Journal of Intelligent & Fuzzy Systems 44 (2023) 2543–2560
DOI:10.3233/JIFS-222520
IOS Press


# Forecasting digital economy of China using an Adaptive Lasso and grey model optimized by particle swarm optimization algorithm

Wenmei Yu, Lina Xia and Qiang Cao*
*School of Finance, Anhui University of Finance and Economics, Bengbu, Anhui, China*

**Abstract**. With the development of big data, Internet finance, the digital economy is developing rapidly and has become an important force to drive the continuous transformation of the global economy and society. China has put forward plans for the development of digital economy from 2021 to 2025, requiring the number of core industries of digital economy to reach 10% of GDP by 2025, while continuously improving China's digital economy to achieve high-quality development of China's digital economy. Aiming at China's digital economy, we use the adaptive lasso method and select feature variables based on quantitative and qualitative perspectives, so as to predict the development trend of China's digital economy from 2021 to 2025 based on the TDGM (1, 1, r) grey model optimized by the particle swarm algorithm. Meanwhile, we have added the comparative analyses with TDGM(1,1), Grey Verhulst, GM(1,1) models and evaluate the prediction results both Ex-ante and Ex-post, demonstrating the feasibility of the proposed model and the accuracy. Finally, we find that the future of China's digital economy will meet the planned objectives in terms of quantity and quality, but the trend of digital economy development in quantity is faster, thanks to the development of digital technology application industry.

Keywords: Digital economy development, adaptive lasso grey model, TDGM(1, 1, r) model, quantity and quality

## 1. Introduction

Recently, digital technologies such as the Internet, big data, cloud computing, artificial intelligence and blockchain have accelerated innovation and increasingly integrated into the whole process of economic and social development in all fields. The digital economy has been on a rise in the global economic development and is becoming a key force in reorganizing global factor resources, reshaping the global economic structure and changing the global competitive landscape. Compared with developed countries such as the United States, China developed its digital economy in a relatively later period, but with the advancement of Internet technology and increasing demand, the scale (quantity) of China's digital economy has grown from 11 trillion yuan in 2012 to 45 trillion yuan in 2021, ranking the second in the world. There is no doubt that the boom in China's digital economy has become an important engine for promoting high-quality economic and social development. However, China still has such problems as insufficient technological accumulation and few laws, regulations and policies to encourage innovation, resulting in a lack of innovation in key areas. This means that China's digital economy is more

*Corresponding author. Qiang Cao, School of Finance, Anhui University of Finance and Economics, Bengbu, Anhui, China. E-mail: caoqiangsh@163.com.

inclined to quantitative growth and imitative follow-up, but less likely to prevail in quality.

In December 2021, China proposed a plan for the development of the digital economy from 2021 to 2025 in terms of both quantity and quality. On the one hand, the plan sets a goal that by 2025, the added value of the core industries of the digital economy will reach 10% of the GDP, with the market system of data elements established and the digital transformation of industries reaching a new level. On the other hand, China will push forward with upgrading the digital efficiency industry and promoting the development of digital economy in terms of quality. The digital economy should interact with the factor chain, industrial chain, value chain and institutional chain, to realize the high-quality development of China's digital economy. Therefore, this paper attempts to predict the development pattern of China's digital economy from quantitative and qualitative perspectives from 2021 to 2025, select the relevant feature variables, and propose policy recommendations for achieving the development goals of China's digital economy and promoting the collaborative quantitative and qualitative development of the digital economy.

## 2. Literature review

The impact of the digital economy in China is enormous. It takes different forms and plays an important role in supporting the national economy and social development. For example, the digital economy has led to the development of tourism [28], enabled the high-quality development of the manufacturing industry [7] and promoted low-carbon development [34]. Similarly, the factors affecting the development of the digital economy are multifaceted and mainly include five aspects: information infrastructure, technological innovation, economic development, social support and demographic characteristics. [24] found that improving information infrastructure development is beneficial to the development of digital economy. [39] argued that technological innovation and increased R&D expenditures by firms would promote the development of digital economy [5, 41] considered that financial development, economic growth, urbanization, and industrial structure upgrading have a positive impact on the digital economy. [12, 26, 37] found that government spending, Internet development, and digital trade have become the main drivers of digital economy development. [20, 41] also argued that income differences and human capital (level of education of the population) affect it.

A broad range of the determinants also make it difficult for us to really grasp the key characteristics of the development of digital economy. [22] suggested that if there exist too many influencing factors, the feature variables will be made obscure, leading to model overfitting. Consequently, several approaches have emerged for selecting factors influencing the digital economy. The method used for early shrinkage variables is ARMA, where the optimal lag order is determined by the AIC and BIC criteria [1]. However, in the era of big data, the dimensionality of variables increases, so more and more academics are turning to machine learning algorithms. Currently, three algorithms can be implemented for variable selection. The first is the decision tree algorithm [2]. However, as a non-parametric method of selecting variables without correlation between feature dimensions, decision tree is too concise and thus weak in selecting variable features [35]. The second is ridge regression [11]. Although it can compress the original variable coefficients to a certain extent, ridge regression cannot compress them to 0, so the final model has to retain all the variables [6]. The lasso model based on penalized regression provides a good solution to the weaknesses of the above two methods [25]. But the lasso algorithm corresponds to the selection bias when calculating highly correlated variables [4, 44]. Therefore, we instead use the adaptive lasso regression, where weights need to be added to automatically adjust for penalty factors [23].

Undoubtedly, accurately predicting the future dynamics of the digital economy can help China achieve its digital economy development planning goals for 2021 to 2025. There are many forecasting methods currently used by scholars. Firstly, the main forecasting methods for large sample data are regression analysis forecasting models and machine learning forecasting models. For example, [33] constructed six different traditional regression models to predict e-commerce sales in EU countries. [42] predicted digital economy consumption psychology by means of a canopy clustering algorithm. [17] used a BP neural network forecasting model and a neural network model optimized by a particle swarm algorithm (PSO-BP) to forecast the Chinese stock market. Second, grey forecasting models are often used in small sample data forecasting, and [9, 18] both employed the traditional GM(1, 1) grey forecasting model to predict US futures returns and bitcoin prices, respectively. To improve the forecasting effect, some scholars also optimize the grey models. [14] used the optimized VOFDGM(1,1) gray model to predict the electricity consumption in China. [30] used the HTS_UGM(1,1) grey model to forecast the gold price demand problem in Sri Lanka. [3, 19, 32] all used the optimized EGM (1,1,$\alpha$, $\theta$) grey model to forecast the number of tourists, trade deficit and palm oil exports from Indonesia, respectively.

Apparently, digital economic development forecasts are typically small-sample based and relate to grey forecasting models. However, the traditional GM(1,1) grey prediction model is difficult to satisfy the criteria of non-homogeneous exponential series [29, 38]. Due to its inherent structural defects. [40] proposed the three-parameter discrete grey forecasting model TDGM(1,1), which added a linear correction term to GM(1,1). [10, 31] found that changing the cumulative order of the TDGM(1,1) model not only preserves the ability to make predictions for non-homogeneous exponential growth series, but also improves the prediction performance. In addition, [8, 16] proposed that the particle swarm algorithm (PSO) can optimize the gray model parameters and obtain more accurate prediction results. Based on the above studies, we consider using the TDGM(1,1,r) model optimized by particle swarm algorithm to predict the future digital economy development in China. However, the TDGM(1,1,r) model is a single variable grey prediction system, and to ensure the stability of this model, a model of the factors affecting digital economic development is needed, and if the results of modeling each feature variable in this model using grey prediction are stable, then the prediction of the dependent variable (digital economic development) is viable [13].

These previous studies investigated future digital economy development from different perspectives, but there is still room for improvement. First, the existing literature seldom selects complex influencing factors of digital economy development in multivariate forecasts, which leads to inaccurate estimates or large errors in the results. Second, China's digital economic development is basically a small sample forecast and machine learning models such as neural networks are not applicable. Third, in predicting China's future digital economy, most researchers have considered the overall development but ignored the quantitative and qualitative imbalances in China's digital economy development.

Hence, the study attempts to make the following contributions. First, an adaptive lasso model based on the weight function to adjust the penalty factor is introduced. As a shrinkage model, the influencing factors of digital economy development are selected and the future development trend of the feature variables is predicted to ensure the reliability of the predicted digital economy. Second, we predict the quantity and quality of digital economy development and its characteristic variables using the TDGM(1,1,r) model based on particle swarm algorithm optimization and compare the results with those f TDGM(1,1), Grey Verhulst, and GM(1,1) models. Third, we conduct a comparative study of digital economy development from both quantitative and qualitative perspectives, analyze different the determining factors and predicts the development trend from 2021 to 2025.

The rest of the paper is organized as below. After the introduction and literature review, Section 3 details the data and methodological steps of the study, Section 4 presents the results, and Section 5 draws conclusions and policy recommendations.

## 3. Data description and methodological steps

### 3.1. Data description

This section provides a characterization of the data. According to the Statistical Classification of the Digital Economy and Its Core Industries (2021) issued by the National Bureau of Statistics of China, this study determines the quantity and quality of digital economy development in terms of "digital industrialization" and "digitalization of industry", including five major categories: digital product manufacturing, digital product service, digital technology application, digital fac-tor-driven industry, and digital efficiency improvement industry. The quantity of digital economy development corresponds to the "digital industrialization" section (the first four categories), reflecting the current development of the core industries of digital economy. The quality of digital economy development corresponds to the digital efficiency improvement industry in the "Indus-try Digitization" section, reflecting the deep penetration and extensive integration of digital economy with various industries of national economy, mainly covering the application scenarios of digital finance.

This study focuses on the development of China's digital economy from the perspective of quantity and quality. First, the value added of the core industry scale of the digital economy measures the quantity of digital economy development [43]. We construct a digital economy measurement framework based on Chinese economic census data by referring [36] and calculate the value added of the core industry scale of China's digital economy[1]. Second, the quality of digital economy development is measured by the degree of digital financial development [21, 27], with data from the Digital Financial Inclusion Index compiled by the Internet Finance Research Center of Peking University[2]. Based on the data provided by Ant Financial, 33 indicators were selected to construct the "Digital Finance inclusion Index (2012–2020)", reflecting the actual situation in China across three dimensions, namely coverage, depth of use and degree of digital support services based on the establishment method of financial inclusion indicators in the existing literature.

Following [39], we select the data of feature variables affecting the quantitative and qualitative development of the digital economy from 2012 to 2020, such as the number of internet broad-band access ports(iap), the number of internet broadband access subscribers(iau), the number of domain names(ndn), total telecommunications business(tts),the technology development lev-el(tec), r&d expenditure(rdf), the number of traditional financial development(fin), economic development level(agdp), industrial structure upgrading(tin),urbanization level(url), government financial spending(gov), internet development level(net), the degree of trade openness(tra), education level(edu), per capita net income level(icl). The specific variables are described in Table 1. In terms of data processing, this paper uses matlab2018a software to take logarithms of all variables when evaluating the model in order to ensure the robustness of the prediction model.

We obtain descriptive statistics of the variable data using R software (version 4.1.1) and report the minimum, maximum, mean, median and standard deviation for each variable, as shown in Table 2. Combined with the raw data and the descriptive statistics in Table 2, it can be found that both the quantity and quality of digital economy have developed better during the sample period, but the quantity of digital economy development is significantly faster than the quality. The standard deviations of the explanatory variables dig and din are as high as 15436.66 and 80.97, indicating that there is a great difference between the data of each year.

[1]http://www.stats.gov.cn/xxgk/tjbz/gjtjbz/202106/t20210603_1818135.htmls.

[2] https://idf.pku.edu.cn/docs/20210421101507614920.pdf.

In addition, when analyzing the feature variables of digital economy development, it can be seen that the feature variables such as iap, iau, ndn, tts, tec, rdf, agdp, tin, url, and icl have been increasing steadily. Meanwhile, in the 9 years from 2012 to 2020, the growth rate is faster, and this shows that these feature variables have a positive and fast development trend during the sample period. Linking the variables Fin, gov, and net, which indicates a fluctuating upward trend and high volatility from 2021 to 2025. However, tra shows a significant fluctuating downward trend over the sample period, and this indicates that China is facing greater uncertainty and possible geopolitical risks in terms of trade opening since 2012.

### 3.2. Methodological steps

In order to accurately predict the quantity and quality of China's future digital economy development, we use the adaptive lasso- TDGM(1,1,r) combination model and compare it with three grey forecasting models. The framework diagram of the empirical research part is detailed in Fig. 1.

#### 3.2.1. Adaptive lasso model

The lasso model is a compressive estimation method that improves the interpretability of the model by adding penalty terms to the generalized linear model, limiting the sum of the absolute values of the parameters to be estimated to a very small threshold, eliminating some characteristic variables with estimated coefficients of zero, and avoiding model overfitting. However, Zou (2006) argued that the Lasso model would have a large bias and inconsistency in the selection of variables, so an improved adaptive Lasso model was proposed, as detailed in Equation (1):

$$\hat{\beta}_{\text{alasso}} = \arg\min_{\beta} \left( \|Y - X\beta\|^2 + \lambda \sum_{j=1}^{p} \widehat{\omega}_j \left|\beta_j\right| \right) \tag{1}$$

Where $\widehat{\omega}_j = 1/\beta_j (j = 1, 2, \cdots, p)$ is the weighting function, $\lambda$ is the penalty parameter, which is the strength of the penalty on the characteristic variables.

Optimization of the base model, let $x_j^* = x_j/\widehat{\omega}_j (j = 1, 2, \cdots, p)$, $\beta^* = \widehat{\omega}\beta$, so

$$\beta_{k+1}^* = \arg\min_{\beta^*} \left\{ f(\beta^*) + \lambda g(\beta^*) \right\} \tag{2}$$

Table 1
Variable definitions

| Indicator | Description | classification |
|---|---|---|
| din | Quantity of digital economy development | Digital Economy |
| dig | Quality of digital economy development | Digital Economy |
| iap | Number of Internet broadband access ports | Infrastructure Factors |
| iau | Number of Internet broadband access subscribers | Infrastructure Factors |
| ndn | Number of domain names per 1,000 people | Infrastructure Factors |
| tts | Total Telecommunications Business | Infrastructure Factors |
| tec | Technology Development Level | Technical Factors |
| rdf | R & D expenditure | Technical Factors |
| fin | Degree of traditional financial development | Economic Factors |
| agdp | Economic Development Level | Economic Factors |
| tin | Industrial structure upgrading | Economic Factors |
| url | Urbanization level | Economic Factors |
| gov | Government Financial Spending | Social Factors |
| net | Internet Development Level | Social Factors |
| tra | Degree of trade openness | Social Factors |
| edu | Education level | Population feature factors |
| icl | Per capita net income level | Population feature factors |

Table 2
Descriptive statistics

| | Min | Max | Mean | Median | Std |
|---|---|---|---|---|---|
| din | 35825.4 | 79637.9 | 55240.6 | 52351.1 | 15436.7 |
| dig | 61.5 | 431.9 | 236.9 | 238.5 | 81.0 |
| iap | 105.6 | 8653.2 | 2174.3 | 1751.5 | 1721.4 |
| iau | 49.9 | 3890.0 | 1039.1 | 774.4 | 843.5 |
| ndn | 0.9 | 294.2 | 24.0 | 11.9 | 40.7 |
| tts | 54.5 | 15025.3 | 1568.7 | 757.1 | 2012.1 |
| tec | 844.0 | 967204.0 | 107308.5 | 55609.0 | 148326.8 |
| rdf | 65029.0 | 25000000.0 | 3705447.0 | 2300000.0 | 4756503.0 |
| fin | 1.8 | 7.6 | 3.4 | 3.2 | 1.1 |
| agdp | 18946.9 | 164158.0 | 55714.0 | 47928.0 | 27267.6 |
| tin | 0.6 | 5.2 | 1.4 | 1.2 | 0.7 |
| url | 0.4 | 0.9 | 0.6 | 0.6 | 0.1 |
| gov | 0.1 | 0.8 | 0.3 | 0.2 | 0.1 |
| net | 57.3 | 189.5 | 101.8 | 98.6 | 24.2 |
| tra | 0.0 | 1.3 | 0.3 | 0.1 | 0.3 |
| edu | 7.5 | 12.8 | 9.3 | 9.2 | 0.9 |
| icl | 9768.0 | 72232.4 | 24480.6 | 21913.4 | 11186.0 |

Where L is a constant, $f(\beta^*) = \left\| y - \sum_{j=1}^{p} x_j^* \beta^* \right\|^2$, $g(\beta^*) = \sum_{j=1}^{p} \left|\beta_j^*\right| = \left\|\beta_j^*\right\|_1$, $z = \beta_k^* - \frac{1}{L}\nabla f(\beta_k^*)$.

Let $\frac{\partial F(x)}{\partial \beta_j^*} = 0$, compute $\beta_j^* = \mathrm{sgn}(z_j) \cdot \max(\left|z_j\right| - \frac{\lambda}{L}, 0)$

Finally, according to Equation (3), an adaptive lasso model is built by R software (version 4.1.1) to select variables for the quantity and quality of China's digital economy development.

$$\widehat{\beta}_j^* = \beta^*/\widehat{\omega}_j,\ j = 1, 2, \cdots, p \tag{3}$$

### 3.2.2. *Ex-ante predictive model evaluation*

To perform ex-ante performance measures of time series forecasting models, we draw on [15] to determine the suitability of the selected model for forecasting by means of the Posterior-Variance Test (PVT). The PVT method includes two metrics, the ratio of root-mean-square deviations (C) and the small-error probability (P), as shown in Equations (4 and 5). In general, with smaller C and larger P, highly accurate predictions can be obtained. The evaluation scale of PVT method is shown in Table 3. If the PVT passes, then the TDGM(1,1,r) grey prediction can be applied, otherwise it means that the model is not suitable for predicting the given data.

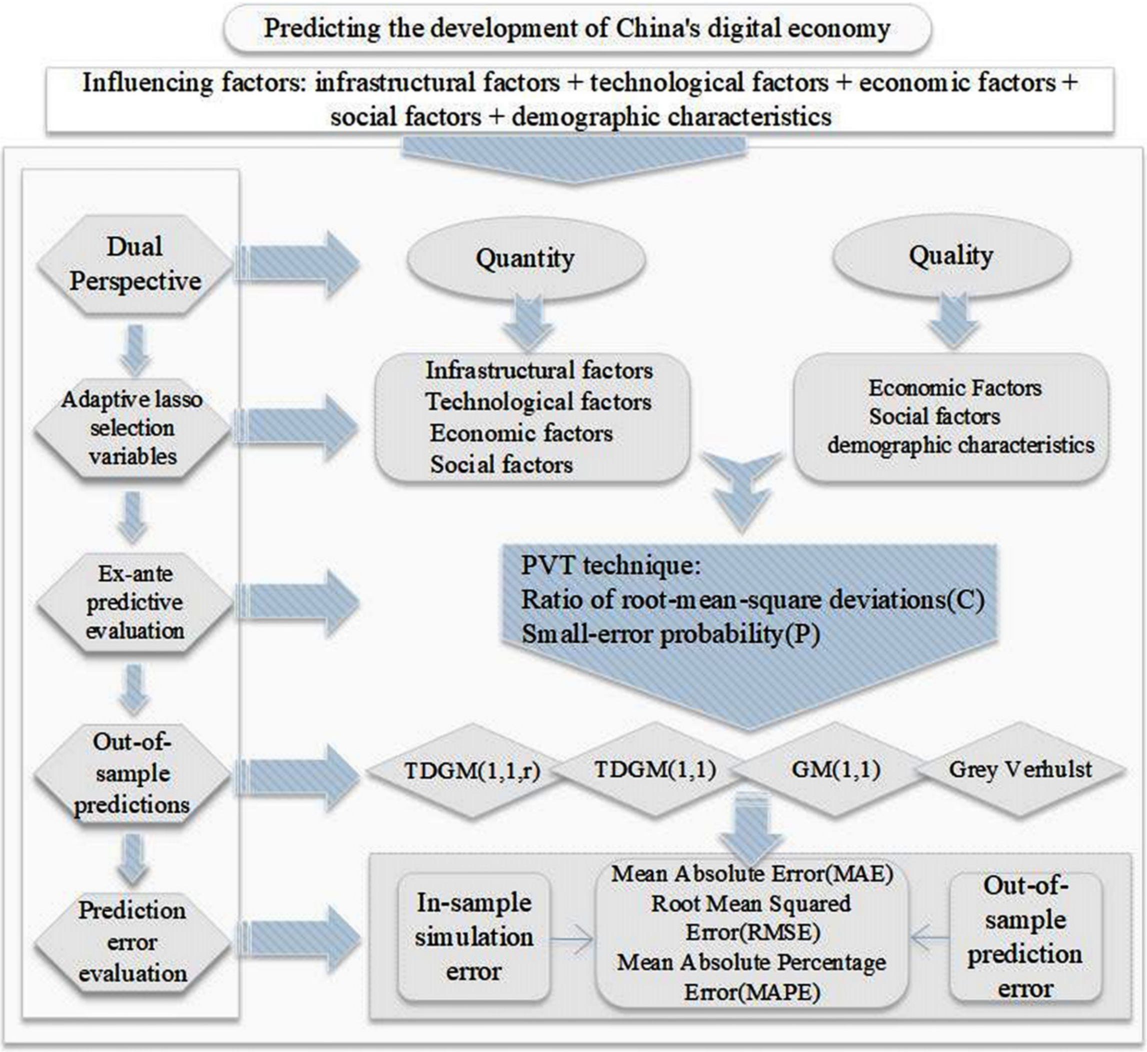


Fig. 1. The framework diagram of the empirical research part.

Table 3
The evaluation scale of PVT method

| Forecast Accuracy | P | C |
|---|---|---|
| Good | >0.95 | <0.35 |
| Qualified | 0.80∼0.95 | 0.35∼0.50 |
| Barely Qualified | 0.70∼0.80 | 0.50∼0.65 |
| Unqualified | ≤0.7 | ≥0.65 |

$$C = \frac{S_2}{S_1} = \frac{\sqrt{\frac{1}{n-1}\sum_{k=2}^{n}[\varepsilon^{(0)}(k) - \bar{E}]^2}}{\sqrt{\frac{1}{n}\sum_{k=1}^{n}[x^{(0)}(k) - \bar{X}]^2}} \tag{4}$$

$$p = P\left\{\left|\varepsilon(k) - \bar{E}\right| < 0.6745\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[x^{(0)}(k) - \bar{X}\right]^2}\right\} \tag{5}$$

Where $\varepsilon^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$, $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} x^{(0)}(k)$, $\bar{E} = \frac{1}{n-1}\sum_{k=2}^{n}\varepsilon^{(0)}(k)$, $k = 2, 3, \cdots, n$.

### 3.2.3. TDGM(1,1,r) grey prediction model

The grey model is a method of forecasting based on small sample data, which does not require a large number of regularly distributed samples in prediction and has high prediction accuracy. The GM(1,1) model is a univariate classical grey prediction model, which is better for modeling sequences with approximately flush exponential growth patterns for prediction. However, most of the real-world system behaviors exhibit approximate non-simultaneous exponential growth characteristics, and satisfactory prediction results will not be obtained if the GM(1,1) model is used to predict them. Therefore, we use the three-parameter discrete grey forecasting model TDGM(1,1), which can forecast non-homogeneous exponential growth series,, and optimize the cumulative order of the TDGM(1,1) model using a particle swarm algorithm to further improve the forecasting ability of the model.

**First,** the initial values of the parameters are obtained. Randomly initialize the position and velocity of the particles in the particle swarm, taking $pBest = 1$, i.e. the mean TDGM(1,1) model. Set $pBest$ in the particles to the current position and $gBest$ to the position of the best particle in the initial population.

If the original data sequence is $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n))$

And $X^{(0)}$ by the r-order accumulation generating operator (r-AGO) can generate the sequence:

$$x^{(r)}(k) = \sum_{i=1}^{k} \frac{\Gamma(r+k-i)}{\Gamma(k-i+1)} x^{(0)}(i), k = 1, 2, \cdots, n. \tag{6}$$

$X^{(0)}$ by the r-order inverse accumulation generating operator (r-IAGO) can generate the sequence:

$$x^{(r-1)}(k) = x^{(r)}(k) - x^{(r)}(k-1) \tag{7}$$

Thus, the Nearest neighbor mean generation sequence of $X^{(r)}$ is

$$Z^{(r)}(k) = \frac{x^{(r)}(k) + x^{(r)}(k-1)}{2}, k = 2, 3, \cdots, n. \tag{8}$$

Substituting $x^{(r-1)}(k)$, $z^{(r)}(k)$ into the TDGM(1,1,r) base model:

$$x^{(r-1)}(k) + az^{(r)}(k) = kb + c$$

And the parameters to be estimated are calculated as

$$\hat{p} = (a, b, c)^T = (B^T B)^{-1} B^T Y. \tag{9}$$

The time response function of TDGM(1,1,r) is $\hat{x}^{(0)}(k)$, from which the simulated and predicted values of the characteristic variable $\hat{X}^{(0)}$ are obtained.

$$\hat{X}^{(0)}(k) = \sum_{i=0}^{k-1} (-1)^i \frac{\Gamma(r+1)}{\Gamma(i+1)\Gamma(r-i+1)} \hat{X}^{(r)}(k-i) \tag{10}$$

Where $\hat{X}^{(r)}(k) = X^{(r)}(1) \cdot \alpha^{k-1} + \sum_{g=0}^{k-2} \left[(k-g) \cdot \beta + \gamma\right] \cdot \alpha^g$, $\alpha = \frac{1-0.5a}{1+0.5a}; \beta = \frac{b}{1+0.5a}; \gamma = \frac{c}{1+0.5a}$

**Second,** based on the above predicted and original values, calculate the average relative simulation error $f(pBest)$ and determine whether $|f(pBest) - f(gBest)|$ is less than the given convergence value $\delta$. If it is satisfied, the optimal value of $gBest$ for the cumulative order r and the simulation and prediction data of the TDGM(1,1,r) model are output and the algorithm run is finished. Otherwise, the position and velocity of the particle are updated. If the particle fitness is better than the fitness of $pBest$, $pBest$ is set as the new position; if the particle fitness is better than the fitness of $gBest$, $gBest$ is set as the new position.

$$\begin{aligned} V = \omega \times V + c_1 \times rand \times (pBest - \Pr esent) \\ + c_2 \times rand \times (gBest - \Pr esent) \end{aligned} \tag{11}$$

Where $\Pr esent = \Pr esent + V$, $\omega = \omega_{\max} - run \times \frac{(\omega_{\max} - \omega_{\min})}{runMax}$

Calculate the population fitness variance $\sigma^2$ and the probability of variation $p_m$, and calculate $f(pBest)$.

$$\sigma^2 = \sum_{i=1}^{n} \left( \frac{f_i - f_{avg}}{f} \right)^2 \tag{12}$$

$$p_m = \begin{cases} k, \sigma^2 < \sigma_d^2 and f(gBest) > f_d, \\ 0, other \end{cases} \tag{13}$$

Where $f = \begin{cases} \max\left\{\left|f_i - f_{avg}\right|\right\}, \max\left\{\left|f_i - f_{avg}\right|\right\} > 1, \\ 1, other \end{cases}$

**Finally,** the random number $\varepsilon \in [0, 1]$ is generated and if $\varepsilon < p_m$, the variation operation is performed according to

$$gBest_k = gBest_k \times (1 + 0.5 \times \eta) \tag{14}$$

If $\varepsilon \geqslant p_m$, the algorithm convergence criterion is judged to be satisfied. If it is satisfied, the optimal value $gBest$ of the cumulative order r and the simulation and prediction data of this TDGM(1,1,r) model are output and the algorithm run is finished, otherwise the run is repeated until the algorithm convergence criterion is satisfied.

*3.2.4. Ex-post forecast accuracy indicators*

TDGM(1,1,r) is a regression model, and we use statistical measures to assess the accuracy of this model's predictions. In this paper, we choose three indicators, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean absolute percentage error (MAPE), to evaluate the accuracy of the model's prediction results. They are all used to measure the deviation between the predicted series and the original series, which can better reflect the actual situation of the error between the predicted and true values. The calculation formulae are detailed in (15), (16) and (17). In general, the smaller the prediction error, the better the TDGM(1,1,r) grey prediction model forecasts.

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|\varepsilon^{(0)}(k)\right| \tag{15}$$

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{k=1}^{\mathrm{n}}\left|\varepsilon^{(0)}(k)\right|^2} \tag{16}$$

$$MAPE = \frac{100\%}{\mathrm{n}}\sum_{i=1}^{n}\left|\frac{\varepsilon^{(0)}(k)}{\bar{X}}\right| \tag{17}$$

## 4. Empirical results

### 4.1. Adaptive-Lasso model for selecting critical feature variables

There are more factors affecting the development of digital economy with higher dimensionality, resulting in more complex feature variables. Therefore, this paper uses the adaptive Lasso model to select the key feature variables of the quantity and quality of digital economy development to achieve the effect of shrinkage variables. The results of the adaptive Lasso model to estimate the coefficients of all the feature variables are detailed in Table 5.

It can be seen from Table 5 that the quantity and quality of digital economy development have different feature variables, which indicates that different factors influence the development of the digital economy in quantitative and qualitative terms in the future. On the one hand, in terms of the quantity of digital economy development, iap, ndn, tts, tec, agdp, tin, url, net, tra, edu, and icl are eliminated because their estimated coefficients are all 0, indicating that the correlation between these variables and the quantity of digital economic development is weak, and they are not considered as influencing factors for the quantity of digital economic development. On the other hand, in terms of the quality of digital economy development, the coefficients of iap, iau, ndn, tts, tec, rdf and icl are also zero, so these variables are not considered as influencing factors for the quality of digital economy development.

In order to verify the validity and rationality of the combined adaptive lasso and TDGM(1,1,r) models, we conduct the comparative tests for four grey models and evaluate the quantitative and qualitative development of the digital economy from in-sample and out-of-sample respectively through three indicators of Ex-post predictive evaluation. And from Table 6, it is evident that all four models can predict the quantity and quality of digital economic development with accuracy.

However, it is worth noting that both for in-sample and out-of-sample forecasting, the estimates of the prediction evaluation of TDGM(1,1,r) are smaller than those of the other three models. This not only suggests that the feature variables selected by the adaptive lasso model represent the relationship with the quantity and quality of digital economic development, but also shows that the TDGM(1,1,r) model based on particle swarm algorithm optimization is reliable.

### 4.2. Out-of-sample prediction of critical feature variables

According to the estimation results of the adaptive lasso selection variables in Table 5, four critical feature variables affecting the quantity of digital economy development and eight critical feature variables affecting the quality of digital economy development are selected in this paper. Considering that the quantity and quality of digital economy development are small sample forecasts, the most appropriate forecasting method is to use the grey model and make out-of-sample forecasts for them to obtain the forecast values from 2021 to 2025. The observation of the original data revealed that the feature variables such as fin, gov, net, edu, and tra have a non-homogeneous exponential growth pattern and are not applicable to the traditional GM(1,1) grey prediction model. Therefore, this paper uses the TDGM(1,1,r) grey model based on the improved particle swarm algorithm to make out-of-sample predictions of digital economic development. Therefore, this paper uses the TDGM(1,1,r) grey model based on the improved particle swarm algorithm for out-of-sample prediction of the critical feature variables of digital economy development in terms of quantity

Table 4
The steps of the empirical framework

Algorithm: Pseudocode for TDGM(1,1,r) empirical framework

1: Variable processing and inputting them into the program.
**Adaptive lasso: (R software, version 4.1.1)**
2: Initialization parameter β.
3: Optimizing the base eq. (1) and finally building an adaptive lasso model based on eq. (3).
**Ex-ante predictive model evaluation: (matlab2018a)**
4: Perform the PVT according to eq. (4) and eq. (5). If the test passes, move to step 5, otherwise it indicates that the model is not suitable for predicting the given data.
**TDGM(1,1,r): (matlab2018a)**
5: Random initialization of the positions and velocities of the particles in the particle swarm.
6: Calculate the parameters to be estimated a, b, c by means of eq. (9).
7: Calculate the predicted value of the initial particle and the average relative simulation error through eq. (10).
8: Determine whether the average relative simulation error is less than the convergence value, if it is satisfied then move to step 13, otherwise move to step 9.
9: Update the position and velocity of the particle by Eq. (11).
10: Calculate the variance of population fitness and the probability of variation by Eqs. (12) and (13).
11: Determine whether to perform the mutation operation by Eq. (14), if not turn to step 13.
12: Determine whether the algorithm convergence criterion is satisfied. If so, perform step 13, otherwise move to step 6.
**Ex-post forecast accuracy indicators: (matlab2018a)**
13: In-sample simulation and out-of-sample prediction accuracy were evaluated by MAE, RMSE and MAPE.
14: Outputs simulation and prediction data for the optimal cumulative order r, TDGM(1,1,r) model and prediction evaluation results.
**TDGM(1,1), GM(1,1) and Grey Verhulst: (matlab2018a)**
15: Compare TDGM(1,1,r) prediction results with TDGM(1,1), GM(1,1) and Grey Verhulst.

Table 5
Adaptive Lasso model estimated coefficients

| Feature Variables | iap | iau | ndn | tts | Tec | rdf | fin | agdp |
|---|---|---|---|---|---|---|---|---|
| din | 0.000 | 1.812 | 0.000 | 0.000 | 0.000 | 0.193 | 0.285 | 0.000 |
| dig | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | –0.085 | 0.044 |
| Feature Variables | tin | url | gov | Net | Tra | edu | icl | |
| din | 0.000 | 0.000 | –0.125 | 0.000 | 0.000 | 0.000 | 0.000 | |
| dig | 0.447 | 2.293 | –0.630 | 0.004 | –0.715 | –0.229 | 0.000 | |

Table 6
Accuracy evaluation results of predicting digital economy development

| din | MAE1 | MAE2 | RMSE1 | RMSE2 | MAPE1 | MAPE2 |
|---|---|---|---|---|---|---|
| TDGM(1,1, r) | 0.006 | 0.009 | 0.007 | 0.012 | 0.064 | 0.078 |
| TDGM(1,1) | 0.009 | 0.015 | 0.01 | 0.022 | 0.092 | 0.136 |
| GM(1,1) | 0.01 | 0.013 | 0.011 | 0.016 | 0.105 | 0.119 |
| Grey Verhulst | 0.007 | 0.034 | 0.009 | 0.034 | 0.068 | 0.301 |
| dig | MAE1 | MAE2 | RMSE1 | RMSE2 | MAPE1 | MAPE2 |
| TDGM(1,1, r) | 0.016 | 0.004 | 0.024 | 0.006 | 0.348 | 0.073 |
| TDGM(1,1) | 0.021 | 0.102 | 0.026 | 0.11 | 0.462 | 1.76 |
| GM(1,1) | 0.023 | 0.131 | 0.03 | 0.139 | 0.493 | 2.252 |
| Grey Verhulst | 0.038 | 0.058 | 0.053 | 0.058 | 0.803 | 1.002 |

Suffix 1: in-sample (2012–2018); Suffix 2: out-of-sample (2019-2020).

and quality. Besides, in order to ensure the reliability of the combined model, three frequently used grey models, TDGM(1,1), GM(1,1), and Grey Verhulst, are chosen for comparison experiments in this paper.

Before making predictions, we drew on the Posterior-Variance Test (PVT) technique used by [15] to determine whether these models were suitable for predicting these variables, the results of which are shown in Table 7. Comparing the empirical results in

Table 7
The results of the posterior-variance test

| | TDGM(1,1,r) | | TDGM(1,1) | | GM(1,1) | | Grey Verhulst | |
|---|---|---|---|---|---|---|---|---|
| | C | *P* | C | *P* | C | *P* | C | *P* |
| Digital Economy Development. | | | | | | | | |
| Din | 0.032 | 1.0 | 0.051 | 1.0 | 0.047 | 1.0 | 0.051 | 1.0 |
| Dig | 0.058 | 1.0 | 0.138 | 1.0 | 0.172 | 1.0 | 0.143 | 1.0 |
| Feature variables of the quantity of digital economy development. | | | | | | | | |
| iau | 0.040 | 1.0 | 0.203 | 1.0 | 0.184 | 1.0 | 0.141 | 1.0 |
| rdf | 0.027 | 1.0 | 0.027 | 1.0 | 0.050 | 1.0 | 0.063 | 1.0 |
| fin | 0.399 | 1.0 | 0.404 | 1.0 | 0.459 | 0.8 | 0.372 | 1.0 |
| gov | 0.376 | 1.0 | 0.893 | 0.4 | 0.986 | 0.3 | 0.956 | 0.6 |
| Feature variables of the quality of digital economy development. | | | | | | | | |
| fin | 0.3989 | 1.0 | 0.404 | 1.0 | 0.459 | 0.8 | 0.372 | 1.0 |
| agdp | 0.1121 | 1.0 | 0.112 | 1.0 | 0.100 | 1.0 | 0.107 | 1.0 |
| tin | 0.0684 | 1.0 | 0.250 | 1.0 | 0.251 | 1.0 | 0.103 | 1.0 |
| url | 0.0349 | 1.0 | 0.073 | 1.0 | 0.071 | 1.0 | 0.060 | 1.0 |
| gov | 0.3759 | 1.0 | 0.893 | 0.4 | 0.986 | 0.3 | 0.956 | 0.6 |
| net | 0.2304 | 1.0 | 0.272 | 1.0 | 0.258 | 1.0 | 0.276 | 1.0 |
| tra | 0.2888 | 1.0 | 0.491 | 0.9 | 0.522 | 0.8 | 0.388 | 0.9 |
| edu | 0.3927 | 1.0 | 0.580 | 0.9 | 0.646 | 0.9 | 0.662 | 0.9 |

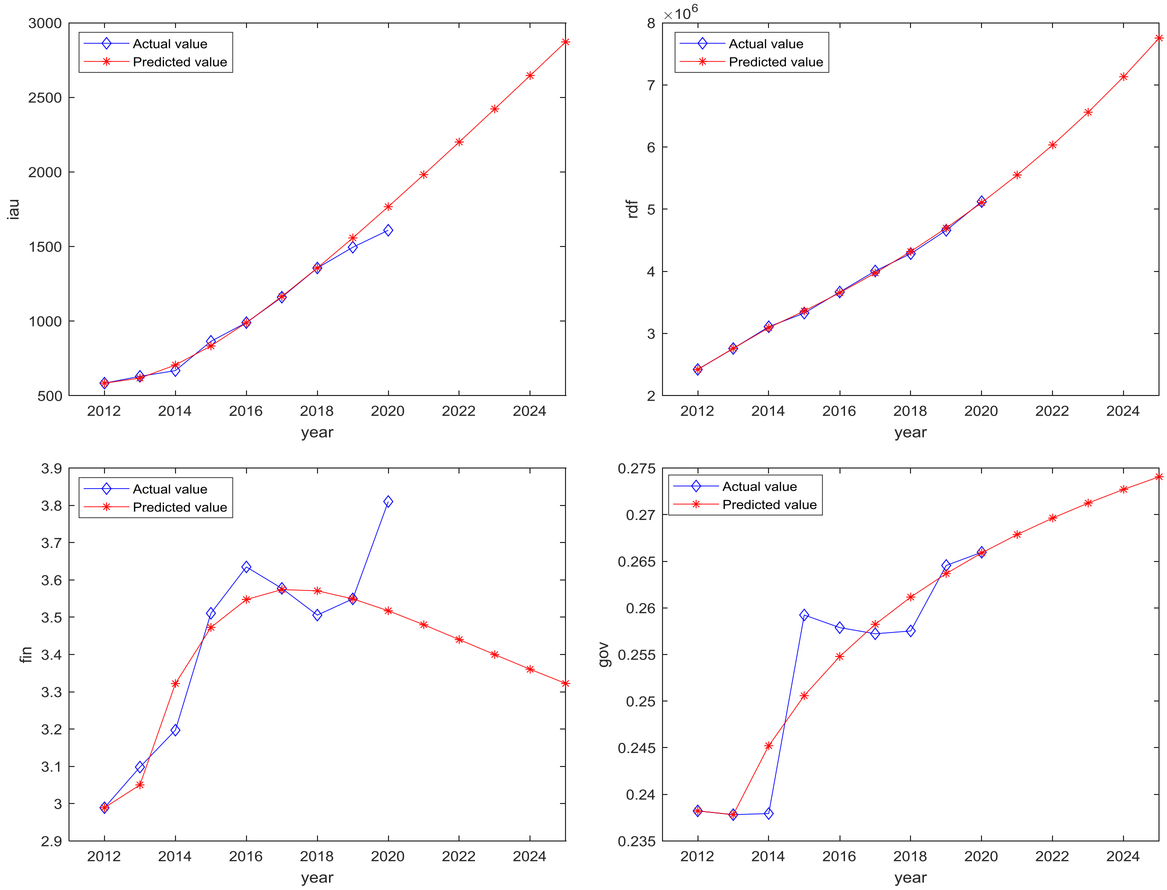


Fig. 2. Performance plots of the feature variables predicting the quantity of digital economy development.

Table 7 with the PVT test evaluation table in Table 3, we found that all four grey forecasting models have good performance when forecasting the quantity and quality of China's digital economy development, validating the choice of grey forecasting models in this paper. However, in forecasting the feature variables, except for TDGM(1,1,r) which passed the PVT tests for all the feature variables, the other three grey models failed the PVT tests for gov, indicating that all three models are not suitable for forecasting gov. More importantly, the TDGM(1,1,r) model performs better than the other three grey models in terms of both the root mean square deviation ratio (C) and the probability of small errors (P). All the characteristic variables had P values equal to 1 and most of them had C values less than 0.35, suggesting good forecast accuracy in accordance to Table 3. Other variables had C less than 0.50, a qualified prediction grade.

The above analysis suggests that the TDGM(1,1,r) grey forecasting model based on particle swarm algorithm optimization is more suitable for forecasting the development of China's digital economy. Therefore, this paper uses this model to make out-of-sample forecasts for all feature variables of the quantity and quality of digital economic development, and the results are shown in Figs. 2 and 3.

#### 4.2.1. *Critical feature variables for the quantity of digital economy development*

Figure 2 shows the performance of the TDGM(1,1,r) grey model predicting the critical feature variables of the quantity of digital economy development in the future (2021–2025).According to the prediction results on the quantitative aspects of digital economy development, we find that infrastructure, technology and economy are the main factors that will promote the quantitative development of China's digital economy in the future. Meanwhile, iau, rdf and gov show a significant rise in the next five years, which indicates that government spending, the number of Internet broadband access users and R&D funding will greatly advance the expansion of the core industries of China's digital economy.

#### 4.2.2. *Critical feature variables for the quality of digital economy development*

Figure 3 shows the performance of the TDGM(1,1,r) grey model predicting the critical feature variables of the quality of digital economy development in the future (2021–2025).According to the prediction results on the qualitative aspects of digital economy development, we find that economic, social and demographic characteristics are the main factors that will promote the qualitative development of China's digital economy in the future. The results show that all the feature variables have significantly increased, except for fin and tra, which have a decreasing or slightly increasing trend in the next five years. This indicates a tendency for China to reduce its traditional financial business in the future, and greater uncertainty and possible geopolitical risks in trade opening. Meanwhile, economic development, urbanization, Internet development, and the population's educational attainment will increase significantly, the industrial structure will tend to be rationalized, and government financial investment will increase, all of which will support the quality development of the digital economy.

#### 4.2.3. *Accuracy evaluation results of the prediction*

The above study proves that the predicted values of both quantity and quality of digital economy development are close to the actual values and that the development pattern is in line with the current situation in China, evidencing the use of the grey model of TDGM(1,1,r) to predict digital economy development. Besides, to further ensure the robustness of the model, we assess the quantity and quality of the development of the digital economy from in-sample and out-of-sample respectively through three statistical indicators of Ex-post prediction evaluation, and add three other grey models for comparison. From Tables 8 and 9, we find that the TDGM(1,1,r) grey model predicts a satisfactory ex-post evaluation indicator for all the feature variables in both the in-sample simulation evaluation and the out-of-sample prediction evaluation, and most of the estimated coefficients are smaller than those from the other three models. This provides further evidence that the adaptive Lasso method of shrinking feature variables with weak correlation is effective. More importantly, the construction of the control group highlights that the TDGM(1,1,r) grey model has better predictive performance and is a good prediction of the different quantitative and qualitative developments of the digital economy in the years 2021–2025.

### 4.3. *Predicting the performance of digital economy development*

Based on the above research results, this paper uses the TDGM(1,1,r) grey model to predict the quan-

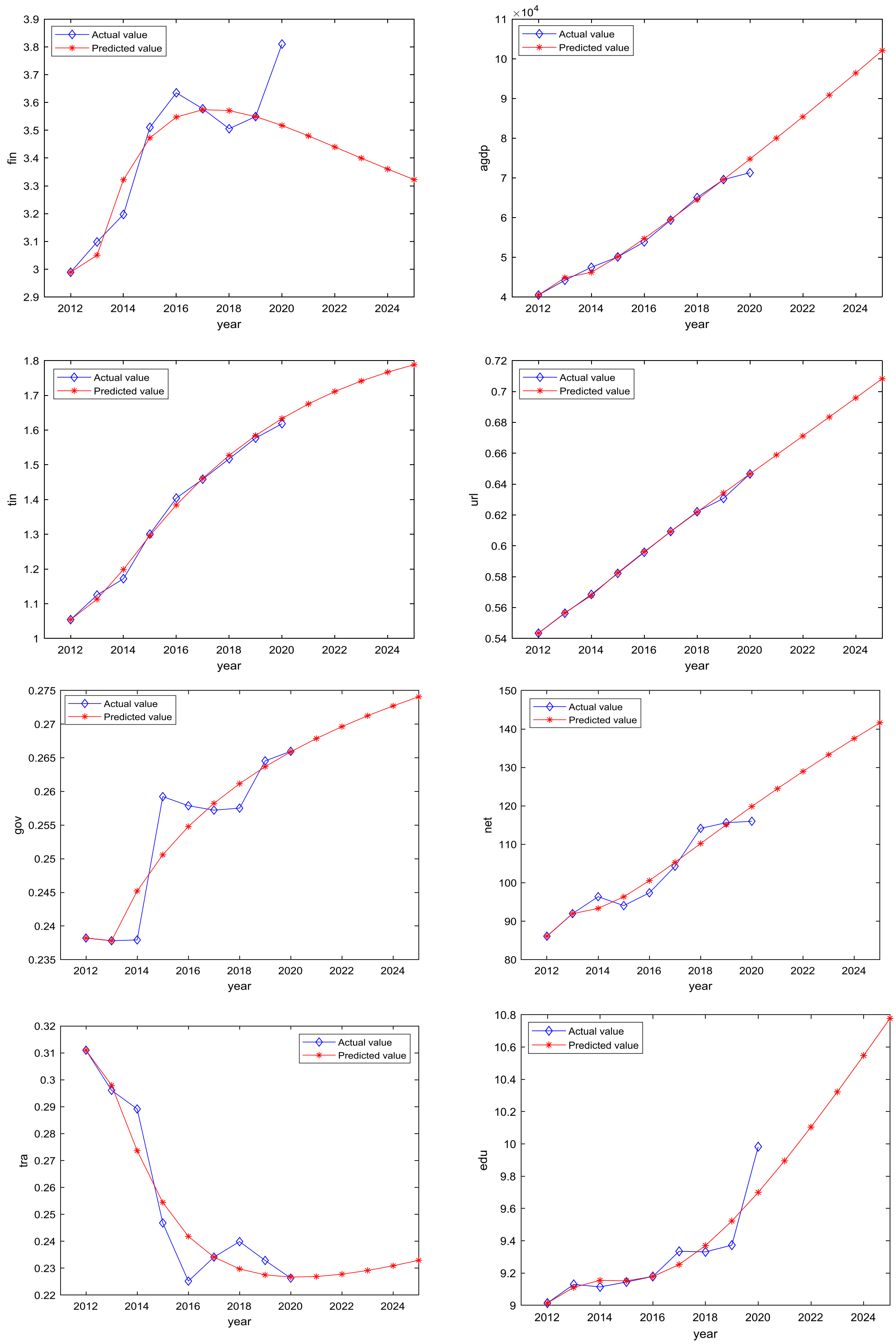


Fig. 3. Performance plots of the feature variables predicting the quality of digital economy development.

Table 8
Accuracy evaluation results for predicting the quantity of digital economy development

| | | MAE1 | MAE2 | RMSE1 | RMSE2 | MAPE1 | MAPE2 |
|---|---|---|---|---|---|---|---|
| Iau | TDGM(1,1,r) | 0.0157 | 0.0177 | 0.0196 | 0.0179 | 0.2697 | 0.2412 |
| | TDGM(1,1) | 0.0211 | 0.1567 | 0.0283 | 0.1660 | 0.3677 | 2.1295 |
| | GM(1,1) | 0.0216 | 0.1350 | 0.0306 | 0.1446 | 0.3805 | 1.8347 |
| | Grey Verhulst | 0.0340 | 0.0798 | 0.0491 | 0.0798 | 0.5786 | 1.0804 |
| Rdf | TDGM(1,1,r) | 0.0055 | 0.0053 | 0.0069 | 0.0073 | 0.0425 | 0.0345 |
| | TDGM(1,1) | 0.0057 | 0.0276 | 0.0075 | 0.0314 | 0.0440 | 0.1789 |
| | GM(1,1) | 0.0092 | 0.0155 | 0.0110 | 0.0157 | 0.0710 | 0.1009 |
| | Grey Verhulst | 0.0087 | 0.0330 | 0.0106 | 0.0330 | 0.0659 | 0.2137 |
| Fin | TDGM(1,1,r) | 0.0167 | 0.0345 | 0.0212 | 0.0381 | 1.5985 | 2.6152 |
| | TDGM(1,1) | 0.0246 | 0.0255 | 0.0311 | 0.0331 | 2.3676 | 2.0047 |
| | GM(1,1) | 0.0278 | 0.0408 | 0.0330 | 0.0456 | 2.6299 | 3.1745 |
| | Grey Verhulst | 0.0219 | 0.0500 | 0.0266 | 0.0500 | 2.0163 | 3.7355 |
| Gov | TDGM(1,1,r) | 0.0046 | 0.0017 | 0.0055 | 0.0024 | 1.9836 | 0.6483 |
| | TDGM(1,1) | 0.005846 | 0.001702 | 0.00711 | 0.001987 | 2.5615 | 0.64154 |
| | GM(1,1) | 0.006109 | 0.006514 | 0.007265 | 0.006515 | 2.6454 | 2.4518 |
| | Grey Verhulst | 0.00512 | 0.001613 | 0.007451 | 0.001613 | 2.1419 | 0.60601 |

Suffix 1: in-sample (2012–2018); Suffix 2: out-of-sample (2019-2020).

Table 9
Accuracy evaluation results for predicting the quality of digital economy development

| | | MAE1 | MAE2 | RMSE1 | RMSE2 | MAPE1 | MAPE2 |
|---|---|---|---|---|---|---|---|
| fin | TDGM(1,1,r) | 0.0167 | 0.0345 | 0.0212 | 0.0381 | 1.5985 | 2.6152 |
| | TDGM(1,1) | 0.0246 | 0.0255 | 0.0311 | 0.0331 | 2.3676 | 2.0047 |
| | GM(1,1) | 0.0278 | 0.0408 | 0.0330 | 0.0456 | 2.6299 | 3.1745 |
| | Grey Verhulst | 0.0219 | 0.0500 | 0.0266 | 0.0500 | 2.0163 | 3.7355 |
| agdp | TDGM(1,1,r) | 0.0062 | 0.0234 | 0.0080 | 0.0331 | 0.0664 | 0.2098 |
| | TDGM(1,1) | 0.0090 | 0.0313 | 0.0110 | 0.0427 | 0.0962 | 0.2798 |
| | GM(1,1) | 0.0098 | 0.0270 | 0.0119 | 0.0353 | 0.1055 | 0.2421 |
| | Grey Verhulst | 0.0085 | 0.0561 | 0.0110 | 0.0561 | 0.0891 | 0.5021 |
| tin | TDGM(1,1,r) | 0.0107 | 0.0112 | 0.0140 | 0.0118 | 0.9821 | 0.6956 |
| | TDGM(1,1) | 0.0194 | 0.0999 | 0.0249 | 0.1046 | 1.7439 | 6.2317 |
| | GM(1,1) | 0.0189 | 0.0989 | 0.0235 | 0.1040 | 1.6543 | 6.1641 |
| | Grey Verhulst | 0.0132 | 0.0186 | 0.0204 | 0.0186 | 1.1900 | 1.1515 |
| url | TDGM(1,1,r) | 0.0003 | 0.0041 | 0.0003 | 0.0042 | 0.0552 | 0.6379 |
| | TDGM(1,1) | 0.0004 | 0.0057 | 0.0004 | 0.0057 | 0.0699 | 0.8952 |
| | GM(1,1) | 0.0003 | 0.0055 | 0.0004 | 0.0055 | 0.0660 | 0.8639 |
| | Grey Verhulst | 0.0010 | 0.0050 | 0.0013 | 0.0050 | 0.1871 | 0.7799 |
| gov | TDGM(1,1,r) | 0.0046 | 0.0017 | 0.0055 | 0.0024 | 1.9836 | 0.6483 |
| | TDGM(1,1) | 0.0053 | 0.0020 | 0.0064 | 0.0028 | 2.5615 | 0.7456 |
| | GM(1,1) | 0.0061 | 0.0065 | 0.0073 | 0.0065 | 2.3033 | 2.4518 |
| | Grey Verhulst | 0.0051 | 0.0016 | 0.0075 | 0.0016 | 2.1419 | 0.6060 |
| net | TDGM(1,1,r) | 0.0188 | 0.0201 | 0.0227 | 0.0283 | 0.4749 | 0.4218 |
| | TDGM(1,1) | 0.0311 | 0.0329 | 0.0381 | 0.0339 | 0.7882 | 0.6924 |
| | GM(1,1) | 0.0230 | 0.0313 | 0.0269 | 0.0313 | 0.5696 | 0.6580 |
| | Grey Verhulst | 0.0225 | 0.0420 | 0.0268 | 0.0420 | 0.5587 | 0.8826 |
| tra | TDGM(1,1,r) | 0.0269 | 0.0195 | 0.0365 | 0.0276 | 1.5533 | 1.0203 |
| | TDGM(1,1) | 0.0433 | 0.0852 | 0.0545 | 0.0854 | 2.4800 | 4.4181 |
| | GM(1,1) | 0.0419 | 0.0982 | 0.0528 | 0.0986 | 2.4391 | 5.0940 |
| | Grey Verhulst | 0.0346 | 0.0842 | 0.0458 | 0.0846 | 2.0557 | 4.3350 |
| edu | TDGM(1,1,r) | 0.0029 | 0.0218 | 0.0042 | 0.0226 | 0.1519 | 0.9557 |
| | TDGM(1,1) | 0.0027 | 0.0234 | 0.0034 | 0.0270 | 0.1402 | 1.0224 |
| | GM(1,1) | 0.0033 | 0.0289 | 0.0040 | 0.0405 | 0.1719 | 1.2544 |
| | Grey Verhulst | 0.0030 | 0.0586 | 0.0040 | 0.0586 | 0.1549 | 2.5489 |

Suffix 1: in-sample (2012–2018); Suffix 2: out-of-sample (2019-2020).

tity of digital economy development and the quality of digital economy development, and the prediction performance is detailed in Fig. 4. It shows that since 2012, the quantity and quality of digital economy development have increased rapidly, but it is obvious that the quantity of digital economy development is faster, while the quality of digital economy development is gradually decreasing. It shows that

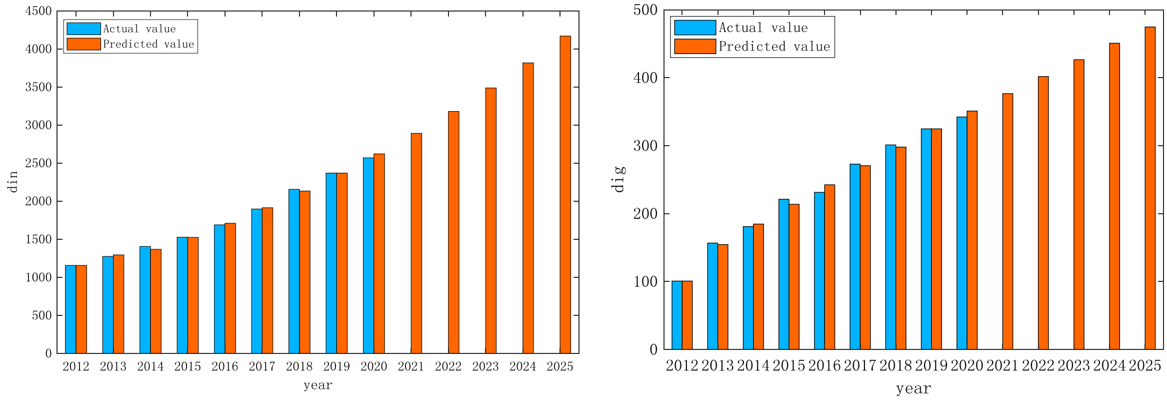


Fig. 4. Predicted quantitative and qualitative performance plots for digital economy development.

since 2012, the quantity and quality of digital economy development have increased rapidly, but it is obvious that the quantity of digital economy development is faster, while the quality of digital economy development is gradually decreasing. Meanwhile, the forecast results show that China's digital economy will have better development in both quantity and quality in the next five years, and the growth rate may be faster than the growth rate from 2021 to 2025.

The specific forecast values and growth rates are shown in Table 10. Two general findings include: First, compared with the development of China's digital economy in recent years, the quantity and quality of the digital economy will rise and expand in the next five years, which indicates that China has better prospects for the development of the digital economy in the future. Meanwhile, China's planning goal of making a bigger and stronger digital economy can also achieve certain results from 2021 to 2025. Second, the average annual growth rate of digital economy development quantity is 10.17%, and the average annual growth rate of digital economy development quality is 6.78%, which means that China still pays more attention to digital economy development quantity growth in the next five years, but the level of digital economy development quality is still at a lower growth rate. The level of digital industrialization is increasing, probably because the digital economy has not yet achieved deep integration with other industries, and thus the efficiency of industrial digitization is still low. In other words, although the digital economy has achieved quantitative growth, it has not deeply penetrated and widely integrated with all industries of the national economy.

Table 10
The specific forecast values and growth rates

| Year | Quantity | | Quality | |
|---|---|---|---|---|
| | Value-added scale | Growth rate | Overall level | Growth rate |
| 2021 | 2890.46 | 12.51% | 376.60 | 10.04% |
| 2022 | 3178.60 | 9.97% | 401.76 | 6.68% |
| 2023 | 3486.68 | 9.69% | 426.48 | 6.15% |
| 2024 | 3816.23 | 9.45% | 450.77 | 5.70% |
| 2025 | 4168.90 | 9.24% | 474.69 | 5.31% |
| Average annual growth rate | 10.17% | | 6.78% | |

### 4.4. Further research

From the above comparative analysis, it is interesting to see that the quantity of China's digital economy development may be faster than the quality of digital economy development in the next five years. Therefore, in order to explore the leading industries that will drive the quantity development of China's digital economy in the future, this paper selects four types of industries, including digital product manufacturing (dpm), digital product service (dps), digital technology application (dta) and digital factor-driven (dfd) industries, according to the classification of the core industries of digital economy by the China Bureau of Statistics, and predicts the development trend of different industries from 2021 to 2025 using the TDGM(1,1,r) grey model, and the prediction results are shown in Table 11.

Several observations can be made. First, all the major core industries of the digital economy show an increasing trend from 2021 to 2025. However, compared to the digital technology application industry and the digital factor-driven industry, which grow at an average annual rate of 20.53% and 14.52%, the digital product services industry and the digi-

Table 11
Forecasting the quantity of digital economy development with classification

| Year | dpm | dps | dta | dfd |
|---|---|---|---|---|
| 2021 | 775.25 | 121.10 | 1383.26 | 654.40 |
| 2022 | 808.14 | 127.79 | 1640.46 | 744.87 |
| 2023 | 841.72 | 134.37 | 1947.37 | 846.76 |
| 2024 | 876.08 | 140.83 | 2313.33 | 961.60 |
| 2025 | 911.30 | 147.18 | 2749.45 | 1091.10 |
| Average annual growth rate | 4.17% | 6.36% | 20.53% | 14.52% |

tal product manufacturing industry grow at a slower average annual rate of 6.36% and 4.17% respectively. Second, the digital technology application industry accounts for 52.22% of the core industry size of the digital economy from 2021 to 2025, and the digital factor-driven industry accounts for 22.37% of the core industry size of the digital economy. Therefore, from 2021 to 2025, the digital technology application industry and the digital factor-driven industry are the main development directions of the core industries of the digital economy with better future prospects. China should focus on supporting economic activities in these two industries, including specifically the digital technology application industry such as software development, Internet-related services and information technology services, and the digital factor-driven industry such as Internet platforms, Internet finance and information infrastructure construction.

## 5. Conclusions and policy recommendations

Aiming at China's digital economy, this article uses the adaptive lasso method to select key feature variables from quantitative and qualitative perspectives, and predicts the development trend of China's digital economy from 2021 to 2025 based on the TDGM(1,1,r) grey model optimized by the particle swarm algorithm. To validate the reliability of the combined model, the other three commonly used grey models (TDGM(1,1), Grey Verhulst, GM(1,1)) are selected for a comparative research on Ex-ante and Ex-post evaluations. The results show some differences in the feature variables that determine the quantity and quality of digital economy development from 2021 to 2025, as more feature variables seem associated with the quality of digital economy development. Furthermore, the combined adaptive lasso and TDGM(1,1,r) models proposed in this paper outperformed the other three grey models in predicting Chinese digital economy and its feature variables. From our analysis, we also find that China's digital economy will achieve the planned quantity and quality though the development is imbalanced in that the economy is developing significantly faster in terms of quantity, with the digital technology application industry the core industry for the quantitative development of the digital economy.

The above conclusions motivate the following suggestions. First, to promote the quantitative and qualitative collaborative development of the digital economy, it is imperative to focus on supporting traditional financial development (fin) and government fiscal spending (gov). On the one hand, traditional financial instruments are needed to accelerate the promotion of digital transformation, deepen the application of new technologies such as big data, artificial intelligence and blockchain in the financial industry, and promote the deep integration of digital technology and traditional finance. On the other hand, the government needs to introduce relevant policies and measures in terms of digital infrastructure, digital market construction, and digital talent incentives. Meanwhile, the government should insist on investing more financial funds in areas such as science and technology innovation and digital economy industry development to empower and strengthen the development of digital economy innovation.

Second, it is necessary to build a long-term mechanism for the high-quality development of the digital economy with equal emphasis on incentives and norms, to promote the rapid development of the quality of the digital economy in the next five years. First of all, it helps to transform and upgrade the traditional economic structure and maintain a healthy economic development environment, while improving the policy system of urban-rural integration and development, effectively promoting the free flow of digital elements between urban and rural areas, and accelerating the integration and development of digital economy and new urbanization. It is imperative to increase the degree of trade openness, through the "spillover effect" to attract cross-border digital trade investment, and promote the trade and investment of SMEs in the digital economy. In addition, China should strengthen the education of micro and small enterprises and low-income people about the digital economy, cultivate high-level financial and technological talents, facilitate the coverage and penetration of the digital economy, and provide a reliable innovation environment.

Third, China should focus on supporting the digital technology application industry and digital factor-driven industry to ensure that the quantity of the digital economy grows as predicted in the next five years. The country should focus on information transmission, software and information technology services, digital content and media, Internet finance, Internet wholesale and retail, and other industry sectors. Then the basic research should be strongly supported to speed up the construction of science and technology innovation and transformation and application of institutional mechanisms. In this way, China can focus its efforts on improving the core key technology research capabilities and supporting the production and application of digital technology products.

In this study, we applied the combined model of adaptive lasso and TDGM(1,1,r) to short-term forecasts. Future efforts can add to our findings by applying the model to simulate the long-term development of digital economy in China and beyond. Uncertainty can be studied in forecasting as there may be drastic changes in the sample data with the increased timespan. In addition, future research can identify and consider possible time-variability when making long-term forecasts of digital economy development, thus providing effective measure to improve the quantity and quality of the digital economy in the country concerned.

## Acknowledgments

We thank the following funds for their support: Humanities Research Project of Anhui Provincial Education Department, Research on the Long-term Mechanism of New Rural Cooperative Finance for High-Quality Service to Common Wealth (No. SK2021A0277).

## References

[1] H. Akaike, New look at statistical-model identification, *IEEE Transactions on Automatic Control* **AC19**(6) (1974), 716–723.
[2] S. Ben Jabeur, N. Stef and P. Carmona, Bankruptcy prediction using the xgboost algorithm and variable importance feature engineering, *Computational Economics* **1**(1) (2022), 1–27.
[3] C.S. Candra, J. Adrian and V.C. Lim, Indonesian trade deficit with China: Background and grey forecasting, **1**(2) (2021), 33–46.
[4] A. Charles and O. Darne, Backcasting world trade growth using data reduction methods, *World Economy* **1**(1) (2022), 1–23.
[5] X.H. Chen, L. Teng and W. Chen, How does fintech affect the development of the digital economy? Evidence from China, *North American Journal of Economics and Finance* **61**(1) (2022), 1–14.
[6] A.B.R. Costa, P.C.G. Ferreira, W.P. Gaglianone, O.T.C. Guillén, J.V. Issler and Y. Lin, Machine learning and oil price point and density forecasting, *Energy Economics* **102**(1) (2021), 105494.
[7] Y.B. Ding, H.Y. Zhang and S.T. Tang, How does the digital economy affect the domestic value-added rate of Chinese exports? *Journal of Global Information Management* **29**(5) (2021), 71–85.
[8] B.C. Ervural and B. Ervural, Improvement of grey prediction models and their usage for energy demand forecasting, *Journal of Intelligent & Fuzzy Systems* **34**(4) (2018), 2679–2688.
[9] M.F.M. Jalali and H. Heidari, Predicting changes in bitcoin price using grey system theory, *Financial Innovation* **6**(1) (2020), 1–13.
[10] X. Gou, B. Zeng and Y. Gong, Application of the novel four-parameter discrete optimized grey model to forecast the wastewater discharged in chongqing China, *Engineering Applications of Artificial Intelligence* **107**(1) (2022), 1–9.
[11] L. Hanzhong and Y. Bin, Asymptotic properties of lasso+Mls and lasso+Ridge in sparse high-dimensional linear regression, *Electronic Journal of Statistics* **7**(none) (2013), 3124–3169.
[12] Y. Hu, H.Q. Zhou, B. Yan, Z. Zou and Y.A. Li, An assessment of China's digital trade development and influencing factors, *Frontiers in Psychology* **13**(1) (2022), 1–10.
[13] J. Huang, B. Zhou, H. Huang, J. Liu and N.N. Xiong, An adaptive lasso grey model for regional fdi statistics prediction, *Cmc-Computers Materials & Continua* **69**(2) (2021), 2111–2121.
[14] M.X. Huang and C.X. Liu, A variable-order fractional discrete grey model and its application, *Journal of Intelligent & Fuzzy Systems* **41**(2) (2021), 3509–3522.
[15] S.A. Javed and D. Cudjoe, A novel grey forecasting of greenhouse gas emissions from four industries of China and India, *Sustainable Production and Consumption* **29** (2022), 777–790.
[16] J.M. Jiang, W.Z. Wu, Q. Li and Y. Zhang, A pso algorithm-based seasonal nonlinear grey bernoulli model with fractional order accumulation for forecasting quarterly hydropower generation, *Journal of Intelligent & Fuzzy Systems* **40**(1) (2021), 507–519.
[17] Y.F. Jiang, Prediction model of the impact of innovation and entrepreneurship on China's digital economy based on neural network integration systems, *Neural Computing & Applications* **34**(4) (2022), 2661–2675.
[18] L.M. Kung and S.W. Yu, Prediction of index futures returns and the analysis of financial spillovers - a comparison between garch and the grey theorem, *European Journal of Operational Research* **186**(3) (2008), 1184–1200.
[19] I.Y. Laksito and I.G.A. Yudiarta, Grey forecasting of inbound tourism to bali and financial loses from the Covid-19, *International Journal of Grey Systems* **1**(1) (2021), 48–57.
[20] E. Lechman and M. Popowska, Harnessing digital technologies for poverty reduction. Evidence for low-income and lower-middle income countries, *Telecommunications Policy* **46**(6) (2022), 1–18.

[21] J. Li, L. Chen, Y. Chen and J. He, Digital economy, technological innovation, and green economic efficiency-empirical evidence from 277 cities in China, *Managerial and Decision Economics* **43**(3) (2022), 616–629.

[22] M.-F. Li, H. Hu and L.-T. Zhao, Key factors affecting carbon prices from a time-varying perspective, *Environmental Science and Pollution Research* **1**(1) (2022), 1–17.

[23] C. Liang, Y.A. Xu, Z.L. Chen and X.F. Li, Forecasting China's stock market volatility with shrinkage method: Can adaptive lasso select stronger predictors from numerous predictors? *International Journal of Finance* & *Economics* **1**(1) (2022), 1–15.

[24] T.C. Liu, Digital policy in european countries from the perspective of the digital economy and society index, *Policy and Internet* **14**(1) (2022), 202–218.

[25] X. Lu, F. Ma, J. Wang and B. Zhu, Oil shocks and stock market volatility: New evidence, *Energy Economics* **103**(1) (2021), 105567.

[26] K. Mossberger, S. LaCombe and C.J. Tolbert, A new measure of digital economic activity and its impact on local opportunity, *Telecommunications Policy* **46**(1) (2022), 1–19.

[27] W. Pan, T. Xie, Z. Wang and L. Ma, Digital economy: An innovation driver for total factor productivity, *Journal of Business Research* **139** (2022), 303–311.

[28] T. Pencarelli, The digital revolution in the travel and tourism industry, *Information Technology* & *Tourism* **22**(3) (2020), 455–476.

[29] W. Qian and J. Wang, An improved seasonal Gm(1,1) model based on the Hp filter for forecasting wind power generation in China, *Energy* **209**(1) (2020), 118499.

[30] R. Rathnayaka and D. Seneviratna, Taylor series approximation and unbiased Gm(1,1) based hybrid statistical approach for forecasting daily gold price demands, *Grey Systems-Theory and Application* **9**(1) (2019), 5–18.

[31] U. Sahin, Future of renewable energy consumption in France, Germany, Italy, Spain, Turkey and Uk by 2030 using optimized fractional nonlinear grey bernoulli model, *Sustainable Production and Consumption* **25** (2021), 1–14.

[32] F.M. Septyari, Grey forecasting of the exports of indonesian palm oil to India, *International Journal of Grey Systems* **1**(2) (2021), 60–68.

[33] G. Soava, A. Mehedintu and M. Sterpu, Analysis and forecast of the use of E-commerce in enterprises of the European Union States, **14**(14) (2022), 8943.

[34] K. Vu, P. Hanafizadeh and E. Bohlin, Ict as a driver of economic growth: A survey of the literature and directions for future research, *Telecommunications Policy* **44**(2) (2020), 101922.

[35] Y. Wu, S. Lin, K. Shi, Z. Ye and Y. Fang, Seasonal prediction of daily Pm2.5 concentrations with interpretable machine learning: A case study of Beijing, China, *Environmental Science and Pollution Research* **29**(30) (2022), 45821–45836.

[36] Z. Xian and W. Tianqi, Measurement and forecast of the scale of China's digital economy core industries, *Statistical Research (In chinese)* **39**(1) (2022), 4–14.

[37] R.F. Xu and E. Guan, Can blockchain innovation promote total factor productivity? Evidence from Chinese-listed firms, *Applied Economics* **1**(1) (2022), 1–18.

[38] M.U. Yousuf, I. Al-Bahadly and E. Avci, A modified Gm(1,1) model to accurately predict wind speed, *Sustainable Energy Technologies and Assessments* **43**(1) (2021), 100905.

[39] S. Yuan, H.O. Musibau, S.Y. Genc, R. Shaheen, A. Ameen and Z. Tan, Digitalization of economy is the key factor behind fourth industrial revolution: How G7 countries are overcoming with the financing issues? *Technological Forecasting and Social Change* **165**(1) (2021), 1–7.

[40] B. Zeng and C. Li, Forecasting the natural gas demand in China using a self-adapting intelligent grey model, *Energy* **112** (2016), 810–825.

[41] S.S. Zhang, X.Y. Ma and Q. Cui, Assessing the impact of the digital economy on green total factor energy efficiency in the post-Covid-19 era, *Frontiers in Energy Research* **9**(1) (2021), 1–18.

[42] Y. Zhang, P. Ruan and J.F. Zhao, Design of digital economy consumer psychology prediction model based on canopy clustering algorithm, *Frontiers in Psychology* **13**(1) (2022), 1–13.

[43] Y. Zhao and Y. Zhou, Measurement method and application of a deep learning digital economy scale based on a big data cloud platform, *Journal of Organizational and End User Computing* **34**(3) (2022), 1–17.

[44] H. Zou, The adaptive lasso and its oracle properties, *Journal of the American Statistical Association* **101**(476) (2006), 1418–1429.

## Appendix

Table A1
Nomenclature

| Nomenclature | |
|---|---|
| (a,b) | The two parameters of GM(1,1) and Grey Verhulst |
| (a,b,c) | The three parameters of TDGM(1,1) and TDGM(1,1,r) |
| GM(1,1) | Univariate grey prediction model for first-order difference equations |
| Grey Verhulst | Univariate first-order non-linear dynamic grey forecasting model |
| TDGM(1,1) | A three-parameter discrete grey forecasting model |
| PSO | particle swarm algorithm optimization |
| TDGM(1,1,r) | TDGM(1,1) model based on PSO to optimize the cumulative order r |
| r | Optimal cumulative order r for the TDGM(1,1, r) model based on PSO |
| PVT | The Posterior-Variance Test |
| MAE | Mean Absolute Error |
| RMSE | Root Mean Squared Error |
| MAPE | Mean Absolute Percentage Error |