

数字经济的量化投资

贺 峰¹, 陈冠杰¹, 刘 磊¹, 刘映伶^{2*}

¹桂林信息科技学院信息工程学院, 广西 桂林

²桂林信息科技学院数学教研部, 广西 桂林

收稿日期: 2024年12月2日; 录用日期: 2025年2月20日; 发布日期: 2025年2月28日

摘要

中国数字经济的发展取得了显著成就, 形成了独特的增长模式。本研究通过综合运用关联度分析与层次分析法, 确立了影响数字经济发展的关键因素指标体系, 并采用灰色关联度分析法和BP神经网络评估了未来资金投入对数字经济整体发展的影响。鉴于研究中面临的数据量庞大、获取难度高及可能存在的数据缺失等问题, 研究团队构建了合理的指标体系, 有效合并了同类项, 确保了数据处理的准确性。该研究不仅深化了对数字经济影响因素的理解, 也为相关政策制定和实践操作提供了科学依据与重要参考。

关键词

数字经济, 灰色关联度, 成交量, BP神经网络

Quantitative Investment in the Digital Economy

Feng He¹, Guanjie Chen¹, Lei Liu¹, Yingling Liu^{2*}

¹School of Information Engineering, Guilin Institute of Information Technology, Guilin Guangxi

²Department of Mathematics, Guilin Institute of Information Technology, Guilin Guangxi

Received: Dec. 2nd, 2024; accepted: Feb. 20th, 2025; published: Feb. 28th, 2025

Abstract

China has made remarkable achievements in the development of its digital economy, forming a unique growth model. In this study, the index system of key factors influencing the development of the digital economy is established by comprehensively using correlation analysis and analytic hierarchy process, and the impact of future capital investment on the overall development of the digital economy is evaluated by using the grey correlation analysis method and BP neural network. In view

*通讯作者。

of the huge amount of data, the difficulty of obtaining and the possible lack of data in the research, the research team constructed a reasonable index system, effectively merged similar items, and ensured the accuracy of data processing. This study not only deepens the understanding of the influencing factors of the digital economy, but also provides a scientific basis and an important reference for relevant policy formulation and practical operation.

Keywords

Digital Economy, Grey Relevance, Volume, BP Neural Network

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

党的二十大对加快建设数字中国作出重要部署。报告强调，要站在统筹中华民族伟大复兴战略全局和世界百年未有之大变局的高度，统筹国内国际两个大局、发展安全两件大事，充分发挥海量数据和丰富应用场景优势，促进数字技术和实体经济深度融合，赋能传统产业转型升级，催生新产业新业态新模式，不断做强做优做大我国数字经济。

努力实现我国数字经济转向深化应用、规范发展、普惠共享的新阶段。为应对新形势新挑战，把握数字化发展新机遇，拓展经济发展新空间，推动我国数字经济健康发展(资料解读于“十四五”数字经济发展规划)。

2. 数字经济下面临的风险

2.1. 什么是数字经济

数字经济(Digital Economy)也称为数字化经济或网络经济，是指基于数字化技术和互联网基础设施的经济活动。它是指利用数字化技术和服务，在全球范围内实现信息和资源的流动、交换和创造价值的经济形态。

2.2. 影响数字经济的指标选取

本研究旨在运用量化投资策略[1]，深入剖析数字经济的发展规律，为投资决策提供科学依据。指标选取遵循全面性、代表性和可操作性原则，以全面、准确且可行地反映数字经济的技术基础、产业发展、市场规模及政策支持等关键方面。

朱波[2]等学者指出，在数字经济时代，伴随互联网、云计算、人工智能等技术革新及新型基础设施加速构建，我国数字经济迅速增长，年增速显著。结合中国统计网数据与文献分析，本研究最终甄选出互联网普及率、移动互联网用户数、电子商务交易额、金融科技投资额、云计算服务收入、政策支持力度及网络安全状况等关键指标，紧密关联“数字经济”领域。

基于 2021 年 7 月 14 日至 12 月 31 日每 5 分钟的“数字经济”板块指数数据作为训练集，本研究采用选定指标预测了 2022 年 1 月 4 日至 1 月 28 日每 5 分钟的板块指数成交量。随后，结合成交量预测与前期指标，进一步构建了板块指数收盘价预测模型。假设以“数字经济”板块指数为交易标的，设定初始资金 100 万元及 0.3% 的交易佣金，利用收盘价预测模型指导高频交易决策(每 5 分钟一次)。最终，计算了该期间交易的总收益率、信息比率及最大回撤率，以全面评估交易策略的表现与风险。

3. 数字经济风险形成的原因

数字经济是社会、国民经济中及其需要重视的问题，研究对象为我国近年来发展经济中的一系列时间序列分析，故我们可以采用关联度分析或者层次分析法，找出有哪些相关的指标对数字经济影响最大，那么我们就可以运用灰色关联度对未来的经费投入对整体的影响进行分析。

难点在于对数据的涉及量较大，指标涉及广，且较多指标的数据难以获取，且缺失。因此在构建指标体系时其个数不能多不能少。我们运用统计学知识合同同类项，将相关联或者有较大共同特征的指标和为一类，最终可以构建出较为合理的指标以便于研究。

4. 数据来源和解决思路

数字经济是社会、国民经济中及其需要重视的问题，通过《国家统计局》中所整理出相关数字经济的数据，对相关的指标数据进行了缺失值处理，来进行分析从而寻找到影响的主要因素和指标并建立数字经济对经济发展质量影响的指标体系。

研究对象为我国今年来发展经济中的一系列时间序列分析，故我们可以采用关联度分析或者层次分析法，找出有哪些相关的指标对数字经济影响最大，那么我们就可以运用灰色关联度对未来的经费投入对整体的影响进行分析。

5. 对数字经济进行相关预测

5.1. 指标排序模型

5.1.1. 建模思路

由于已给出数据有缺失，为保证分析结果的准确性，对于附件中各指标缺失项进行补全，我们采用了三次 Hermite 修正 Akima 插值法，将缺失项进行补全。

要求我们对已给出的相关数据进行分析，寻找“数字经济”板块有关的主要指标。我们接下来对所构建的 5 个指标分别与数字经济进行关联性分析，我们分别运用灰色关联度和相关系数两种方法进行分析，并将最后的结果进行综合排序得到排序指标模型，即数值从大到小的指标即为“数字经济”板块有关的最主要的指标。

5.1.2. 理论准备

灰色关联度

灰色系统理论是一种用灰色关联度顺序来描述因素之间关系强度，大小和次序的。其基本思想为：以因素的数据为依据，用数学的方法找到研究因素之间的几何关系。即曲线越接近，相应序列之间的关联度就越大，反之就越小[3]。

$$l_{\text{关联系数}} = \frac{\min_{i,k} |x_0(k)' - x_i(k)'| + l * \max_{i,k} |x_0(k)'' - x_i(k)'|}{\Delta_{0ij(k)} + l * \max_{i,k} |x_0(k)' - x_i(k)'|} \quad (1)$$

如果关联度越大，也就意味着 $x_i(k)$ 与最优指标 $x_0(k)$ 最为接近，即第 i 个被评价的对象优于其他指标，那么可以推出评价对象的优劣程度。

相关系数

相关系数是值描述变量之间相关程度的指标，用 $\zeta_{\text{相关系数}}$ 表示，取值为 $[-1,1]$ 。 $|\zeta_{\text{相关系数}}|$ 的值越大，误差越小，同理关联度。

$$l_{\text{相关系数}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

5.1.3. 模型建立

灰色关联度分析与传统相关系数在数据处理上存在显著差异。灰色关联度分析擅长处理信息不完备和不确定性问题,适用于数据不充分情境,且针对性强,无需变量正态性假设;而传统相关系数,如 Pearson 相关系数,则更适用于数据充分且明确、变量分布多样的相关性判定。

故本文为了更全面反映各个指标与“数字经济”板块有关的主要影响指标之间的相关性。我们综合考虑灰色关联度 $l_{\text{关联度}}$ 与相关系数 $l_{\text{相关系数}}$ 两方面,故我们可以得到如下模型:

$$l = \frac{1}{2} \times (l_{\text{关联度}} + l_{\text{相关系数}}) \quad (3)$$

之后用 MATLAB 将各个指标的值带入并进行排序,得出“数字经济”板块有关的最主要的指标。

5.1.4. 模型的求解

根据对初始数字经济相关数据的分析,本文对其指标进行数据可视化具体结果(见图 1)。

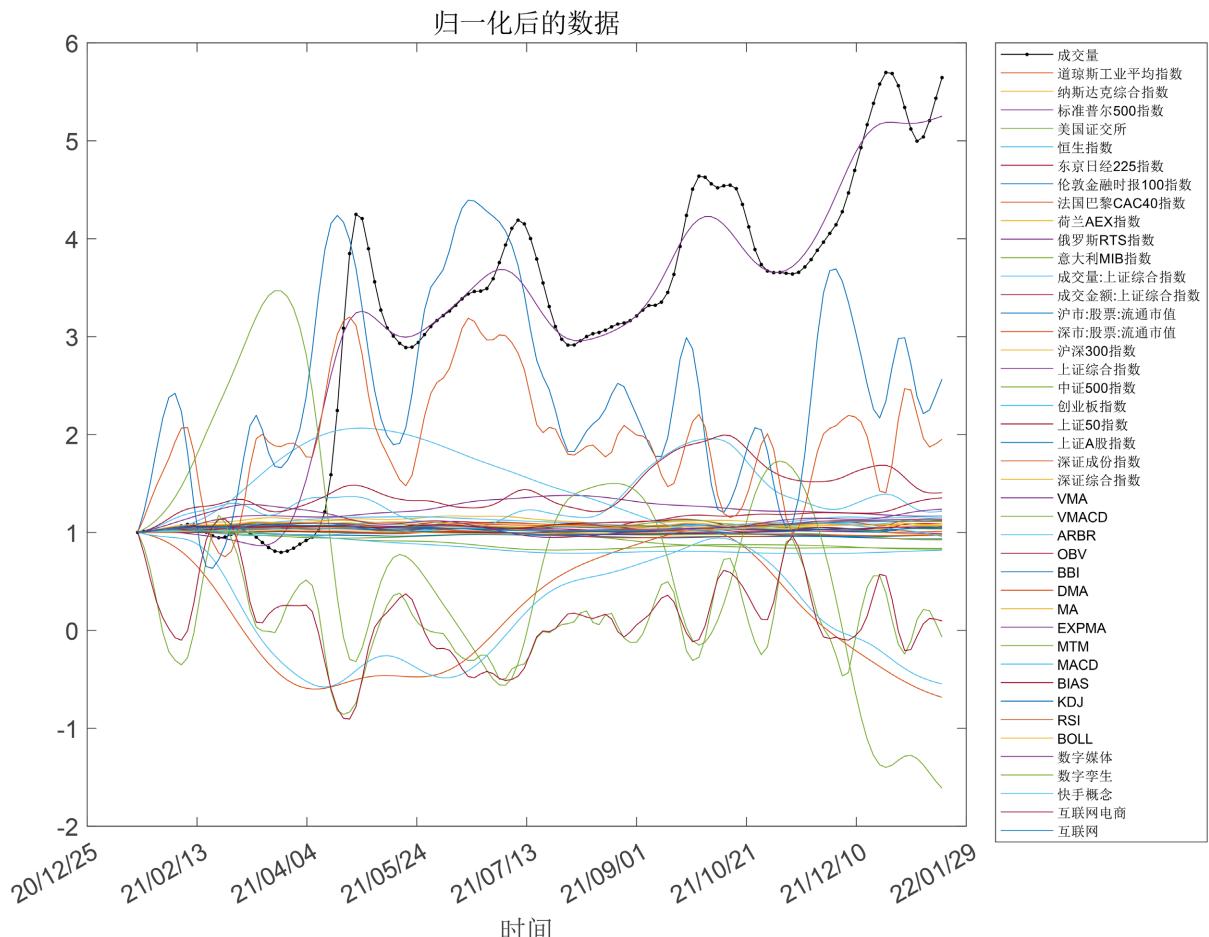


Figure 1. A graph of the normalized data from the original data

图 1. 原始数据的归一化后的数据图

图 1 为原始数据进行数据预处理归一化后的图像,由此我们可以看出一部分曲线光滑但一部分曲线呈波动状且无规律变化故我们用上文办法将数据进行关联度计算可得图 2:

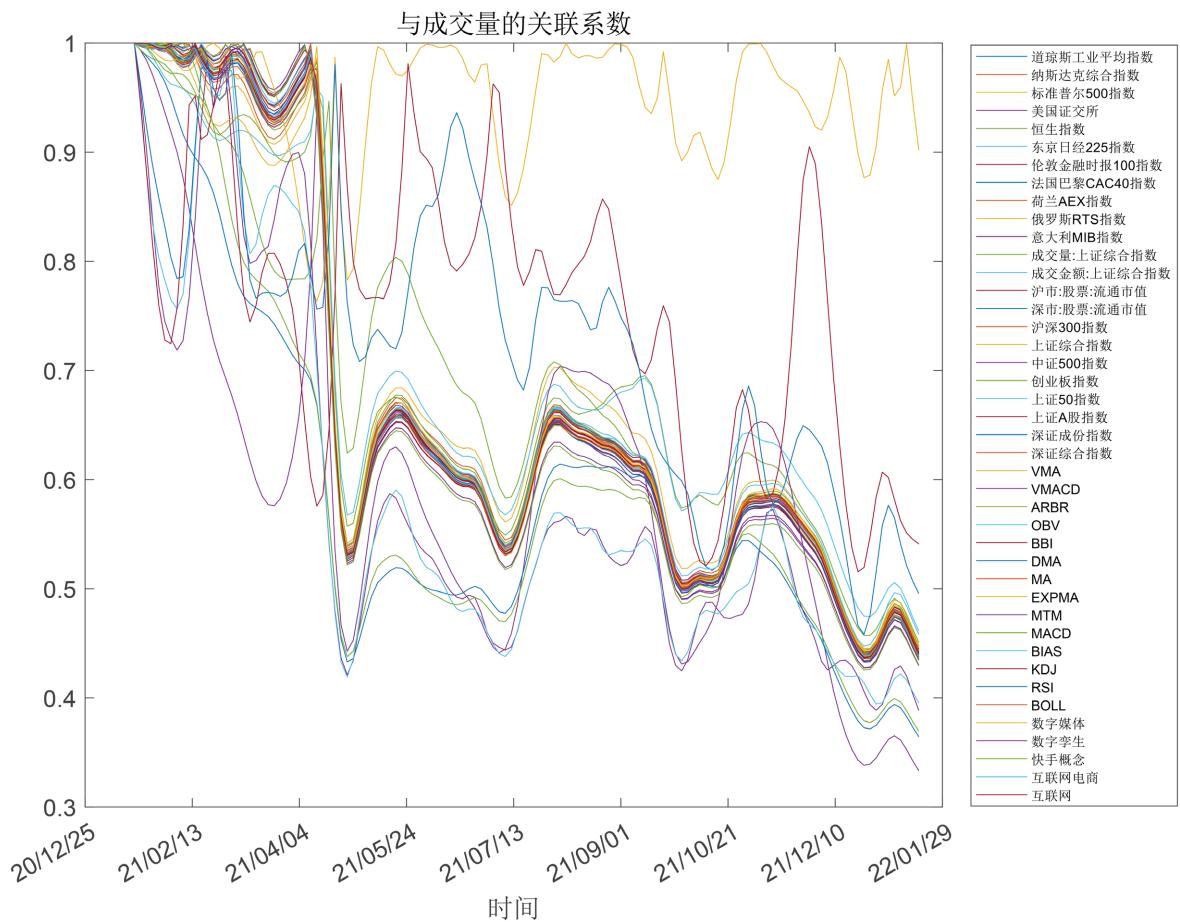


Figure 2. Grey correlation factor
图 2. 灰色关联系数

处理后的数据所展示的关联度较于原始数据更加的平滑和准确,那么我们可以知道灰色关联值排序情况(见表 1)。

Table 1. Grey correlation values of the most important factors related to the “Digital Economy” sector from December 2020 to January 2022

表 1. 2020 年 12 月到 2022 年 1 月“数字经济”板块有关的最主要因素的灰色关联值

名次	1	2	3	4	5
指标 ζ 值	俄罗斯 RTS 指数	ARBR	互联网电商	创业板指数	纳斯达克综合指数
	0.6826	0.6824	0.6756	0.6744	0.6728

5.1.5. 关键指标的选取

根据图 2 和表 1 我们可以清晰的观察到。在 2020 年 12 月到 2022 年 1 月这段时期不同指标对“数字经济”板块有关的最主要因素的关联性大小及排序,可以发现各个指标都有一定的关联性。可知在 2020 年 12 月到 2022 年 1 月,“数字经济”板块有关的最主要因素影响是俄罗斯 RTS 指数、ARBR、互联网

电商、创业板指数、纳斯达克综合指数。

5.2. 对“数字经济”板块指数每五分钟的成交量进行预测

5.2.1. BP 神经网络预测模型的建立与求解

数据预处理

在数据导入模型之前，我们实施了精细的数据清洗与归一化处理流程。此流程不仅包括常规的缺失值补全(采用三次 Hermite 修正 Akima 插值法)，还增设了异常值检测与处理环节，以确保输入数据的高质量。

针对数字经济领域的信息需求，本文旨在预测“数字经济”板块指标每五分钟的成交量。为实现此目标，我们采用了 BP 神经网络预测模型[4]。BP 神经网络(Back Propagation Neural Network)，即基于误差反向传播算法(Back Propagation Algorithm)的人工神经网络，是应用最为广泛的神经网络模型之一。它能够有效处理分类、回归、模式识别及数据挖掘等多种复杂问题。

方法一：

构成示意图：

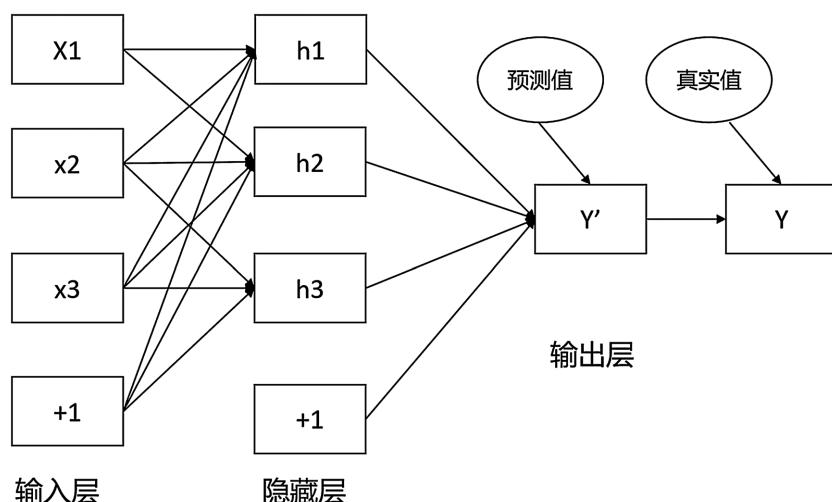


Figure 3. The basic structure of a BP neural network

图 3. BP 神经网络的基本构成

这是典型的三层结构：输入层(input layer)、隐藏层(hidden layer)及输出层(output layer)。输入层与输出层各为一层，而隐藏层数量依据具体学习任务可灵活设置。数据经输入层加权与非线性变换传递至隐藏层，再经类似处理最终输出预测值(见图 3)。

神经网络由输入层、隐藏层、输出层构成，各层分别配置有相应数量的神经元。预测值通过计算得出后，与真实值进行比较。若预测值与真实值高度一致，则表明所设权重与偏置参数较优，无需调整；反之，若差异显著，则需依据偏导数计算调整权重与偏置参数[5]，通过反向传播算法迭代更新参数，直至预测值逼近真实值(见图 4)。

BP 神经网络的整体流程见图 4。

方法二：

BP 神经网络作为一种经典的人工神经网络模型，在诸多领域展现出了强大的预测能力。然而，在实际应用中，模型的性能往往受到多种因素的影响，其中激活函数与损失函数的选择尤为关键。针对回归问题，合理的激活函数和损失函数能够加速模型收敛，提高预测精度。

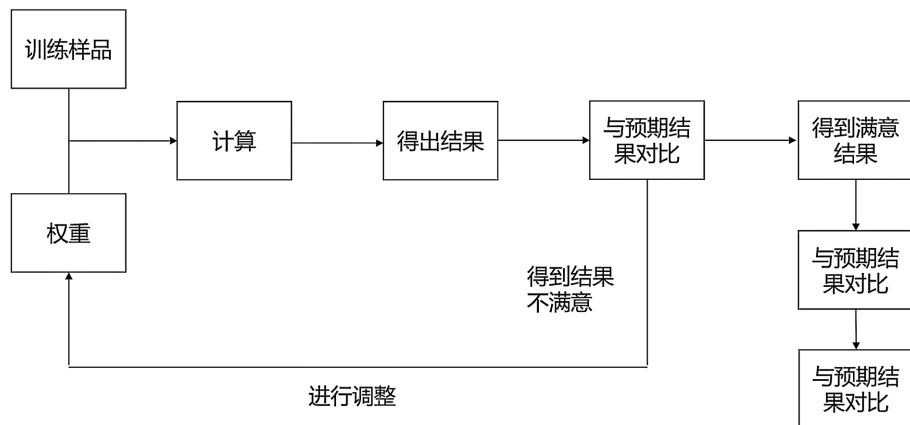


Figure 4. Overall flowchart of the BP neural network
图 4. BP 神经网络的整体流程图

1、激活函数的作用与选择

激活函数是神经网络中的非线性变换单元，它决定了神经元是否应该被激活以及激活的程度。在 BP 神经网络中，常用的激活函数包括 Sigmoid、Tanh 和 ReLU 等。其中，ReLU (Rectified Linear Unit, 修正线性单元) 函数因其简单高效、梯度消失问题较小等优点，在回归问题中得到了广泛应用。ReLU 函数的表达式为：

$$f(x) = \max(0, x) \quad (4)$$

2、损失函数的作用与选择

损失函数用于衡量模型预测值与真实值之间的差异，是模型训练过程中的优化目标。在回归问题中，均方误差(Mean Squared Error, MSE)是最常用的损失函数之一。MSE 计算预测值与真实值之间差的平方的平均值，能够反映预测的整体误差水平。然而，在某些情况下，MSE 的变种(如加权 MSE、对数 MSE 等)可能更适合特定问题的需求。

BP 神经网络模型构建

在模型构建阶段，本研究选用了 ReLU 作为隐藏层的激活函数。ReLU 函数不仅能够加快模型的收敛速度，还能在一定程度上缓解梯度消失问题，从而提高模型的训练效率和预测性能。

同时,针对回归问题的特性,本研究选用了均方误差 MSE 的变种作为损失函数。具体地,考虑到数据中的异常值可能对 MSE 产生较大影响,本研究采用了加权 MSE,即对不同的预测误差赋予不同的权重,以减小异常值对模型训练的影响。加权 MSE 的表达式为:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (5)$$

5.2.2. BP 神经网络的误差函数

BP 神经网络的误差函数为均方差函数:

$$E(w, b) = \frac{1}{m} \sum_{i=1}^m \frac{1}{m} \sum_{j=1}^k (\hat{y}_{ij} - y_{ij})^2 \quad (6)$$

其中, m 为训练样本个数, k 为输出个数, $\hat{y}_{i,j}$ 为第 i 样本, 第 j 输出的预测值, $y_{i,j}$ 为对应的真实值。

对于数字经济指标模块，本文选取了 2021 年 7 月 14 日到 2022 年 1 月 28 日每天 10 点的成交量数据，再将成交量的数据进行预处理。

将经过处理后的指标数据导入到 MATLAB 中, 通过 BP 神经网络训练, 得到了成交量的预测值。

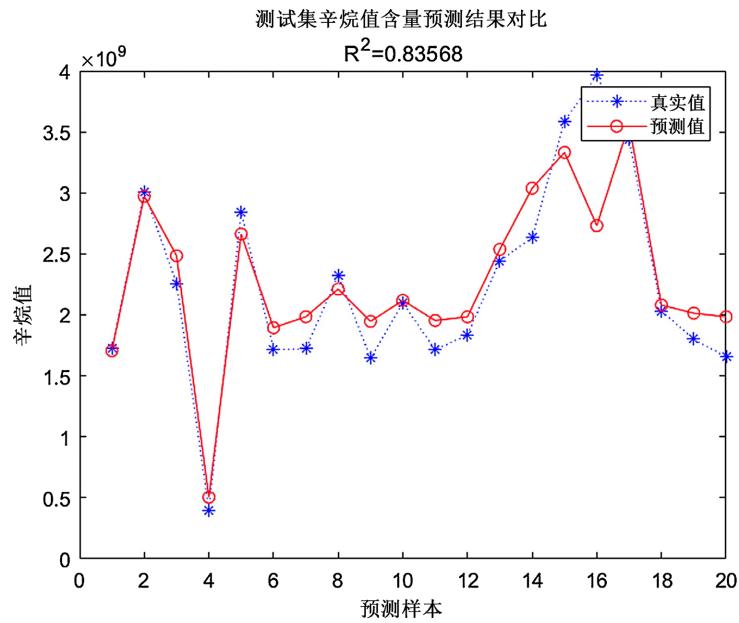


Figure 5. Predicted vs actual values with error analysis (NN-Based)
图 5. 基于神经网络的预测值与实际值的误差分析图

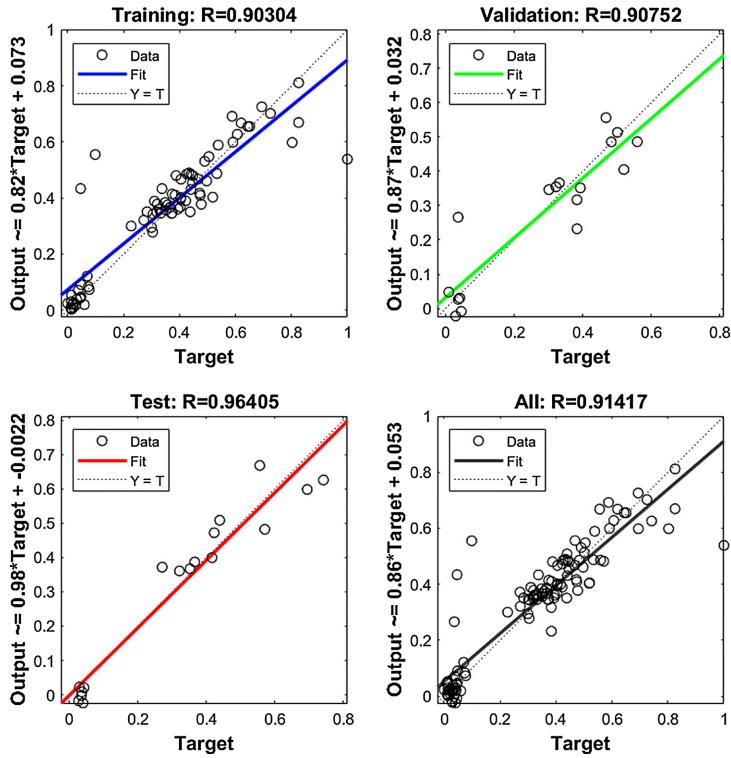


Figure 6. Correlation analysis images for individual sample sets and populations
图 6. 各个样本集和总体的相关性分析图像

我们使用 2021 年 7 月 13 日至 12 月 30 日“数字经济”板块指数数据训练预测模型，并生成误差分析(图 5)和相关性分析(图 6)图像，验证了模型可靠性。基于此模型，我们预测了 2022 年 1 月 4 日至 28 日该板块指数每 5 分钟成交量的部分值(见表 2)。

Table 2. Test values, predicted values, and errors
表 2. 测试值、预测值和误差

成交量测试值	成交量预测值	相互误差
1,727,292,600	1,706,433,351	0.012076268
3,006,247,000	2,971,398,223	0.01159212
2,258,221,200	2,484,278,963	0.100104349
400,096,410	505,039,432	0.262294335
2,839,991,900	2,661,728,023	0.06276915
1,716,299,100	1,894,056,146	0.10356997
1,721,823,200	1,983,840,127	0.152174118
2,324,209,100	2,211,529,100	0.048481008
1,644,826,400	1,946,441,966	0.183372279
2,095,599,900	2,118,426,228	0.010892503
1,712,966,400	1,953,669,013	0.140518
1,835,702,500	1,984,034,070	0.080803709
2,444,089,100	2,536,332,220	0.037741308
2,631,970,800	3,038,145,701	0.154323483
3,588,650,300	3,331,186,442	0.071743925
3,961,806,300	2,731,144,809	0.310631414
3,438,208,900	3,561,478,016	0.035852713
2,030,196,100	2,079,953,725	0.024508778
1,801,269,800	2,013,738,928	0.117955193
1,653,453,100	1,983,701,597	0.199732606
1,727,292,600	1,706,433,351	0.012076268

5.3. 对每 5 分钟的“数字经济”板块指数(收盘价)进行预测

通过上述的方法进行以收盘价为主体的预测,得出五个指标,对指标进行 BP 神经网络预测模型。为了分析与研究相关数据,更好地预测股票的收盘价[6],本文继续选用 BP 神经网络模型进行解决问题。

对于数字经济指标模块,本文选取了 2021 年 7 月 14 日到 2022 年 1 月 28 日每天 10 点的收盘价数据,再将收盘价数据与上述选取的主要指标数据选取相同的时间区域,随后将选取的数据导入到 MATLAB 中,通过 BP 神经网络算法得到预测数据。最后将预测数据和实际数据进行关联性分析和可视化。

如图 7 所示,为原始数据进行数据预处理归一化后的图像,由此我们可以看出一部分曲线光滑但一部分曲线呈波动状且无规律变化故我们用灰色关联度分析处理可得:

如图 8 所示,处理后的数据所展示的关联度较于原始数据更加的平滑和准确,那么我们可以知道灰色关联值排序情况(见表 3)。

将预处理后的指标数据输入至 MATLAB 环境中,利用 BP 神经网络模型进行训练,从而获得了收盘价的预测结果。

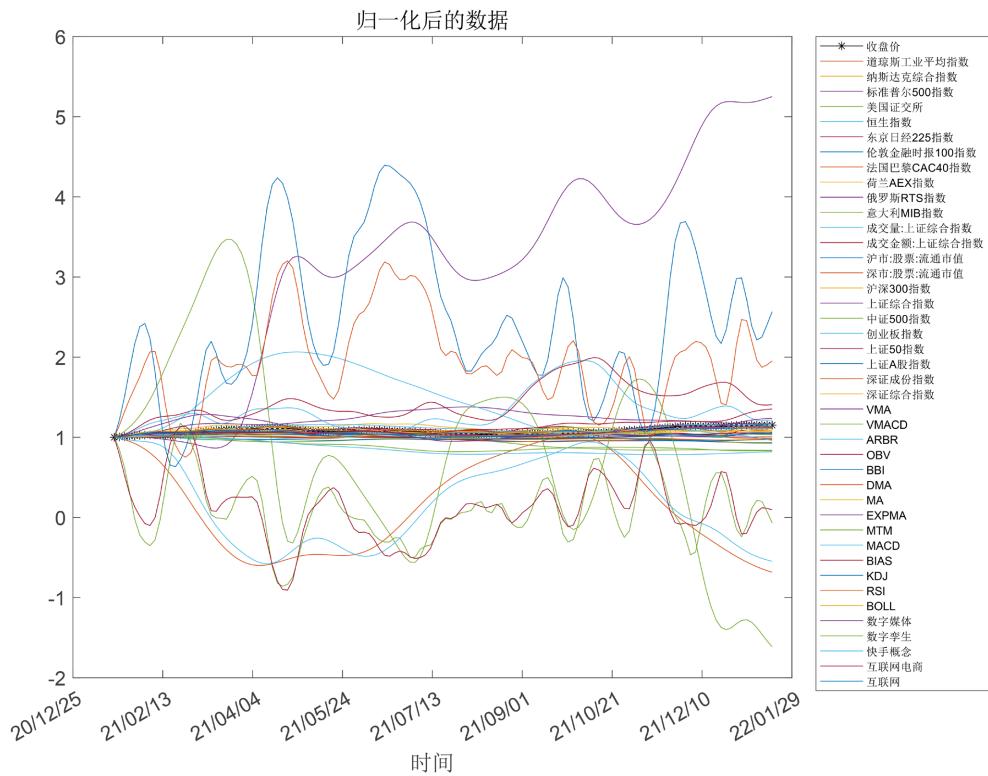
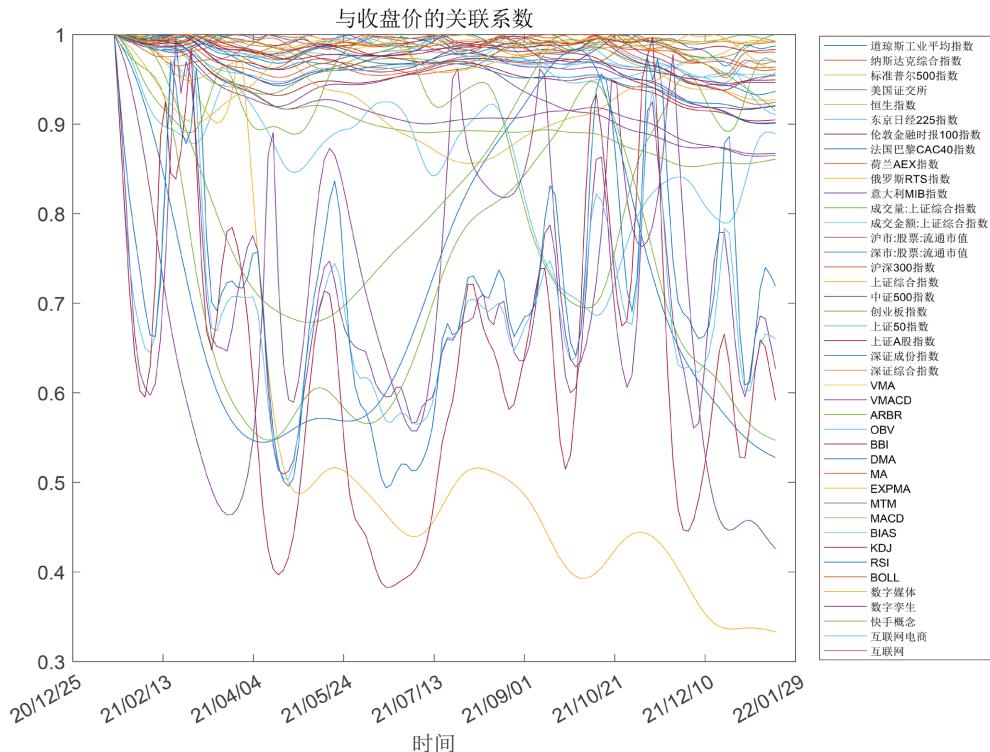
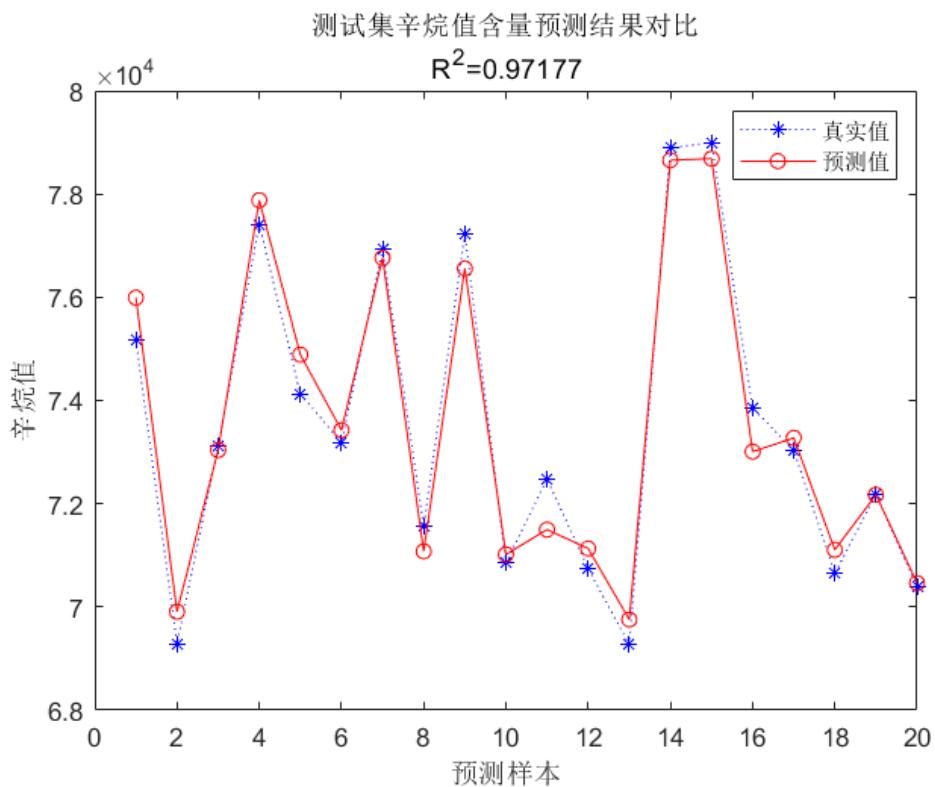
**Figure 7. Correlation thermodynamics graph****图 7. 原始数据的归一化后的数据图****Figure 8. Grey correlation factor****图 8. 灰色关联系数**

Table 3. Grey correlation values of the most important factors related to the “Digital Economy” sector from December 2020 to January 2022**表 3.** 2020 年 12 月到 2022 年 1 月 “数字经济” 板块有关的最主要因素的灰色关联值

名次	1	2	3	4	5
指标 ζ 值	EXPMA	MA	深证成份指数	深证成份指数	沪深 300 指数
	0.9927	0.9910	0.9897	0.9872	0.9872

**Figure 9.** Predicted vs actual values with error analysis (NN-Based)**图 9.** 基于神经网络的预测值与实际值的误差分析图

如图 9、图 10 所示对真实值和测试值的误差分析和相关性分析，使用 2021 年 7 月 14 日到 2022 年 1 月 28 日每天 10 点的收盘价数据通过 BP 神经网络的预测模型进行训练，从而得出 2022 年 1 月 4 日至 2022 年 1 月 28 日每天 10 点的收盘价数据的部分预测值，并且求出计算误差(见表 4、表 5)。

5.4. 计算在 2022 年 1 月 4 日至 2022 年 1 月 28 日期间交易的总收益率、信息比率、最大回撤率

我们采用 100 万元作为初始投资，以 0.3% 的交易佣金对“数字经济”板块指数进行高频交易。基于数据降维技术，我们构建了一个旨在最大化投资收益的模型，该模型能够综合考虑交易成本与预期收益，为投资者提供最优的交易策略。为评估投资效果，我们引入了信息比率和最大回撤率两个核心指标。利用 2022 年 1 月 4 日至 1 月 28 日的实际收盘价数据，我们计算了交易的总收益率、信息比率和最大回撤率，并与基于模型预测值的结果进行了对比分析，以验证模型的预测能力和投资策略的有效性[7] [8]。

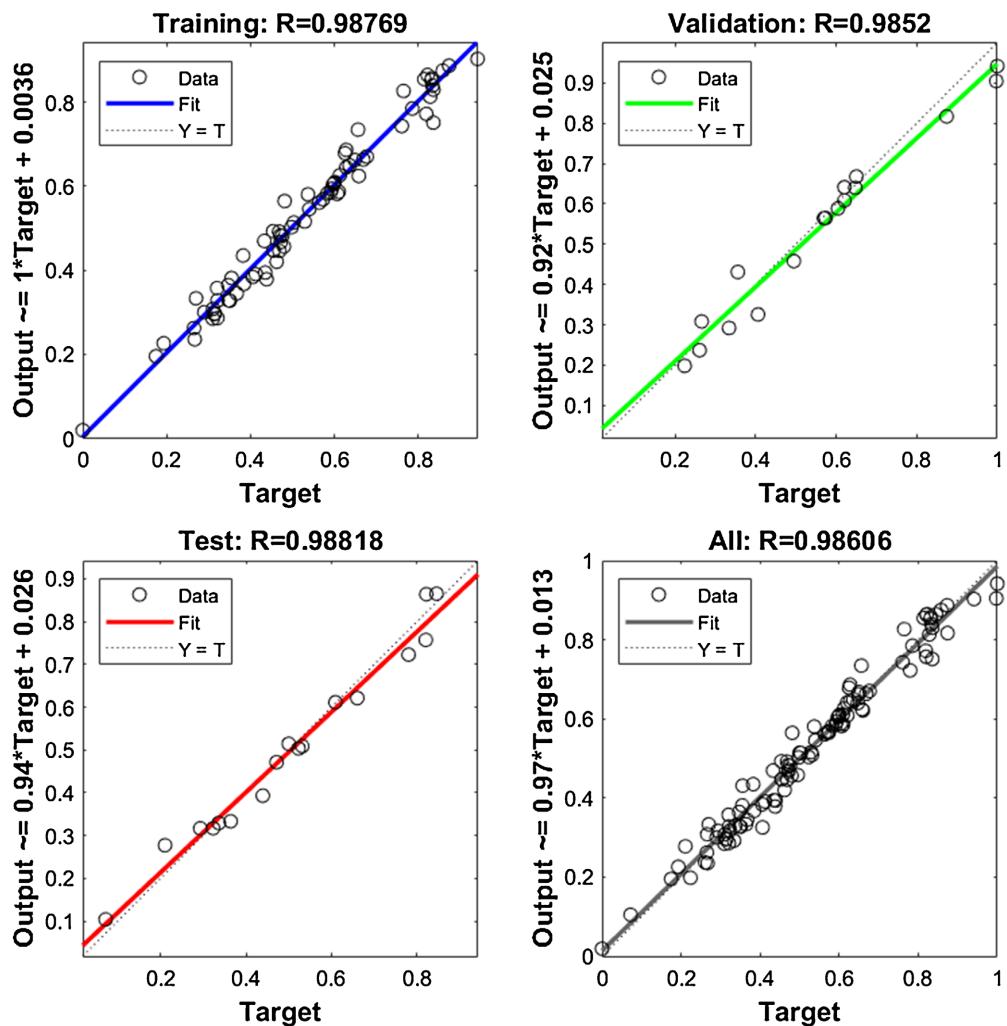


Figure 10. Correlation analysis images for individual sample sets and populations
图 10. 各个样本集和总体的相关性分析图像

Table 4. Error calculation
表 4. 误差计算

平均绝对误差 MAE	均方误差 MSE	均方根误差 RMSE
0.11615	0.00095896	0.030967

Table 5. Test values, predicted values, and errors
表 5. 测试值、预测值和误差

成交量测试值	成交量预测值	相互误差
1556.930302	1558.392854	0.000939382
1538.602998	1518.918232	0.012793922
1549.922904	1538.872345	0.007129748
1577.948029	1570.741318	0.004567141
1515.342785	1518.04661	0.001784299

续表

1637.984404	1631.687116	0.003844535
1574.80206	1564.834671	0.006329297
1552.890415	1548.348261	0.002924967
1560.875263	1545.853616	0.009623861
1489.891275	1490.162645	0.000182141
1520.352642	1519.228129	0.000739639
1443.492054	1454.72805	0.007783899
1612.044163	1601.945198	0.006264695
1476.01789	1481.571375	0.003762478
1503.072238	1502.227618	0.000561929
1501.099988	1498.97622	0.001414808
1461.47949	1488.819362	0.018706983
1460.371798	1473.375593	0.008904441
1473.266079	1477.623248	0.00295749

Table 6. Related calculation formulas**表 6. 相关计算公式**

总收益率 = $(P - 1) \times 100$, $P = (1 + \text{第1日当日收益率}) \times (1 + \text{第2日当日收益率}) \times \dots \times (1 + \text{第}N\text{日当日收益率})$

当日收益率 = $\frac{\text{当前收盘总资产市值} - \text{前一日收益总资产市值}}{\text{当日收盘总资产市值}}$, **信息比率** = $\frac{\text{日均超额收益率}}{\text{超额收益率标准差}}$

日均超额收益率 = $\frac{\text{第}i\text{天超额收益率}}{T}$, 其中 T 表示按交易日计算的天数

超额收益率 = 当日收益率 - 当日中证 500 指数收益率 $\times 90\%$,

超额收益率标准差: T (按交易日计算的天数)日内的每日超额收益率序列求得的标准差。

设 D_i 为第 i 天的产品净值, D_j 是 D_i 后面第 j 天的净值, 那么**最大回撤率** $Dr = \max \frac{D_i - D_j}{D_i}$

通过计算公式(见表 6), 我们使用 MATLAB 计算得出总收益率为 100.20%, 日均超额收益率为 0.0035, 超额收益率标准差为 0.0386, 信息比率 = $\frac{0.0035}{0.0386} = 9.19\%$, 最大回撤率为 11.17%。

基金项目

桂林信息科技学院 2024 年校级培育项目: 两类带终止的复发事件的半参数模型在生物医学领域的预测研究(XJ2024107)。

参考文献

- [1] 马萌. MATLAB 量化金融分析基础与实战[M]. 北京: 机械工业出版社, 2018.
- [2] 朱波, 曾丽丹. 数字经济时代区域经济下行风险防范与应对: 基于 GaR 模型的实证分析[J]. 中国软科学, 2024(6): 90-100.

- [3] 吴雪萍, 许贤丽. 新常态下金融与技术协同创新的实证研究——基于灰色关联度模型的分析[J]. 中小企业管理与科技, 2022(3): 59-61.
- [4] 罗雨晴, 张舒娟. 基于 PCC-BP 神经网络模型的碳配额价格预测[J]. 科技创业月刊, 2023, 36(S1): 76-80.
- [5] Zhang, X., Xu, N., Dai, W., et al. (2024) Turbofan Engine Health Prediction Model Based on ESO-BP Neural Network. *Applied Sciences*, **14**, 1996. <https://doi.org/10.3390/app14051996>
- [6] 胡智涛, 郝立丽. 基于神经网络模型的中国碳中和股票收盘价预测[J]. 统计与咨询, 2023(5): 16-19.
- [7] 肖奎喜. 我国开放式证券投资基金的业绩评价[D]: [博士学位论文]. 杭州: 浙江大学, 2005.
- [8] 韩立强. 基于 boosting 模型的逆向投资策略研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2019.