Supplementary Information:
Analysis of the Spatial–Temporal Evolution of the Digital Economy and Its Impact on the Employment Structure in China from 2001 to 2020

S1 Entropy method

Due to the wide variety of indicators involved and inconsistent units of measurement, this cannot be directly compared. Therefore, the following formula is used to treat the indicators dimensionless to eliminate the influence of indicators on the evaluation results due to the different units of measurement. The specific weight calculation procedure is as follows.

(1) Normalize the original data:

$$x_{ij}^{'} = \frac{x_{ij} - minx_{ij}}{maxx_{ij} - minx_{ij}} \text{ (S1)}$$

where $x_{ij}^{'}$ is the index value after standardization of the indicator, x_{ij} is the original index value, i represents the province, j represents the indicator, $maxx_{ij}$ and $minx_{ij}$ respectively represent the maximum value and the minimum value of a certain index value in all areas.

(2) Calculate the weighting of each indicator:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}^{i}} \text{ (S2)}$$

where *n=30* indicates the 30 provinces (municipalities and autonomous regions) selected in this paper.

(3) Calculate the information entropy for each indicator:

$$e_j = -\frac{1}{lnn} \sum_{i=1}^{n} p_{ij} lnp_{ij} \text{ (S3)}$$

where *n=30* indicates the 30 provinces (municipalities and autonomous regions) selected in this paper.

(4) Calculate the variance factor for each indicator:

$$g_j = 1 - e_j \text{ (S4)}$$

(5) Calculate the weights for each metric:

$$w_j = \frac{g_j}{\sum_{j=1}^{m} g_j} \text{ (S5)}$$

Where *m=20* indicates the 20 indicators selected in this paper to build digit1 and *m=23* indicates the 23 indicators selected in this paper to build digit2.

The weight of each indicator of China's digital economy development level from 2001 to 2020 is calculated by the traditional entropy value method described above, and the weight values of a total of 20 sets of indicators are obtained, which are recorded as $w_{j2001}, w_{j2002}, \ldots, w_{j2020}$. In order to achieve the comparability of indicators between different years, this paper draws on the treatment methods of scholars such as Zhu Facang (2019), Wang Ruyi (2018) and other scholars, and based on Newton's cooling law, the weight values of 20

groups of indicators of China's digital economy development level are time-decayed, and finally the total weight values of each index are obtained. The weight value of each indicator after time decay treatment is calculated as follows:

$$w_i = w_{j2020} * e^{-0.2*1} + w_{j2019} * e^{-0.2*2} + \ldots + w_{j2001} * e^{-0.2*20} \text{ (S6)}$$

Where w_i represents the total weight value of each indicator after time decay processing. $w_{j2001}, \ldots, w_{j2020}$ respectively represent the weight values of each indicator from 2001 to 2020, the time function weight is represented by the equation $e^{-0.2*t}$. The closer the time is to the present, the greater its weight should be. Therefore, this paper sets the time *t* value in 2020 to 1, the time *t* in 2019 to 2, and so on, the time *t* in 2001 is set to 20.

The comprehensive evaluation index of China's digital economy development level can be obtained by comprehensively weighting the values of each index of the above index system and the weights corresponding to the indicators, and the specific calculation formula is:

$$Y_i = \sum_{i=1}^{m} w_i * x_i \text{ (S7)}$$

Where Y_i is the comprehensive index of digital economy development level, w_i is the weight value of each index after time decay and normalization, and x_i is the index to measure the development level of digital economy.

Quality tests of data including NULL values test and Volume tests and Uniqueness tests, were introduced as below.

NULL values test: The Stata command “misstable summarize” was used to perform the test, which returned "variables nonmissing or string." This indicates that the data used in this study does not have any missing values. (Section S1 of Supplementary Information).

Volume tests and Uniqueness tests: The commands “unique province” and “unique year” were used to obtain statistics on the sample size in this study. It was found that there are a total of 30 provinces, 20 years, and 600 observations. The “sum year” command revealed that the sample period in this study spans from 2001 to 2020. (Section S1 of Supplementary Information).

S2 Spatial correlation analysis

Local Moran's *I* was used to determine the categories of spatial clusters for each province in order to differentiate the spatial dependence among regions. There are four categories of local spatial autocorrelation: High-High (provinces with high values of statistics clustered with neighboring provinces with similar values), Low-Low (provinces with low values of statistics clustered with neighboring provinces with similar values), High-Low, and Low-High (Anselin et al., 2009). Strong spatial correlations are indicated by the first two categories, and great spatial dispersion is indicated by the other two. Local Moran's I is calculated utilizing Equation (S8) and Local Indicators of Spatial Association (LISA) cluster maps are utilized to visualize the characteristics of each province's spatial clusters.

$$Local\ Moran's\ I_i = \frac{n^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\frac{(x_i-\bar{x})\sum_{ij}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}(x_j-\bar{x})}{\sum_{ij}w_{ij}} \quad \text{(S8)}$$

S3 Digital economy

We evaluate the development of digital economy in terms of digital basis, digital industry, digital innovation and digital efficiency. Specifically, infrastructure and popularization are used to describe digital basis. Infrastructure is measured by length of long-distance optical cable lines and capacity of mobile phone exchanges, while popularization is measured by number of mobile phone subscribers at year-end. Scale and industry inputs are used to describe the digital industry. The scale is measured by business volume of telecommunication services, average business volume of telecommunication services per capita, revenue from principal business of electronic and telecommunication equipment manufacturing, the proportion of electronic and telecommunication equipment manufacturing to GDP, pieces of express mail services and average pieces of express mail services per capita. Industry inputs is measured by annual average employees of electronic and telecommunication equipment manufacturing and number of enterprises of electronic and telecommunication equipment manufacturing. The innovation investment and output are used to describe digital innovation. Among them, the investment is calculated using full-time equivalent of R&D personnel of industrial enterprises above designated size and expenditure on R&D of industrial enterprises above designated size and the output is calculated using transaction value in technical markets, the proportion of inventions applications to the number of patent applications of industrial enterprises above designated size, number of patent application certified and

number of patent application certified for invention. Digital efficiency is measured in three aspects: agriculture, industry and services. Among them, agriculture is measured by electricity consumed in rural area and the gross domestic product on agriculture, forestry, animal Husbandry and fishery, industry is measured by new products of industrial enterprises above designated size and services is measured by total retail sales of consumer goods. The development level of China's digital economy is comprehensively evaluated by using the entropy method.

Table S1. Correlation coefficient between variables.

	lnemployment	*digit1*	*digit2*	*lnpgdp*	*lnurban*	*lngovernment*	*lnindustry*
lnemployment	1						
digit1	0.521	1					
digit2	0.528	0.997	1				
lnpgdp	0.767	0.600	0.589	1			
lnurban	0.854	0.467	0.472	0.858	1		
lngovernment	-0.125	-0.228	-0.230	0.171	0.028	1	
lnindustry	0.838	0.440	0.446	0.728	0.760	0.121	1

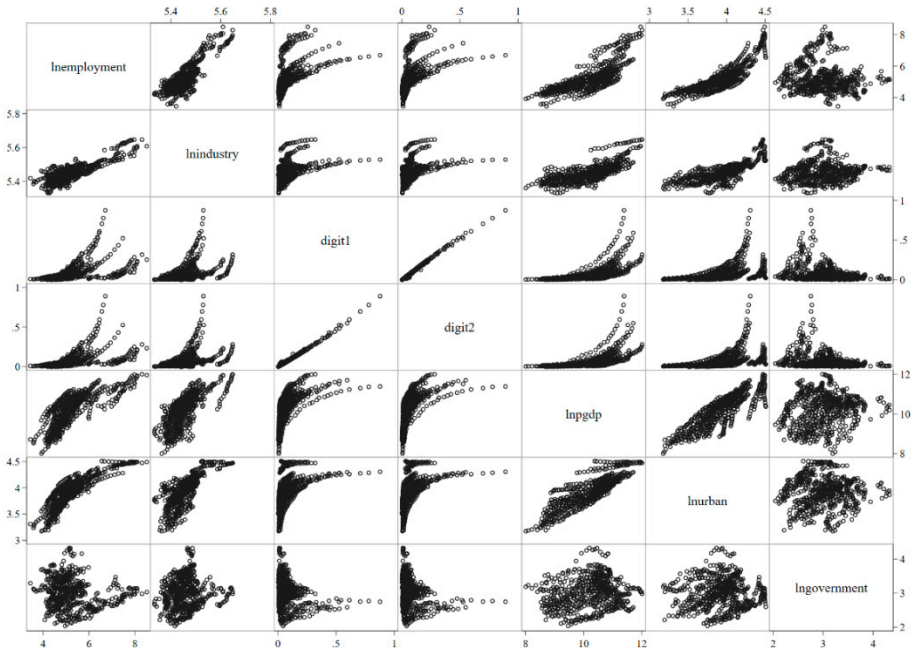

Figure S1. Correlation coefficient between variables.

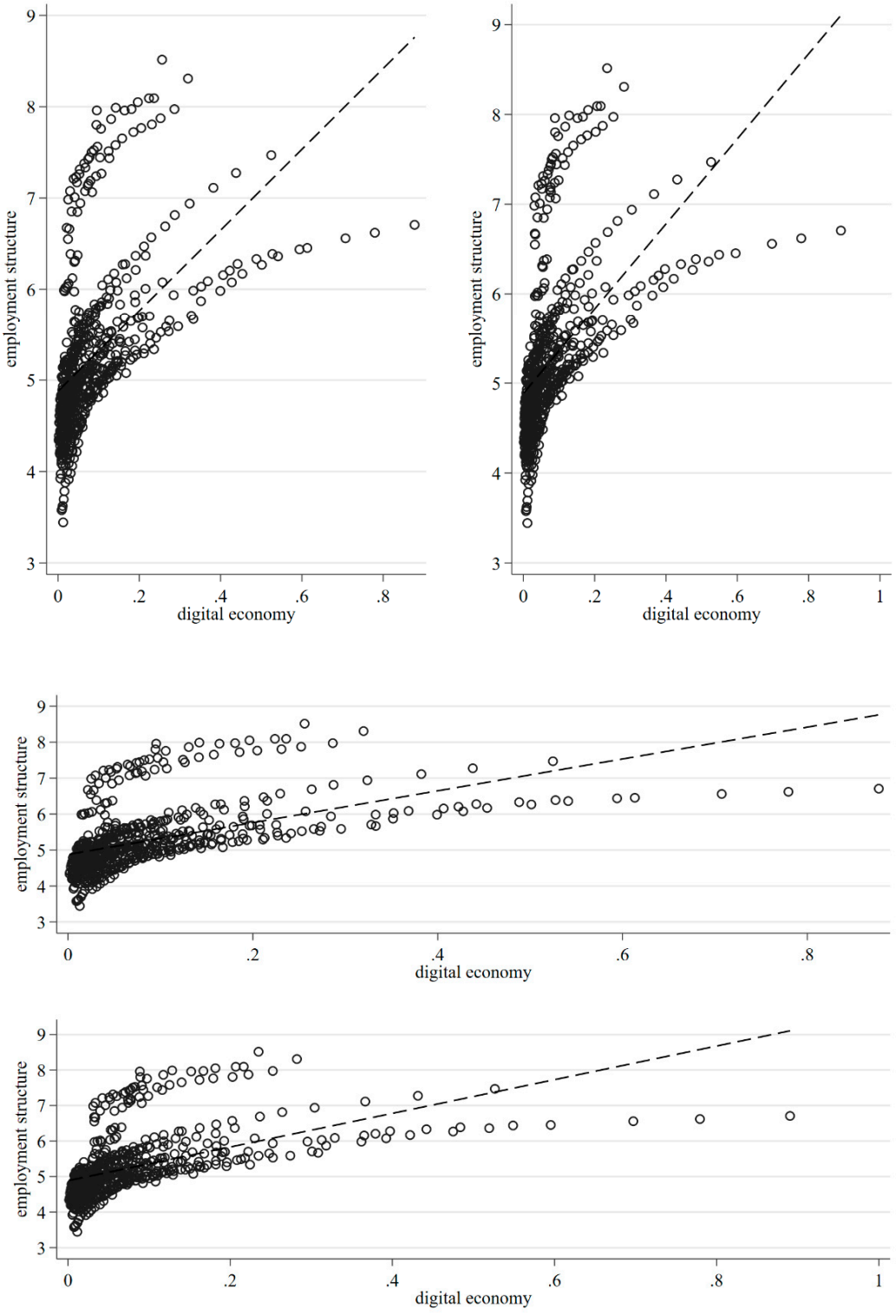

Figure S2. Digital economy and employment structure analysis diagram.