap Between Urban and Rural Residents Based on Copula Function and Kernel Density Estimation***

**Yifei Xu**

*School of Mathematical Sciences, Dalian University of Technology, Dalian, China  
xyf@mail.dlut.edu.cn*

**Abstract:** With the advancement of technology, the digital economy is booming, a change that has many implications for the urban-rural divide. In this study, the Thiel index is used to measure the income gap between urban and rural residents, establish an evaluation system to measure the digital economy development index, and explore the relationship between the two based on the Copula function. Firstly, the marginal distributions of the two variables were modelled using non-parametric kernel density estimation and tested for goodness-of-fit using root-mean-square error. The Copula function was then used to simulate the joint distribution, and the tail correlation coefficients were calculated by comparing the function to the original data Spearman rank correlation coefficients and calculating the Euclidean distance to test the goodness-of-fit and hence the tail correlation coefficients. The results show that (1) Epanechnikov kernel function, rectangular kernel function, and t-Copula are the best fits for modelling the digital economic development index, Terre index, and joint distribution, respectively. (2) The upper and lower tail correlation coefficients of the two variables are -0.6931 and -0.6784 respectively, indicating that when there is an extremely large value of the Digital Economy Development Index, the income gap between urban and rural residents tends to be extremely small; and vice versa. The findings of this study will be useful in helping policy makers to formulate measures to promote coordinated urban-rural economic development and narrow the urban-rural gap.

**Keywords:** KDE, Copula, Correlation analysis, Marginal distribution

## **1. Introduction**

People's need for convenience and quality of life has led to the creation of the digital economy, and the development of technology and the Internet has provided support for the popularisation of the digital economy. The digital economy is booming and has become closely related to everyone's life. However, it is impossible to measure the level of development of the digital economy with a single indicator. This paper assigns weights to a number of indicators that are strongly related to the digital economy, and establishes a "scoring" system called the Digital Economy Development Index (DEDI) to quantify the level of digital economy development in different regions in different years.

The uneven distribution of resources between urban and rural areas, the relatively backward development of the rural economy and technology, and the relative lag in industrial upgrading have

led to a significant income gap between urban and rural residents. The impact of the development of digital economy on the income gap between urban and rural residents is mainly reflected in the following aspects. Through the Internet, the digital economy has provided a platform for the promotion and sale of agricultural products, which has brought agricultural products into the public's view and increased sales; on the other hand, the development of the Internet has enabled farmers to receive more knowledge and technology and apply them to production, which has improved their motivation and production efficiency. However, due to their low level of knowledge, lagging and limited access to information, farmers are unable to take full advantage of the opportunities brought by the Internet compared to urban residents. This may increase the income gap. In digital transactions, the development of the digital economy has facilitated the development of e-commerce and the popularity of the logistics industry, which has facilitated the transport of rural products out of the countryside and increased rural incomes. The development of the digital economy provides more employment opportunities, which will increase incomes in rural areas, but due to the backwardness of education and the lack of knowledge and talent, farmers can mostly only choose jobs with low technical requirements and low pay, which may increase the income gap. There is already literature confirming the correlation between the two: Liu Jiadong et al. [1] used threshold regression analysis and dynamic panel generalised moments estimation to study the specific effect of the digital economy on the urban-rural income gap, and found that the threshold effect between the two variables; Xu Weixiang et al. [2] used the generalised spatial nested model (GNS) to study the potential of the development of the digital economy on urban-rural income gap spatial effects, and further found that digital economy has significant spatial spillover effects.

In the processing of digital economic development index weights, Sun Tiantian [3], Zhao Tao et al. [4], Meng Weifu et al. [5] chose the principal component analysis method, Yuan Lemon [6], Wang Jie et al. [7], Wang Xing et al. [8] used the entropy method, Liu Jun et al. [9] used the NBI index weights determination method and linear weighting method. In terms of urban-rural residents' income gap, Zhang Yujing et al. [10] used the absolute income gap between urban and rural residents to measure it, and Zhang Shaohua et al. [11] used the ratio of urban and rural residents' income to measure it. In contrast, the Thiel index takes into account population changes and decomposes the urban-rural income gap into intra-group and inter-group gaps, which can better demonstrate the changes in the urban-rural income gap. Therefore, existing studies generally choose the Thiel index to represent the urban-rural income gap, such as: Wang Liangjun et al. [12], Liang Shuanglu et al. [13], Cheng Mingwang et al. [14], Zhang He et al. [15].

In this paper, we adopt a non-parametric estimation method that uses kernel density estimation to simulate the marginal distribution of the digital economy index and the income gap between urban and rural residents. This method does not rely on any distributional assumptions. The article uses the copula function to study the joint distribution between the two variables and the correlation between the two variables in the tails. Chapter 2 of the article labels the sources of data and provides descriptive statistics of the data. Chapter 3 describes the methodology used in the article and its formulae. Chapter 4 gives the results of calculations and analyses. Chapter 5 presents the conclusions. Chapter 6 presents the discussion.

## 2. Data resource

In measuring the digital economy development index, this paper considers the data of 30 provinces (municipalities and autonomous regions) except Tibet during the period of 2007-2021, and selects 12 indicators to establish the evaluation index system, including the value added of the financial industry, the turnover of the technology market, and the funding of R&D, from the perspective of three aspects, namely, the development of the information and network, the development of the digital transaction, and the capacity of digital innovation. The data were mainly obtained from the

National Bureau of Statistics, statistical yearbooks of provinces, the China Statistical Yearbook, the China Science and Technology Statistical Yearbook, and the State Postal Administration.

### 3. Method

#### 3.1. Entropy weighting

In thermodynamics, entropy is a measure of the chaos or disorder of a system, with greater entropy indicating a more chaotic system. Information entropy refers to the amount of information contained in a source, and is used to describe the amount of information in an event on average, and mathematically refers to the expectation of the amount of information contained in an event(Eq.1).

$$H(X) = - \sum_{i=1}^n p(x_i) \log(p(x_i)) \quad (1)$$

For a certain indicator, entropy can be used to judge its degree of dispersion, the smaller the information entropy, the greater the degree of dispersion of the indicator, the greater the weight of the indicator.

There are  $m$  indicators  $X_1, X_2 \dots, X_m, X_i = \{x_{i1}, x_{i2} \dots, x_{in}\}$ . These data were first standardised (Eq.2, 3).

$$y_{ij} = \frac{x_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)} \quad (2)$$

$$y_{ij} = \frac{\max(X_i) - x_{ij}}{\max(X_i) - \min(X_i)} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (3)$$

where Equation 2 is the normalisation formula for positive indicators and Equation 3 is the normalisation formula for negative indicators.

For the same indicator, the value of the data is calculated as a proportion of the total value(Eq.4).

$$p_{ij} = \frac{y_{ij}}{\sum_{j=1}^n y_{ij}} \quad (4)$$

The following calculates the information entropy of each indicator based on the definition of information entropy. To ensure that each logarithmic calculation is meaningful, the information entropy is calculated by first adding 0.001 to the weight calculated above(Eq.5). In turn, the weights of the indicators are calculated (Eq.6).

$$E_i = - \frac{\sum_{j=1}^n (p_{ij} + 0.001) \cdot \ln(p_{ij} + 0.001)}{\ln(n)} \quad (5)$$

$$w_i = \frac{1 - E_i}{\sum_{i=1}^m 1 - E_i} \quad (6)$$

Finally, a composite score is calculated(Eq.7).

$$S_i = \sum_{i=1}^m w_i \cdot y_{ij} \quad (7)$$

#### 3.2. Kernel Density Estimate(KDE)

When estimating probability densities, non-parametric estimation methods are often used for data with uncertain distributions, the more common of which is the kernel density estimation (KDE) method. KDE simulates the marginal distributions by selecting the kernel function to compute the corresponding probability density function.

For the variable  $X = \{x_1, x_2 \dots, x_n\}$ , calculate the probability density at a given point  $x$ (Eq.8).

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (8)$$

where  $h$  refers to the bandwidth and  $K(\cdot)$  refers to the kernel function. Common kernel functions are Gauss, Epanechnikov, Triangle, Uniform, etc.

### 3.3. Binary copula functions

Sklar pointed out that for any  $n$ -dimensional joint distribution function, there exists a Copula function  $C$  such that this joint distribution can be represented by its marginal distributions and the Copula function. The converse is also true. Thus, the Copula function can combine the marginal distributions of two variables into a joint distribution, which can be used to study the dependence between the two variables.

Assume that the marginal distributions of  $X$  and  $Y$  are  $F_X(x)$  and  $F_Y(y)$ , respectively, and the joint distribution is  $F(x, y)$ . By Sklar's theorem, there is the following equation (Eq.9):

$$F(x, y) = C(F_X(x), F_Y(y)) \quad (9)$$

where  $C$  is the Copula function. Some common binary Copula functions are Binary Normal Copula function,  $t$ -Copula function, Clayton-Copula function, Frank-Copula function, Gumble-Copula function, and so on.

Firstly, the appropriate Copula function is selected based on the marginal distribution, and the values of each parameter in the function are calculated to obtain the complete Copula function. Then the Copula function is tested for goodness of fit.

### 3.4. Correlation coefficient

#### 3.4.1. Spearman's rank correlation coefficient

The Spearman rank correlation coefficient is a nonparametric measure of rank correlation that investigates the monotonic relationship between two random variables.

Let the two variables  $X = \{x_1, x_2, \dots, x_n\}$  and  $Y = \{y_1, y_2, \dots, y_n\}$ . Calculate the Spearman's rank correlation coefficient for the two variables (In Eq. 10).

$$\rho = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)} \quad (10)$$

Let the marginal distributions of  $X$  and  $Y$  be  $F_X(x)$  and  $F_Y(y)$ , respectively, and the joint distribution be  $F(x, y)$ . Denote  $u = F_X(x)$ ,  $v = F_Y(y)$  and the Copula function is  $C(u, v)$ . Calculate the Spearman's rank correlation coefficient for the Copula function (In Eq. 11).

$$P = 12 \iint_{[0,1]^2} C(u, v) du dv - 3 \quad (11)$$

#### 3.4.2. Tail correlation coefficient

The tail correlation coefficient is used to measure the correlation between two variables at the extremes. Let the two variables be  $X, Y$ . A threshold  $\alpha$  (expressed as a percentage) is first set for an extreme event, i.e., the largest  $\alpha$  portion and the smallest  $\alpha$  portion of the data are taken as extreme events (tails). The data points in which both variables exceeded the high threshold in all data points were counted, and the upper-tailed correlation coefficients were obtained by calculating the Spearman's rank correlation coefficients between  $X$  and  $Y$  in the data points. Similarly, the data points that appear in both  $X$  and  $Y$  lower-tailed portions are counted, and the Spearman rank correlation coefficient between  $X, Y$  in the data points is calculated to obtain the lower-tailed

correlation coefficient. The formula is calculated with reference to Equation 10 for the rank correlation coefficient of the two variables.

### 3.5. Calculation of goodness-of-fit

#### 3.5.1. Root Mean Square Error (RMSE)

It is used to characterize the goodness of fit of the kernel function in kernel density estimation (Eq.12).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [F_e(i) - F_i(i)]^2}{n}} \quad (12)$$

where  $F_e(i)$  is the empirical frequency and  $F_i(i)$  is the modelled frequency of the kernel density function.

### 3.6. Euclidean distance

Used to describe the goodness of fit of a function (Eq.13).

$$d(C, C_{emp}) = \sqrt{\sum_{i=1}^n [C(x_i, y_i) - C_{emp}(x_i, y_i)]^2} \quad (13)$$

$$C_{emp}(u, v) = \frac{1}{n} \sum_{i=1}^n I_{[F_X(x_i) \leq u]} I_{[F_Y(y_i) \leq v]} \quad (14)$$

Where  $c_{emp}$  is the empirical Copula function;  $R(x_i) = \text{rank}(x_i)$  denotes the order of  $x_i$  among all  $x$ 's;  $I[\cdot]$  is the indicator function, which returns 1 when the condition is valid, and 0 otherwise; and  $F_X, F_Y$  denote the marginal distribution functions of the two variables, respectively.

## 4. Result

### 4.1. Measurement of the digital economy development index

This paper calculates the digital economy development index (DEDI) for 30 provinces (municipalities and autonomous regions) except Tibet. Based on the existing literature, this paper starts from the three primary indicators and selects 12 secondary indicators to measure the DEDI.

The data were first Min-Max normalised to convert all data into numbers between  $[0,1]$  for subsequent calculations. The entropy weight method was adopted to calculate the weight of each secondary indicator, as shown in the following table (Table 1). The weights of the indicators were weighted to calculate the DEDI for each province for different years.

Table 1: Indicator weights

| Indicators                          | Weights | Secondary indicators                                                                                                     | Weights | Properties |
|-------------------------------------|---------|--------------------------------------------------------------------------------------------------------------------------|---------|------------|
| Information and network development | 0.1649  | Average wage of persons employed in urban units in the information transmission, computer services and software industry | 0.0345  | Positive   |
|                                     |         | Number of Internet broadband access ports / size of population                                                           | 0.0423  | Positive   |
|                                     |         | Cell phone year-end subscribers / size of population                                                                     | 0.0171  | Positive   |
|                                     |         | Value added of the financial sector                                                                                      | 0.0710  | Positive   |
| Digital                             | 0.4220  | Internet business density                                                                                                | 0.0306  | Positive   |

Table 1: (continued)

|                                 |        |                                                     |        |          |
|---------------------------------|--------|-----------------------------------------------------|--------|----------|
| trading development             |        | Online shopping density                             | 0.0354 | Positive |
|                                 |        | Technology market turnover                          | 0.1589 | Positive |
|                                 |        | Express mail volume                                 | 0.1971 | Positive |
| Digital innovation capabilities | 0.4131 | R&D funding                                         | 0.0903 | Positive |
|                                 |        | Local finance expenditure on science and technology | 0.0909 | Positive |
|                                 |        | Domestic Patent Applications Received               | 0.1100 | Positive |
|                                 |        | Domestic Patent Applications Granted                | 0.1219 | Positive |

Descriptive statistical analysis of the digital economy development index reveals that the digital economy development index increases year by year, with the average value rising from 0.0304 to 0.2255. this indicates that the digital economy continues to develop within these 30 provinces. The standard deviation of the digital economy development index keeps increasing, reaching 0.187 in 2021, indicating that the development of the digital economy in different regions of different provinces is not balanced and differs greatly. The descriptive statistics of the digital economy development index are shown in the table below (Tables 2).

Table 2: Descriptive statistics of the digital economy development index

| year     | average | median | Std.   | min    | max    | N   |
|----------|---------|--------|--------|--------|--------|-----|
| 2021     | 0.2255  | 0.1752 | 0.1870 | 0.0701 | 0.8738 | 30  |
| 2020     | 0.1984  | 0.1599 | 0.1650 | 0.0619 | 0.7665 | 30  |
| 2019     | 0.1757  | 0.1361 | 0.1414 | 0.0606 | 0.6548 | 30  |
| 2018     | 0.1599  | 0.1234 | 0.1247 | 0.0582 | 0.5727 | 30  |
| 2017     | 0.1378  | 0.1027 | 0.1040 | 0.0468 | 0.4570 | 30  |
| 2016     | 0.1246  | 0.0951 | 0.0934 | 0.0429 | 0.3980 | 30  |
| 2015     | 0.1089  | 0.0838 | 0.0822 | 0.0362 | 0.3245 | 30  |
| 2014     | 0.0921  | 0.0725 | 0.0682 | 0.0326 | 0.2721 | 30  |
| 2013     | 0.0845  | 0.0639 | 0.0674 | 0.0262 | 0.2519 | 30  |
| 2012     | 0.0740  | 0.0512 | 0.0599 | 0.0225 | 0.2209 | 30  |
| 2011     | 0.0616  | 0.0427 | 0.0502 | 0.0164 | 0.1946 | 30  |
| 2010     | 0.0512  | 0.0345 | 0.0421 | 0.0144 | 0.1590 | 30  |
| 2009     | 0.0422  | 0.0299 | 0.0380 | 0.0097 | 0.1438 | 30  |
| 2008     | 0.0352  | 0.0232 | 0.0318 | 0.0082 | 0.1260 | 30  |
| 2007     | 0.0304  | 0.0194 | 0.0289 | 0.0071 | 0.1100 | 30  |
| umbrella | 0.1068  | 0.0707 | 0.1135 | 0.0071 | 0.8738 | 450 |

## 4.2. Urban-rural income gap

Referring to the existing literature [12], the Thiel index can better portray the changes in the income gap between urban and rural residents. In this paper, the Thiel index is used to express the urban-rural income gap.

Descriptive statistical analysis of the Terrell index of income disparity between urban and rural residents shows that the Terrell index has been decreasing year by year, with the average value decreasing from 0.1323 to 0.0656, which indicates that the income gap has been gradually narrowing in these 30 provinces. The standard deviation of the income gap gradually decreases,

decreasing to 0.0283 in 2021, indicating that regions with originally large gaps have seen more gap reduction in these 15 years, and the urban-rural gap has effectively shrunk. In addition, combining the descriptive statistics of the two variables, there is a largely negative correlation between the income gap between urban and rural residents and the digital economy development index. The descriptive statistics of the income gap between urban and rural residents are shown in the table below (Table 3).

Table 3: Descriptive statistics of income disparity between urban and rural residents

| year     | average | median | Std.   | min    | max    | N   |
|----------|---------|--------|--------|--------|--------|-----|
| 2021     | 0.0656  | 0.0625 | 0.0283 | 0.0180 | 0.1350 | 30  |
| 2020     | 0.0697  | 0.0670 | 0.0299 | 0.0180 | 0.1440 | 30  |
| 2019     | 0.0763  | 0.0725 | 0.0319 | 0.0190 | 0.1530 | 30  |
| 2018     | 0.0804  | 0.0760 | 0.0337 | 0.0190 | 0.1580 | 30  |
| 2017     | 0.0838  | 0.0785 | 0.0352 | 0.0190 | 0.1640 | 30  |
| 2016     | 0.0866  | 0.0805 | 0.0366 | 0.0200 | 0.1690 | 30  |
| 2015     | 0.0894  | 0.0845 | 0.0376 | 0.0200 | 0.1710 | 30  |
| 2014     | 0.0923  | 0.0890 | 0.0391 | 0.0210 | 0.1780 | 30  |
| 2013     | 0.0976  | 0.0950 | 0.0415 | 0.0210 | 0.1870 | 30  |
| 2012     | 0.1037  | 0.1015 | 0.0440 | 0.0220 | 0.1970 | 30  |
| 2011     | 0.1070  | 0.1050 | 0.0454 | 0.0220 | 0.2020 | 30  |
| 2010     | 0.1157  | 0.1145 | 0.0471 | 0.0220 | 0.2100 | 30  |
| 2009     | 0.1266  | 0.1250 | 0.0514 | 0.0240 | 0.2300 | 30  |
| 2008     | 0.1273  | 0.1245 | 0.0522 | 0.0250 | 0.2260 | 30  |
| 2007     | 0.1323  | 0.1285 | 0.0563 | 0.0240 | 0.2510 | 30  |
| umbrella | 0.0970  | 0.0910 | 0.0457 | 0.0180 | 0.2510 | 450 |

### 4.3. Determination of edge distribution

In this paper, we use non-parametric kernel density estimation. Since too small a bandwidth will lead to too rough a fitted kernel function; too large a bandwidth will make the estimated results too smooth and too far from the original data to represent the real data. Therefore, in this paper, the smoothness of the function and the gap between the function and the data are considered to select the bandwidth. The simulation estimation using four kernel density functions, Gaussian, Epanechnikov, triangular kernel function, and rectangular kernel function, respectively, is shown below (Fig. 1).

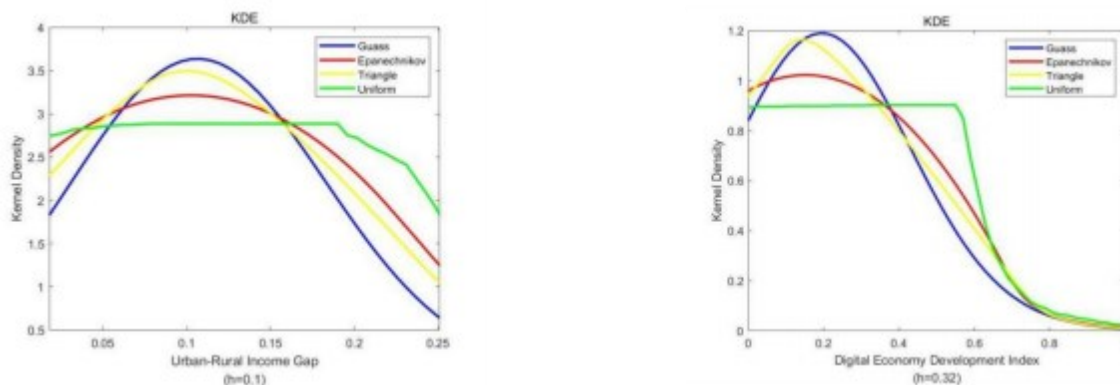


Figure 1: Kernel density estimates for both variables

The Root Mean Square Error (RMSE) between the simulation results and the actual values of the four functions are further calculated as follows (Table 4).

Table 4: Root-mean-square error of the kernel function with respect to the actual data

|                        | Gaussian | Epanechnikov | Triangle | Uniform |
|------------------------|----------|--------------|----------|---------|
| DEDI                   | 0.1445   | 0.0894       | 0.1312   | 0.1244  |
| Urban-rural income gap | 2.1864   | 2.0382       | 2.1631   | 1.8587  |

Through comparison, it can be found that when simulating the digital economic development index, the root-mean-square error of the four kernel functions does not differ much, and the error of Epanechnikov kernel function is the smallest, so Epanechnikov kernel function is chosen to estimate the marginal distribution of the digital economic development index; when simulating the urban-rural income disparity, the root-mean-square error of the rectangular kernel function is relatively small, so the rectangular kernel function is chosen to simulate the marginal distribution of the urban-rural income gap. The distribution of kernel density in different years was plotted as follows (Figures 2 and 3).

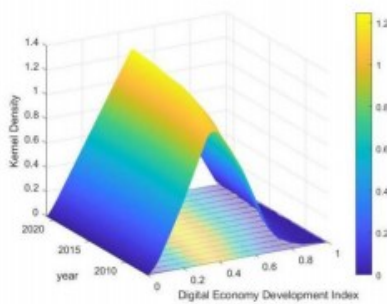


Figure 2: Epanechnikov KDE of the digital economy development index

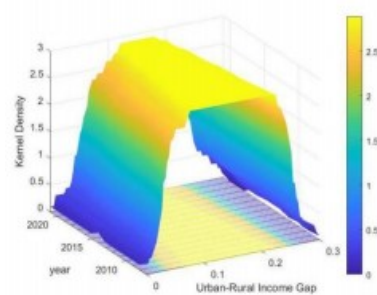


Figure 3: Rectangular KDE of the urban-rural income gap

The DEDI modelled by the Epanechnikov kernel function has a single peak distribution with a gradually increasing peak and a peak kernel density that increases year by year. This indicates that the DEDI has an upward trend and the degree of aggregation increases. The high-density region of the urban-rural income gap simulated by the rectangular kernel function gradually becomes narrower and tends to have a single-peaked distribution; the data are distributed in 0-0.3, concentrated in 0.07-0.22, and have symmetry, and the data are distributed more evenly. On the other hand, comparing the two images, the urban-rural income gap has a small width and a large peak of 2.9, with a more concentrated data distribution.

#### 4.4. Fitting of Copula functions and calculation of tail correlation coefficients

Based on the level of digital economy development in each province and the marginal distribution of urban-rural income gap, the parameters of five functions, Gaussian-Copula, t-Copula, Clayton-Copula, Frank-Copula and Gumbel-Copula, are calculated and fitted to the joint distribution respectively. The Spearman's correlation coefficients for the raw data and the five functions were further calculated and compared as follows (Table 5).

Table 5: Spearman's rank correlation coefficients of raw data with Copula function

| raw data | Gaussian | t       | Clayton | Frank   | Gumbel                 |
|----------|----------|---------|---------|---------|------------------------|
| -0.7836  | -0.5373  | -0.8242 | 0.4041  | -0.8527 | $2.051 \times 10^{-6}$ |

The Euclidean distances of the five function simulations were further calculated to determine their goodness of fit as follows (Table 6).

Table 6: Euclidean distance of Copula function

|       | Gaussian | t      | Clayton | Frank  | Gumbel |
|-------|----------|--------|---------|--------|--------|
| Value | 0.5465   | 0.0530 | 8.0495  | 0.0717 | 3.9347 |

It is not difficult to find that the correlation coefficients of the two functions, binary t-Copula and Frank-Copula, are close to the original data, and the Euclidean distance of the binary t-Copula function is the smallest, which indicates that the binary t-Copula function has the best fitting effect. Therefore, in this paper, the binary t-Copula function is chosen to model the joint distribution and the probability density function is used to portray the relationship between the two variables. The distribution function and density function of the joint distribution are shown below (Figures 4 and 5).

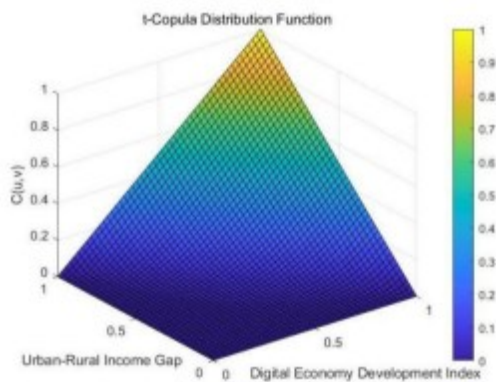


Figure 4: T-Copula distribution function

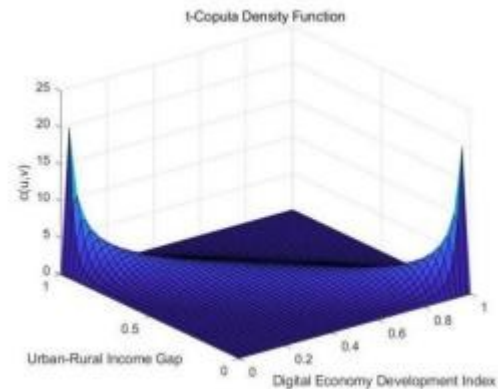


Figure 5: T-Copula density function

From the figure, the density function is t-distributed and the probability density increases abruptly in the upper and lower tails, almost to 20, which indicates that both variables are more sensitive in the upper and lower tails. In addition, the function image has high symmetry.

The upper and lower tail correlation coefficients of the level of development of the digital economy and the urban-rural income gap are further calculated. Setting the extreme event threshold at 10%, the upper and lower tail correlation coefficients of the two variables were calculated to be -0.6931 and -0.6784, respectively, with high tail correlations. This shows that when the DEDI has a very large value, the income gap between urban and rural residents tends to have a very small value; when the DEDI has a very small value, the income gap between urban and rural residents tends to have a very large value. There is a negative correlation between the two variables.

## 5. Discussion

In determining the marginal distributions of the two variables, this paper uses a non-parametric kernel density estimation. Four kernel functions, Gaussian kernel function, Epanechnikov kernel function, trigonometric kernel function, and rectangular kernel function, were used to fit the data of the two variables, and the root mean square error was used to judge the goodness of fit. For the DEDI, the root mean square error of the Gaussian kernel function, Epanechnikov kernel function, triangular kernel function, and rectangular kernel function are 0.1445, 0.0894, 0.1312, and 0.1244, respectively, and the Epanechnikov kernel function has the smallest error, so the Epanechnikov kernel function is chosen to simulate the DEDI. For the Terrell index, the root mean square errors of

the four functions are 2.1864, 2.0382, 2.1631, and 1.8587, respectively, and the rectangular kernel function has the smallest error, so the rectangular kernel function is chosen to fit the Terrell index. Gaussian-Copula, t-Copula, Clayton-Copula, Frank-Copula, and Gumbel-Copula were used to simulate the joint distribution of the two variables, respectively. The Spearman rank correlation coefficients of the five functions were calculated as -0.5373, -0.8242, 0.4041, -0.8527, and  $2.051 \times 10^{-6}$ , respectively, which were found to be close to t-Copula and Frank-Copula when compared with Spearman rank correlation coefficients of -0.7836 for the original data. The Euclidean distances of the five functions were further calculated as, 0.5465, 0.0530, 8.0495, 0.0717, and 3.9347, and the t-Copula was found to be the smallest. Hence t-Copula was chosen as the joint distribution function of the two variables. The Copula function image has significant peaks near A tends to 0, B tends to 1 and A tends to 1, B tends to 0, which indicates high tail correlation of the joint distribution. Further tail correlation coefficients were calculated. With an upper tail correlation coefficient of -0.6931 and a lower tail correlation coefficient of -0.6784, there is a strong negative correlation between the DEDI (DEDI) and the urban-rural income gap in the tails, which suggests that the probability of a union of the two variables occurs at these extreme values (i.e., a very high or low DEDI and a very low or very high urban-rural income gap) is high. Whereas the function images are relatively low in the intermediate region (i.e., A and B are close to 0.5), suggesting that the probability of the union of these two variables occurring is low in the case of these intermediate values. This could mean that in regions with extreme (very high or very low) levels of digital economy development, the urban-rural income gap also tends to be extreme (very small or very large).

The impact of the digital economy on the urban-rural income gap shows an inverted "U" shape, which indicates that the impact of the digital economy on the urban-rural income gap is complex, and that it may not be conducive to narrowing the gap at the initial stage, but will contribute to the narrowing of the urban-rural income gap and to the promotion of common prosperity after it has reached a certain stage of development.

## 6. Conclusion

This paper selects the Thiel index of urban-rural residents' income gap in 30 provinces (municipalities and autonomous regions) except Tibet from 2007 to 2021 for research, and selects 12 indicators to construct the DEDI. The marginal distributions of the two variables were modelled using kernel density estimation, and then the joint distribution was modelled using the t-Copula function, and a strong tail correlation was found between the two variables. Further calculation of upper tail correlation coefficient of -0.6931 and lower tail correlation coefficient of -0.6784 confirms a strong negative correlation between the two variables in the tails. This study contributes to helping policymakers understand the relationship between the development of the digital economy and the urban-rural income gap, so that they can formulate more effective policies to promote balanced economic development and reduce income inequality. However, due to data limitations, the selection of indicators for the DEDI is small and not comprehensive enough, and indicators such as e-commerce sales and online retail sales can be added for consideration. The negative correlation between the DEDI and the income gap between urban and rural residents in the tails may be due to the fact that industries such as "Internet+" are concentrated in large cities, which are more attractive, with most of the resources flowing into urban areas, providing many opportunities for urban residents to improve their salaries. On the other hand, rural areas around big cities are backward in terms of infrastructure and education, traditional industries are hardly subject to digital transformation, and the income gap between urban and rural residents is even greater. Based on the above analyses, it is possible to promote the development of the digital economy in

rural areas by upgrading the rural learning environment, thereby narrowing the gap between urban and rural areas.

## References

- [1] LIU Jiadong, AIHUMBATI Abuti. *Research on the dynamic impact of digital economy on urban-rural income gap* [J]. *Times Economy and Trade*, 2025(2): 177-181
- [2] XU Weixiang, SHI Rougang, ZHOU Jianping. *Mechanism and spatial effect of digital economy on urban-rural income gap* [J]. *East China Economic Management*, 2025(2): 84-94.
- [3] Sun Tiantian. *Research on the impact of digital economy on farmers' income based on principal component analysis* [J]. *Journal of Harbin Institute*, 2024(12): 36-41.
- [4] ZHAO Tao, ZHANG Zhi, LIANG Shangkun. *Digital Economy, Entrepreneurial Activity and High Quality Development - Empirical Evidence from Chinese Cities* [J]. *Management World*, 2020 (10): 65-75.
- [5] Meng Weifu, ShiKejing, Li Jianying. *Impact and mechanism of digital economy development on agricultural carbon emissions* [J/OL]. *Rural Finance Research*.
- [6] Yuan Lemon. *Research on the impact of digital economy on industrial carbon emission intensity* [J]. *Modern Industrial Economy and Informatisation*, 2024 (6): 10-16.
- [7] J. Wang, S. Zhang. *Research on the impact effect of digital economy on wages* [J/OL]. *Lanzhou Journal*.
- [8] Wang Xing, Ding Hua, Han Xuefeng. *The impact of digital economy on new quality productivity - a test based on double fixed and panel threshold models* [J/OL]. *Business and Management*.
- [9] Liu Jun, Yang Yuanjun, Zhang Sanfeng. *Research on the measurement and driving factors of China's digital economy* [J]. *Shanghai Economic Research*, 2020(6): 81-96.
- [10] Zhang Yujing, Tu Shengwei. *Digital Inclusive Finance, Agricultural Factor Allocation Efficiency and Rural-Urban Resident Income Gap-Evidence from 1572 Counties in China* [J]. *Journal of Nanjing Agricultural University (Social Science Edition)*, 2025(1):182-192
- [11] Zhang Shaohua, Qiao Bin. *Can the Unified Big Market Promote the Common Wealth of Urban and Rural Areas?-Another discussion on the moderating effect of digital economy* [J]. *World Agriculture*, 2025(3):69-81
- [12] WANG Liangjun, XIA Mengke, LIN Zihan. *Digital economy, entrepreneurial activity and urban-rural residents' income gap: evidence from Chinese cities* [J]. *Shanghai Management Science*, 2025(2): 73-79
- [13] Liang Shuanglu, Liu Peipei. *Digital inclusive finance and urban-rural income gap* [J]. *Journal of Capital University of Economics and Business*, 2019(1): 33-41
- [14] CHENG Mingwang, ZHANG Jiaping. *Internet penetration and urban-rural income gap: theory and empirical evidence* [J]. *China Rural Economy*, 2019(2):19-41
- [15] Zhang He, Bai Qinxian. *Has digital financial inclusion reduced the urban-rural income gap?* [J]. *Exploration of Economic Issues*, 2018(10):122-129