

基于时序数据的CPI组合预测分析

孙菲阳, 王智宇, 彭乐颖, 杨世花, 李文博

上海工程技术大学数理与统计学院, 上海

收稿日期: 2025年4月27日; 录用日期: 2025年5月19日; 发布日期: 2025年5月29日

摘要

居民消费价格指数(Consumer Price Index, 简称为CPI)是客观反映消费品及服务价格的波动情况的一种宏观经济指标。CPI水平直接体现通货膨胀或紧缩状况,它在经济发展中起着重要作用,为政府制定宏观经济政策、规划经济市场业务以及分析货币或债券市场提供重要参考依据。因此,对CPI的预测分析研究具有重要现实意义。首先,采用三种单项模型,即自回归滑动平均求和(ARIMA)模型、最小二乘支持向量(LSSVM)回归模型和BP神经网络模型,分别对上海市2004年1月至2024年1月间的CPI月度数据进行拟合预测,并对单一模型预测误差进行比较分析。其次,建立两种组合模型对CPI进行预测分析。第一种由ARIMA模型和BP神经网络模型集成,此组合模型将居民消费价格指数时间序列分解为线性自相关主体和非线性残差两部分,利用ARIMA模型预测其线性主体部分,继而用BP神经网络模型预测其非线性残差部分,将两部分相加得到最终预测值。第二种将ARIMA模型和LSSVM回归模型进行赋权组合预测。最后,对比单一模型和不同组合模型的预测效果,从而找到优化的模型来对上海市CPI进行预测。研究发现,单一模型中ARIMA模型在其特定条件下表现出了最为出色的预测能力。而组合模型ARIMA-BP和ARIMA-LSSVM则各结合了两种模型的优点,使得其在复杂情况下的预测效果优于单一模型,其中ARIMA-BP模型优于ARIMA-LSSVM模型。总体上,ARIMA-BP模型预测表现最为卓越,预测结果最为准确。

关键词

CPI预测, ARIMA模型, LSSVM模型, 组合模型

Time-Series Data Based Composite Forecasting Analysis of CPI

Feiyang Sun, Zhiyu Wang, Leying Peng, Shihua Yang, Wenbo Li

School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai

Received: Apr. 27th, 2025; accepted: May 19th, 2025; published: May 29th, 2025

Abstract

The Consumer Price Index (CPI) is a macroeconomic indicator that objectively reflects the fluctuation

文章引用: 孙菲阳, 王智宇, 彭乐颖, 杨世花, 李文博. 基于时序数据的 CPI 组合预测分析[J]. 统计学与应用, 2025, 14(5): 158-172. DOI: 10.12677/sa.2025.145134

of consumer goods and services prices. The level of CPI directly indicates inflation or deflation conditions, playing a crucial role in economic development by providing significant references for government policy-making, planning economic market operations, and analyzing monetary or bond markets. Therefore, research on forecasting analysis of CPI has substantial practical significance. Firstly, three single models—namely, the Autoregressive Integrated Moving Average (ARIMA) model, Least Squares Support Vector Machine (LSSVM) regression model, and Backpropagation (BP) Neural Network model—are employed to fit and forecast the monthly CPI data of Shanghai from January 2004 to January 2024. Comparative analysis of prediction errors among these single models is conducted. Secondly, two composite models are established for CPI forecasting. The first composite model integrates ARIMA and BP Neural Network models, decomposing the time series of the consumer price index into linear autocorrelation components and nonlinear residuals. The linear component is predicted using the ARIMA model, while the nonlinear residual part is predicted using the BP Neural Network model; their sum constitutes the final prediction value. The second composite model combines ARIMA with LSSVM through weighted forecasting. Finally, the forecasting effectiveness of single models versus different composite models is compared to identify the optimized model for predicting Shanghai's CPI. The study found that among single models, the ARIMA model exhibited superior forecasting capability under specific conditions. Composite models ARIMA-BP and ARIMA-LSSVM each combine the advantages of two models, making them more effective in complex scenarios than any single model, with the ARIMA-BP model outperforming the ARIMA-LSSVM model. Overall, the ARIMA-BP model demonstrated the most outstanding performance in terms of prediction accuracy.

Keywords

CPI Forecasting, ARIMA Model, LSSVM Model, Composite Model

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

自从改革开放以来,中国经济的迅猛发展导致了经济体制的重大变革,从而引发了我国居民消费结构的根本性变化。居民消费价格指数(Consumer Price Index, CPI),作为关键的统计指标,旨在精确衡量消费者在购买商品时所面临的价格水平。作为宏观经济的重要参考依据,该指数深刻揭示了消费品和服务价格的变动趋势。经济发展的过程中,居民消费价格指数(CPI)已经成为衡量国家或地区经济状况的重要标准之一。本文针对CPI时序数据,通过运用单一模型和组合模型进行预测及比较分析,旨在构建出合适的数学模型,以更精确地预测CPI。这将对相关部门进行宏观经济分析和决策提供重要的参考价值,有助于经济的稳健发展。

时间序列分析旨在研究随时间变化的事物,发现规律并预测未来趋势。1970年,Box和Jenkins提出了一套基于纯随机理论的时间序列建模方法,ARIMA模型因此也被称为Box-Jenkins模型[1]。国内外许多学者利用时间序列分析方法对CPI的预测进行过长期深入的研究。Yeko Mwanga针对Uganda地区2010年1月至2020年7月的CPI数据进行了分析,建立ARIMA模型对未来12个月份的CPI预测结果得到了当地统计局的认可[2]。另外,Nyoni和Thabani建立ARIMA(1, 1, 1)模型分析德国1960年至2017年的CPI年度时间序列数据,预测结果显示德国的居民消费价格指数在未来十年内将持续呈上升趋势[3]。Wedding提出了一种将传统的ARIMA模型和径向基神经网络相结合所形成的集成模型对时间序列进行

预测[4]。

国内不少学者运用统计分析、机器学习及综合评价等模型方法对 CPI 进行预测分析。朱颜杰等人构建 SARIMA 模型, 以我国 1990 至 2011 年 CPI 数据进行预测相对误差均控制在 1% 以内[5]。此外, 蔡敏建立了 SARIMA 和 BP 神经网络模型, 结果表明组合模型分析要更优于单一模型分析, 得出的结论也更具有参考意义[6]。黄树花证实了由自回归滑动平均求和(ARIMA)模型与误差反向传播(BP)模型组成的集成模型用于预测居民消费价格指数的有效性[7]。姚金海、邹家骏构建基于小波分解的 SVM-ARIMA 组合模型对 CPI 进行精准预测证明该组合模型的预测分析框架具有较强的适应性和扩展性[8]。山东大学的闫泓水构建了混合模型 EEMD-SARIMA-LSTM, 该模型对 CPI 指数预测有较高的精度[9]。

潘静等[10]学者在研究中构建了 ARIMA 模型与 GM(1, 1)模型, 旨在对居民消费价格指数进行预测。这两种模型在 CPI 预测中的应用为我国政策制定者提供了有效的决策依据。彭乃驰等[11]采用了小波分析与 BP-SARIMA 模型对 CPI 进行预测。通过对比分析, 他们发现 BP-SARIMA 模型在预测 CPI 方面具有较高的准确性。许明燕[12]构建了 ARIMA 模型与 BP 神经网络模型对江苏省 GDP 进行预测分析发现 BP 神经网络模型的预测效果优于 ARIMA 模型。此外, 山东大学郑雪平与水庆象[13]构建了分段线性 Lagrange 插值 Grey-Markov 模型用于预测 CPI, 为我国 CPI 预测提供了新的视角。

随着互联网技术的不断发展, 已有较多学者运用网络搜索数据预测研究 CPI, 并取得了一定的成效。如唐玉辉基于百度搜索指数对居民消费价格指数进行预测研究[14]。应虹存利用网络搜索数据的及时性和组合预测模型的精确度进行建模预测 CPI [15]。

通过分析国内外学者的研究结论可以得知, 在针对 CPI 预测的研究中, 国内外学者采用了多种方法和技术, 其中包括时间序列分析、机器学习、综合评价等方法。这些方法在预测 CPI 方面具有一定的优势, 但同时也存在一定的局限性[16][17]。常见的组合预测方法包括简单平均、加权平均、Stacking 等[18]。组合预测方法在一定程度上克服了单一模型的局限性, 但如何选择和组合合适的模型仍然是一个挑战。

2. 相关知识与理论基础

2.1. 时间序列

时间序列(Time series)是一种重要的统计方法, 时间序列的主要应用广泛而多样, 包括系统描述、系统分析、未来预测、决策制定以及控制。其分析的核心目的有两个: 一是通过分析数据建立数学模型以揭示其内在规律; 二是根据历史数据和相关因素, 对未来的趋势进行预测和预报。

2.2. 预测精度的评价指标

为了对组合后的模型进行深入有效的评估与讨论, 我们需采用各种性能评价指标, 旨在对所研究的组合模型进行深入有效的评估与讨论, 从而得出有益的结论。

通常评价预测精度的方法有以下几种:

(1) 预测误差的方差(MSE)和均方根误差(RASE)

$$MSE = \frac{\sum_{i=1}^n e_i^2}{n} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

$$RASE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

(2) 平均相对误差(MPE)和平均相对误差绝对值(MAPE)

$$\text{MPE} = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} \quad (3)$$

因为 MPE 存在正负抵消的问题, 因此 MAPE 应用更为广泛, 公式如下

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

均方根误差(RASE)是一种衡量模型预测结果与实际值之间差异的指标, 它对模型总体性能的评估具有较高的敏感度。而平均绝对百分比误差(MAPE)则是一种相对指标, 它能更好地反映模型在预测过程中的相对误差情况。在进行评估过程中, 我们还需要关注一些可能影响模型组合效果的因素, 如模型之间的相关性、组合方式等。通过对比实验和分析, 我们可以找到一种或多种有效的组合策略, 以提高组合模型的预测准确性和稳定性。

2.3. 组合模型简介

在实际应用中, 我们面临着复杂多变的数据挑战, 预测数据往往伴随着许多不确定的风险。这就导致了我们需要针对不同数据选择合适的单一模型。如 ARIMA 模型、LSSVM 模型、BP 神经网络模型等, 虽然在特定领域具有较高的预测准确性, 但它们各自针对的数据类型和条件范围具有很大的局限性。组合预测模型的基本原理在于, 根据各类单一模型的特性和权重进行整合, 构建出一个预测能力更强的模型。在预测过程中, 此类方法能有效降低信息损耗, 提升模型精度与稳定性。

本文的数据为上海市的 CPI 同比数据, 一般蕴含线性分量和非线性分量的不同组成部分, 所以, 本文将利用线性模型应用性较好的 ARIMA 模型与非线性模型应用性较好的 LSSVM 模型和 BP 神经网络模型来拟合, 以便最大程度地提取 CPI 中的有效信息。

3. 基于时序数据 CPI 的定性分析

3.1. 数据来源与描述性分析

本文的研究对象为上海市的居民消费价格指数, 在获取数据的过程中发现, CPI 数据在 CEIC 中国统计数据库里找到的有两种, 同比数据与环比数据。在查阅国内外文献的过程中, 注意到大部分国内研究 CPI 的文献倾向于使用 CPI 同比数据, 而在国外, 部分发达国家倾向于采用经过季节调整的 CPI 环比数据。本文选取上海市 2004 年 1 月至 2024 年 1 月的上海市 CPI 月度同比数据, 所有数据均来自 CEIC 中国统计数据库。

根据图 1, 我们可以看到从 2004 年到 2024 年间, CPI 指标的发展变化呈现出显著的波动。在初始阶段, 指标数值较低, 但在接下来的几年中呈现出稳步上升的趋势。特别是在 2006 年 7 月至 2007 年 8 月期间, 增长幅度尤为明显。然而, 在随后的几年里, 指标值有所回落, 并在一段时间内保持相对稳定。

3.2. 数据预处理

在时间序列分析中, 首先进行数据平稳性检验, 若为非平稳序列, 出现了明显的增长或者下降的趋势, 则要进行差分处理, 转换为平稳序列。常用的稳定性分析方法包括通过绘制时序图来直观展示数据的变化趋势。本文先利用时序图进行初步判断, 再通过相关性图进行深入验证, 以确保分析结果的准确性和可靠性。

依据图 1 的初步分析, 我们可以推断该序列具有非平稳特性, 同时峰值的出现呈现出一定的周期性, 因此还具有季节性特征。然而, 由于时序图判断的主观性, 为了进一步确认该序列的非平稳特性, 我们

将对原始序列进行 ADF 单位根检验。接下来，我们看表 1 中的 ADF 检验结果。表中列出了变量、差分阶数、T 检验结果、AIC 值等信息。在这些信息中，最为关键的是 T 检验结果和 AIC 值。

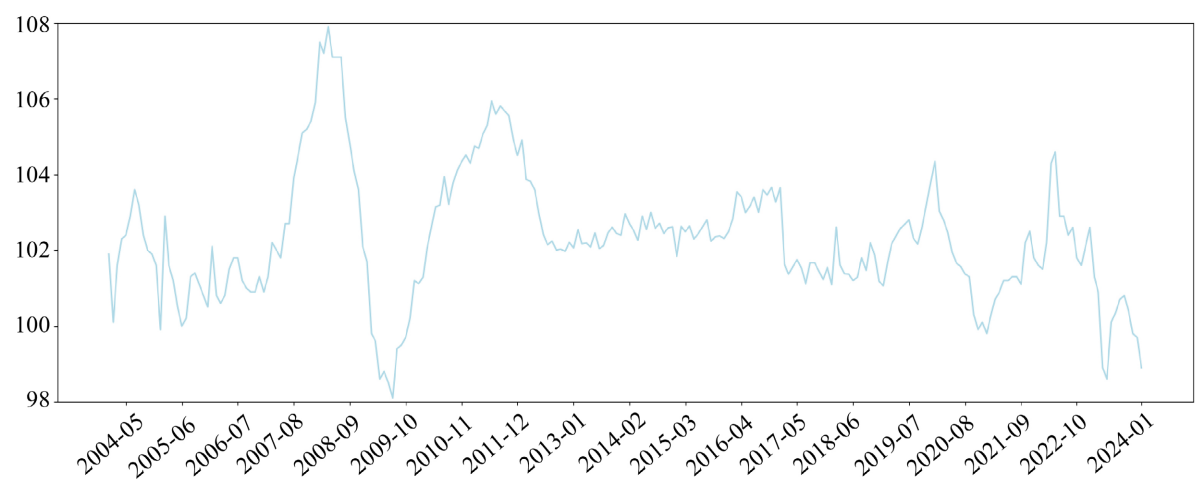


Figure 1. The time series chart of Shanghai’s monthly CPI (year-on-year) from 2004 to 2024
图 1. 2004 年到 2024 年间上海市月度 CPI (同比)时序图

Table 1. ADF unit root test table
表 1. ADF 单位根检验表

变量	差分阶数	t	p	AIC	临界值(1%)	临界值(5%)	临界值(10%)
CPI (同比)	0	-3.158	0.023**	322.594	-3.46	-2.874	-2.574
	1	-7.721	0.000**	330.321	-3.459	-2.874	-2.574
	2	-6.858	0.000**	351.153	-3.46	-2.875	-2.574

在差分为 0 阶时，显著性 P 值为 0.023**，这意味着在 5%的显著性水平上，根据统计学原理在差分为 0 阶时，我们可以得出结论：在 5%的显著性水平上，拒绝原假设，表明该时间序列具有平稳性。

在差分为 1 阶和 2 阶时，显著性 P 值皆为 0.000***，这意味着在 1%、0.1%的显著性水平上，根据统计学的原理，我们可以拒绝原假设，认为该序列为平稳时间序列。然而，与差分为 0 阶时的结果相比，我们需要注意到，随着差分阶数的增加，AIC 值也在逐渐减小。这表明，差分为 2 阶时，模型对数据的拟合程度更好。

综上所述，通过对表 1 中的 ADF 检验结果进行分析，我们可以得出以下结论：

- (1) 在差分为 0 阶时，该时间序列在 5%的显著性水平上拒绝原假设，说明它是一个平稳的时间序列；
- (2) 在差分为 1、2 阶时，该时间序列在 1%、0.1%的显著性水平上拒绝原假设，同样说明它是一个平稳的时间序列。

4. 基于单一模型的上海市 CPI 预测

4.1. 基于 ARIMA 模型的预测分析

时间序列数据可通过实验或统计部门获取。经数据预处理，本文已知该时间序列数据平稳，无需差分，可直接确定 ARIMA (p, d, q)模型的 d 为 0。若模型基本解释系统数据相关性且噪声序列为白噪声，则模型验证通过。本文论述的时序数据具备平稳性，无需对其进行进一步的差分处理。在理论层面，差

分次数的增加能够从时序信息中提取出更多不稳定的确定性信息，然而在实操过程中，差分次数并非越多越好。每一次的差分计算都会导致信息损失，因此，应防止过度差分现象。

基于上述步骤分析，差分结果如下图 2。

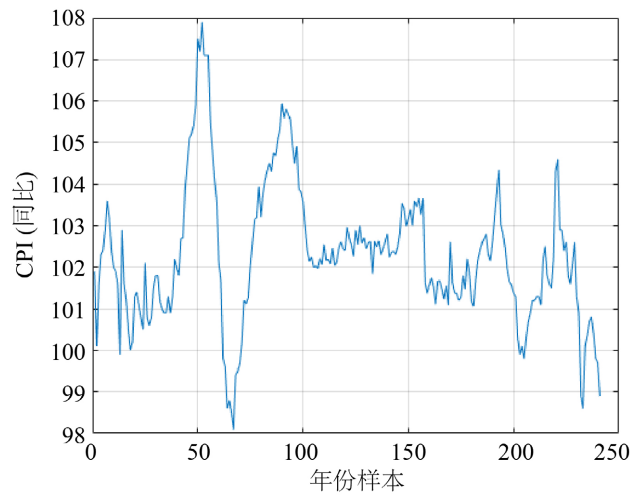


Figure 2. The difference chart of Shanghai’s monthly CPI (year-on-year) from 2004 to 2024

图 2. 2004 年到 2024 年间上海市月度 CPI (同比)差分图

选定模型为 ARIMA 模型(1, 0, 0)最终得到真实值和拟合值之间的对比图，如图 3 所示：

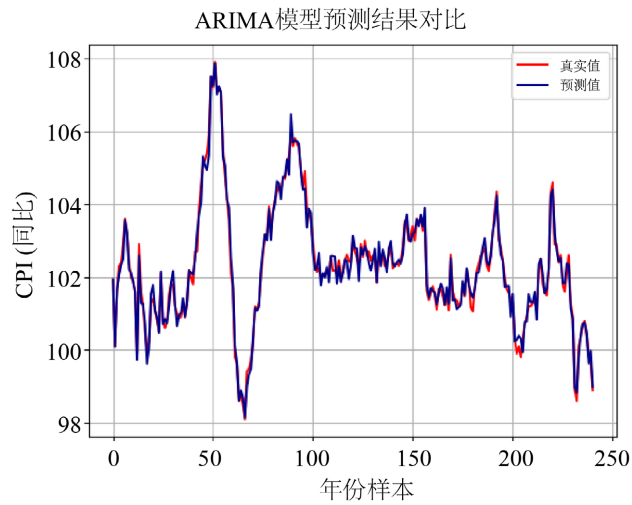


Figure 3. ARIMA model fitting and prediction chart

图 3. ARIMA 模型拟合预测图

为确保 ARIMA 模型的准确性，必须验证其残差是否为白噪声，即无自相关特性。

通过分析模型表格中的 Q 统计量及其对应的 P 值，若 P 值大于 0.1，则可以认为残差符合白噪声要求，从而确认模型的有效性。在多次分析模型对比过程中，信息准则 AIC 与 BIC 值用于评估(越低表明表现越好)，同时， R^2 值反映时间序列的拟合效果，接近 1 则表示效果越好。

表 2 详细列出了模型检验的各项指标，如自由度、Q 统计量、样本规模以及模型的拟合优度。在 ARIMA 模型(1, 0, 0)的检验中，我们观察到 Q 统计量的结果。根据这一结果，ARIMA 模型(1, 0, 0)的 P 值在统计上不显著，这意味着我们无法否定该模型的残差序列符合白噪声特征的假设。除此之外，该模型的拟合优度 R^2 值为 0.855，显示出模型具有较高的拟合效果，基本符合我们的预期要求。

Table 2. ARIMA model (1, 0, 0) diagnostic table
表 2. ARIMA 模型(1, 0, 0)检测表

项	符号	值
样本数量	Df Residuals	239
	N	241
Q	Q6 (P)	0.002 (0.966)
	Q12 (P)	9.072 (0.170)
	Q18 (P)	77.782 (0.000***)
	Q24 (P)	89.929 (0.000***)
	Q30 (P)	98.761 (0.000***)
信息准则	AIC	476.378
	BIC	486.833
拟合优度	R^2	0.855

Table 3. ARIMA model parameter table
表 3. ARIMA 模型参数表

	系数	标准差	t	$P > t $	0.025	0.975
常数	7.282	2.203	3.305	0.001	2.964	11.601
Ar.L1	0.929	0.022	43.081	0	0.886	0.971
Sigma2	0.409	0.025	16.232	0	0.36	0.458

基于上海市月度 CPI (同比)，最终模型结果为 ARIMA 模型(1, 0, 0)，根据 ARIMA 模型参数检验表 3，得到的模型公式如下：

$$y_t = 7.282 + 0.929y_{t-1} \tag{5}$$

4.2. 基于最小二乘支持向量回归模型的预测分析

在支持向量机(SVM)的基础上，最小二乘支持向量回归(LSSVM)是一种基于统计学习理论的机器学习技术。鉴于 CPI 数据的非线性和复杂性，我们可以利用 LSSVM 回归构建上海市 CPI 预测模型。

在 LSSVM 回归中，首先需要定义目标函数。目标函数通常是一个凸优化问题，用于寻找一个超平面，使得预测值与实际值之间的误差最小。假设训练集为 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ，其中 x_i 为第 i 个样本的特征向量，为第 i 个样本的真实值。

在 LSSVM 回归的优化问题中，存在特定约束条件。这些约束条件有助于确保算法的有效性和稳定性，训练样本的真实值 y_i 与预测值 $f(x_i)$ 之间的差距需小于等于给定的容忍度 ϵ 。这个约束条件可以表示为：

$$-\epsilon \leq f(x_i) \leq \epsilon \quad (6)$$

最后，将约束条件代入目标函数，我们可以得到 LSSVM 回归的对偶问题。在实际应用中，我们还需要根据问题的特点选择合适的核函数和参数，以获得更好的预测效果。在评估 LSSVM 回归模型性能时，通常无法像传统模型那样获得确定性方程，而是通过测试数据预测精度来评估模型性能。

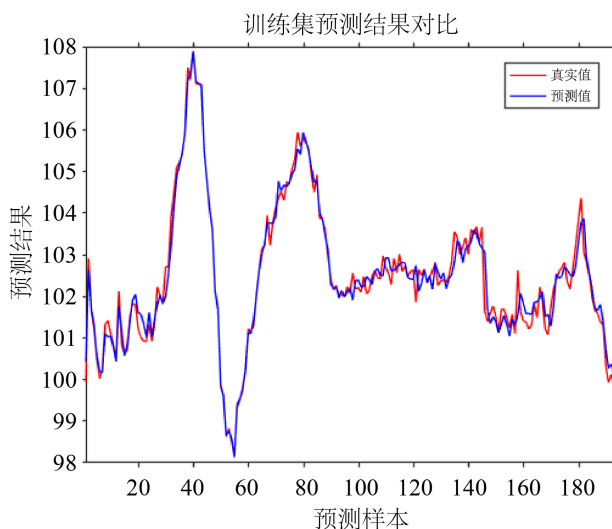


Figure 4. LSSVM model fitting effect on the training set

图 4. LSSVM 模型拟合训练集效果图

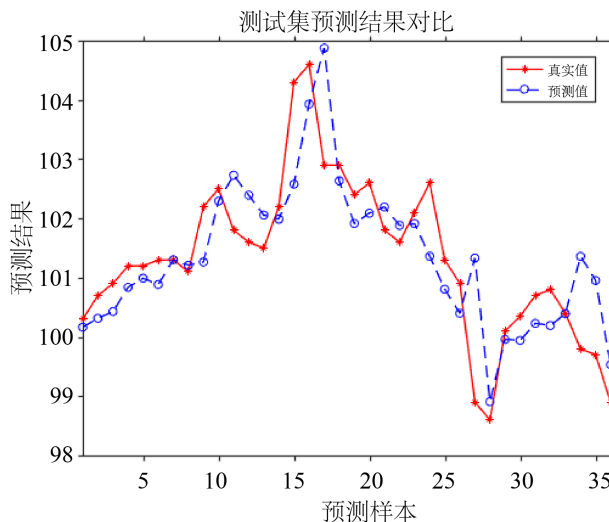


Figure 5. LSSVM model fitting effect on the test set

图 5. LSSVM 模型拟合测试集效果图

图 4、图 5 揭示了 LSSVM 回归模型预测值与数据真实值之间的误差对比分析。结果显示，LSSVM 回归模型具有较好的拟合效果，因此可将其纳入组合模型筛选之列。

4.3. 基于 BP 神经网络模型的预测分析

在当今世界，数据驱动的决策越来越受到重视，预测分析作为一种重要的数据挖掘方法，在各行各

业中都发挥着至关重要的作用。在众多预测分析模型中，BP (Back Propagation)神经网络模型由于其强大的自学能力和较高的预测精度，受到了广泛关注和应⤵。单隐层网络由输入层、隐含层和输出层三个核心层次组成，这一结构布局在图 6 中得到了清晰的展示。单隐层网络在数据处理和决策支持方面发挥着至关重要的作用，为现代科技发展和智能化进程提供了坚实的基础。

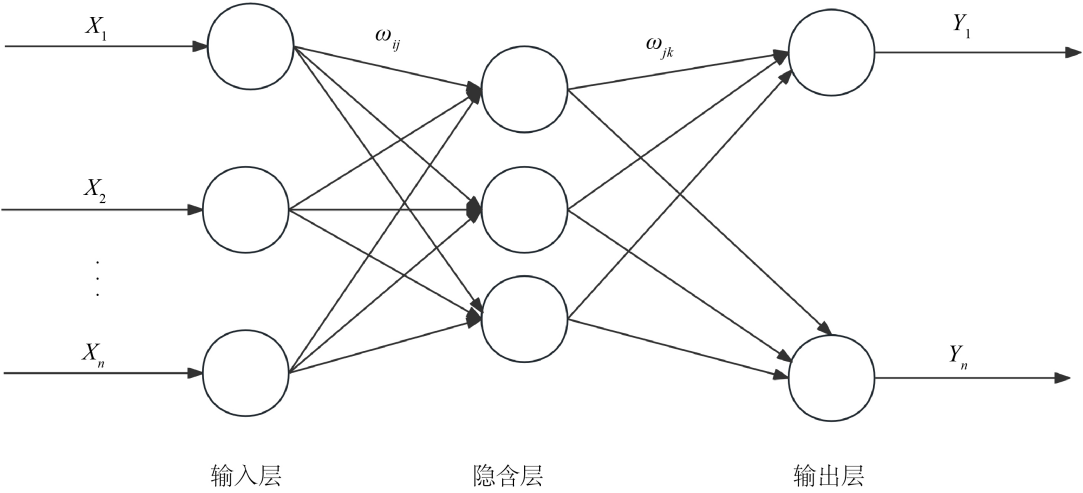


Figure 6. Topology of the BP neural network
图 6. BP 神经网络拓扑结构

BP 神经网络的学习过程主要包括两个阶段：正向传播和逆向传播。在正向传播阶段，输入信号历经输入层，继而传递至隐含层。在隐含层，信号经过一系列复杂处理与计算，进而传递至输出层。若输出层的结果与预期存在偏差，即存在误差，神经网络便会进入误差逆向传播阶段。在此阶段，依据输出层的误差信号，网络自输出层逐层向前计算，并调整各层间的连接权值及神经元阈值。这一调整过程旨在最小化输出误差，使网络的实际输出更接近预期目标。通过不断交替进行正向传播与逆向传播，神经网络的性能得以逐步优化。

根据 Kolmogorov 定理，BP 网络具有出色的非线性映射功能。理论上，一个三层 BP 神经网络(包括输入层、输出层和中间层)足以拟合任意非线性函数。

根据本文所选时间序列数据特征，将已知上海市历史 CPI 数据设为输入变量，上海市 CPI 预测值设为输出变量。我们可以构建一个三层的 BP 神经网络，输入层包含若干上海市历史 CPI 数据，隐藏层设置为 15 个神经元，输出层只有一个神经元，即把上海市前 15 月的 CPI 数据作为输入值，第 16 月的 CPI 数据作为输出值，则 2004 年至 2024 年的月度 CPI 数据最终会滑动生成 226 个训练样本。通过训练和测试数据集的划分，对网络进行训练，并不断调整权值，直至达到预设的学习精度。最后，利用训练好的网络对未来的 CPI 进行预测。

图 7、图 8 揭示了 BP 神经网络模型预测值与数据真实值之间的误差对比分析。可以看出 BP 的拟合效果较好，可以将其纳入组合模型的选择中。

4.4. 单一模型效果对比

从模型的效果来看，ARIMA 模型、LSSVM 回归模型、BP 神经网络模型的 RMSE 和 MAPE 对比效果如表 4。

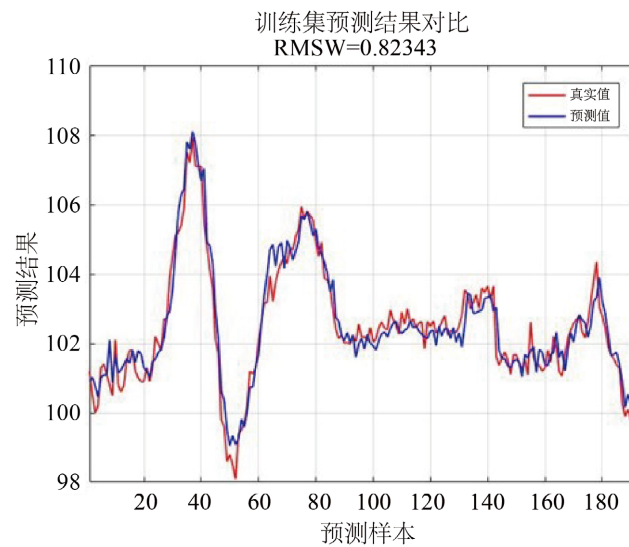


Figure 7. BP model fitting effect on the training set
图 7. BP 模型拟合训练集效果图

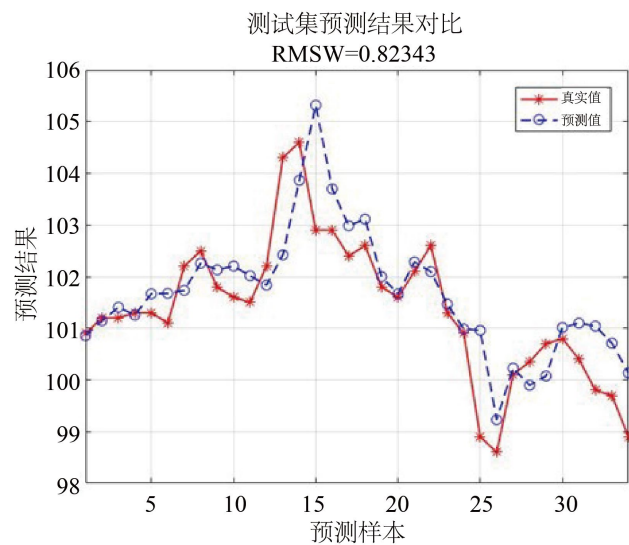


Figure 8. BP model fitting effect on the test set
图 8. BP 模型拟合测试集效果图

Table 4. Comparison of errors among single models
表 4. 单一模型误差对比

模型	RMSE	MAPE
ARIMA 模型	0.51	0.0045
LSSVM 模型	0.69	0.0061
BP 神经网络模型	0.82	0.0142

根据表 4，分析单一模型对比效果，可知在三种单一模型中，ARIMA 模型对原始序列的信息提取最为充分，其 RMSE 为 0.51，MAPE 为 0.0045。相较而言，LSSVM 回归模型预测误差较小，预测结果较

为优良，其 RMSE 为 0.69，MAPE 为 0.0061。鉴于 BP 神经网络对样本具有较强依赖性，需充足样本进行训练，尽管其中期预测方面表现较为优秀，但从长期和短期效果来看，其预测效果却不如 LSSVM 模型和 ARIMA 模型，其 RMSE 为 0.82，MAPE 为 0.0142。综合考虑，单一模型中 ARIMA 模型最优，BP 神经网络模型的预测效果相较于其他两种单一模型略显不足。

5. 基于组合模型的上海市 CPI 预测

5.1. 基于 ARIMA-BP 组合模型的预测分析

ARIMA 和 BP 神经网络模型在 CPI 数据的短期预测中均表现出一定的有效性，为我们提供了有价值的参考信息，然而也存在明显的不足之处。具体而言，BP 神经网络模型在预测方面的表现并不理想，其预测结果呈现出较大的不稳定性。相对而言，ARIMA 模型的预测效果则更为出色，其预测相对误差低于 1%，展现出了较高的预测精度。与一般的预测模型相比，ARIMA 模型在 CPI 数据的短期预测上确实表现不俗。

综上所述，ARIMA 模型和 BP 神经网络在 CPI 数据的短期预测中各有优劣。ARIMA 模型凭借其稳定的线性特征提取能力，在预测精度上更胜一筹；而 BP 神经网络虽然在非线性特征提取方面表现出色，但在预测稳定性上还有待提升。在实际应用中，我们可以根据数据的具体特点和预测需求，选择合适的模型进行预测。

组合模型是一种创新性的建模方法，它巧妙地将 ARIMA 模型和 BP 神经网络模型相结合，以实现和时间序列数据的更精确预测。上海市 CPI 数据兼具线性和非线性特征。

- (1) 针对线性部分我们采用 ARIMA 模型进行预测，并已在上文得到预测结果。
- (2) 非线性部分则利用残差序列来表示，并构建 BP 神经网络模型对其进行预测，得到新的残差预测结果。
- (3) 将 ARIMA 模型预测的线性部分与 BP 神经网络模型预测的残差部分相加将，得到组合模型的预测结果。

通过以上三个步骤，我们构建了一个具有较高预测精度的组合模型。在实际应用中，可以根据数据特点和需求灵活调整 ARIMA 模型和 BP 神经网络模型的参数，以实现更好的预测效果。总之，组合模型为时间序列预测提供了一种有效的方法，有望在各类领域取得良好的应用前景。

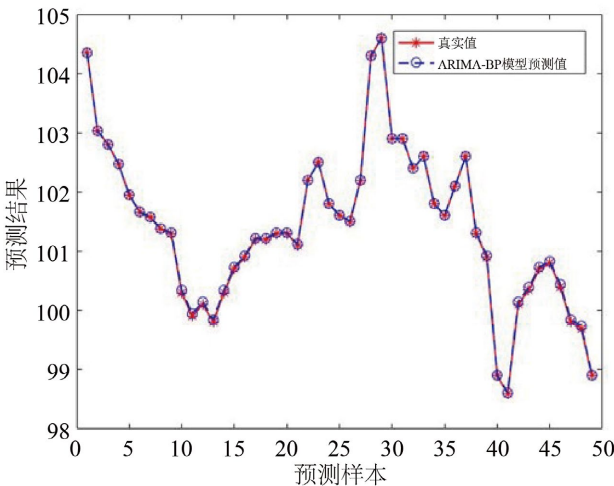


Figure 9. Fitting effect chart of the ARIMA-BP model
图 9. ARIMA-BP 模型拟合效果图

利用构建完成的 ARIMA-BP 组合模型, 对未来数据进行了预测分析, 并将组合模型的预测结果与基础模型数据进行了详尽的对比, 对比结果如表 5 所示。

Table 5. Prediction error of the ARIMA-BP model

表 5. ARIMA-BP 模型误差

日期	预测值	真实值	误差	RMSE	MAPE
2023-07	100.33	100.35	0.02		
2023-08	100.76	100.7	-0.06		
2023-09	100.86	100.8	-0.06		
2023-10	100.40	100.4	0	0.04	0.0003
2023-11	99.83	99.8	-0.03		
2023-12	99.74	99.7	-0.04		
2024-01	98.87	98.9	0.03		

首先, 我们分析一下图 9 中的两条曲线以及表 5 的具体误差。通过仔细观察, 我们可以发现这两条曲线的走势相当接近, 这种现象并非巧合, 而是与生成组合模型的算法密切相关。组合模型是通过将多个模型结合在一起, 以提高预测精度。

尽管整体上看, 预测线基本保留了原曲线的走势, 但在整个过程中, 预测值与实际值之间始终存在一定的偏差。ARIMA 模型和 BP 神经网络模型的组合有效地降低了这种偏差。

通过以上分析, 我们可以得出结论: ARIMA-BP 模型在预测复杂动态数据时具有较高的精确度和稳定性, 可以有效地降低预测值与实际值之间的偏差。因此, 在实际应用中, 采用组合模型进行预测具有较高的实用价值。

5.2. 基于 ARIMA-LSSVM 组合模型的预测分析

在组合模型领域, 权值的选择被视为核心环节。这是因为各类单一模型在提取序列信息时具有差异性, 如何合理地确定权重, 从而提高组合模型的预测精度, 成为关键性问题。

假设观测数据为 $X = (x_1, x_2, \dots, x_n)$, 选择了 k 个预测模型对数据进行建模, 那么这 k 个预测模型 $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$ 在 t 时刻的预测值记为 $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t)\}$ 。设其权重的向量 $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ 是组合模型的权重, 因此, 组合预测模型的表达式为:

$$\varphi(t) = \lambda_1 \varphi_1(t) + \lambda_2 \varphi_2(t) + \dots + \lambda_k \varphi_k(t) = \sum_{i=1}^k \lambda_i \varphi_i(t) \quad (7)$$

根据上文 ARIMA 和 LSSVM 回归模型与原始数据的不同拟合程度, 我们将给予两者不同的权重, 鉴于 ARIAM 模型和 LSSVM 模型的预测结果以及相应的均方根误差和平均绝对误差, 根据平均绝对误差我们得到 MAPE 系数组合法的权重 $a = 0.43$, $a = 0.87$, 可知, MAE 系数组合法的公式为:

$$\varphi(t) = \lambda_1 \varphi_1(t) + \lambda_2 \varphi_2(t) \quad (8)$$

利用构建完成的 ARIMA-LSSVM 组合模型, 对未来数据进行了预测分析, 并将组合模型的预测结果与基础模型数据进行了详尽的对比, 对比结果如表 6 所示。

Table 6. Prediction error of the ARIMA-LSSVM model
表 6. ARIMA-LSSVM 模型误差

日期	预测值	真实值	误差	RMSE	MAPE
2023-07	100.1	100.35	0.025	0.61	0.0051
2023-08	100.4	100.7	0.3		
2023-09	100.5	100.8	0.3		
2023-10	100.6	100.4	-0.2		
2023-11	100.9	99.8	-1.1		
2023-12	100.4	99.7	-0.7		
2024-01	99.7	98.9	-0.8		

观察发现该组合模型在预测准确度上具有较高的稳定性。七个预测值几乎都在 1.0 以内，且预测结果波动较小，具有良好的预测效果。这种稳定性在很大程度上提高了组合模型的实用价值。

从中长期的预测结果来看，组合模型在长期的拟合效果非常好。这表明，在解决实际问题时，组合模型能够更好地适应时间序列的变化趋势。

综上所述，ARIMA-LSSVM 组合模型在预测精度、长期预测效果以及稳定性等方面具有显著优势。因此，在实际应用中，可以考虑将 ARIMA-LSSVM 组合模型作为一种有效的预测工具，为决策者提供有力支持。同时，在选择模型时，应根据具体问题的特点，充分比较各类模型的优缺点，以实现更精确、更稳定的预测效果。

5.3. 组合模型效果对比

从模型的效果来看，ARIMA-BP 模型、ARIMA-LSSVM 模型的 RMSE 和 MAPE 对比效果如表 7。

Table 7. Comparison of errors among combined models
表 7. 组合模型误差对比

模型	RMSE	MAPE
ARIMA-BP 模型	0.04	0.0003
ARIMA-LSSVM 模型	0.61	0.0051

根据表 4 与表 7 的结果可知，通过对比各类单一模型以及组合模型的预测效果，我们发现在两种组合模型中，ARIMA-BP 模型将原始序列的信息分解为线性自相关主体和非线性残差两部分，预测结果最为精确，其 RMSE 为 0.04，MAPE 为 0.0003，模型拟合效果卓越。

值得注意的是，ARIMA 模型和 BP 神经网络模型生成的拟合值不仅“记忆”了原数据的发展趋势，而且与单一模型相比，偏差显著降低。尤其是在曲线的结尾部分，拟合效果更为理想。这表明，该组合模型在处理复杂和非线性数据时具有更高的准确性和稳定性。

相较而言，ARIMA-LSSVM 模型通过赋权组合的方式预测误差较大，其 RMSE 为 0.61，MAPE 为 0.0051。在均方根误差(RMSE)和平均绝对百分比误差(MAPE)方面，ARIMA-LSSVM 组合模型虽然略高于单一的 ARIMA 模型，但相较于其他单一模型而言，其数值较低。这表明 ARIMA-LSSVM 组合模型在预测精度上具有一定的优势，其预测效果在整体上优于其他单一模型。

综上所述，组合模型在复杂情况下的预测效果优于单一模型，其中 ARIMA-BP 模型优于 ARIMA-

LSSVM 模型。

6. 总结与研究展望

6.1. 总结

本文运用三种单一(ARIMA 模型、LSSVM 回归模型、BP 神经网络模型),以及两种组合模型(ARIMA-LSSVM 模型和 ARIMA-BP 模型)在对上海市 CPI 时间序列预测上都表现出较好的拟合能力。

当时间序列展现出线性趋势时,ARIMA 模型通常能够取得良好的拟合效果;对于具有非线性趋势的时间序列,LSSVM 与 BP 模型相较于 ARIMA 模型更为有效。在本文中,我们采用 ARIMA 模型对上海市 CPI 的线性部分进行拟合,对于非线性部分则利用 LSSVM 回归模型、BP 神经网络模型进行拟合。通过比较发现,集成模型的效果优于单一的 BP 模型。

单一模型中 ARIMA 模型在其特定条件下表现出了优秀的预测能力,其预测效果最为出色,而 LSSVM 较为优异,但 BP 神经网络模型的预测效果相较于其他两种单一模型略显不足。而组合模型 ARIMA-BP 和 ARIMA-LSSVM 则各结合了两模型的优点,使得其在复杂情况下的预测效果优于单一模型,其中 ARIMA-BP 模型优于 ARIMA-LSSVM 模型。总体上,ARIMA-BP 模型预测表现最为卓越,预测结果最为准确。

6.2. 研究展望

在本文中,由于上海市 CPI 同比数据同时具备线性特征与非线性特征,我们采用 ARIMA-LSSVM 模型和 ARIMA-BP 模型对上海市 CPI 进行预测。然而,由于数据量有限,通过组合系数法确定权重提高预测精度存在一些问题需要进一步探讨和优化。并且评估指标较少,未来研究可以使用更多指标来对模型的预测性能进行评估。

预测效果与精度会受到样本影响,本文所选数据量相对较少,而 BP 神经网络模型通常需要大量数据来优化网络结构并达到理想训练效果。在数据有限情况下,采用 BP 神经网络对上海市 CPI 数据进行预测可能引入较大误差,导致预测效果不佳,未来的研究可以考虑采用改进的 BP 算法和小波分析等方法,以进一步提升模型的性能和准确性。

最后,本文主要关注 CPI 数据序列本身信息,而未考虑可能影响 CPI 序列值波动的外部因素。例如,政策变动、原油价格波动等市场不稳定因素可能对 CPI 产生影响。因此,未来研究应将这些外部因素纳入考虑,以实现更全面、深入的 CPI 研究,提高研究意义和现实价值。

6.3. 非技术能力达成

在当前社会经济环境下,消费者价格指数(CPI)是衡量物价水平变动的重要指标,在推进 CPI 组合预测分析的研究和应用过程中,必须严格遵守相关法律法规和政策要求,确保工作的严谨性、规范性和准确性。要严格遵守数据保密规定,确保数据使用的合法性和合规性;要尊重个体权益,避免不当披露;要遵循伦理规范,保护当事人隐私,维护社会稳定、保障民生福祉作出积极贡献。

基金项目

上海工程技术大学(SUES)国家级大学生创新训练计划(项目编号:202410856038)。

参考文献

- [1] 蒋泽迪. 基于时间序列分析的我国房地产业经济增长预测模型研究[D]: [硕士学位论文]. 苏州: 苏州科技大学, 2018.

-
- [2] Mwanga, Y. (2020) Arima Forecasting Model for Uganda's Consumer Price Index. *American Journal of Theoretical and Applied Statistics*, **9**, 238-244. <https://doi.org/10.11648/j.ajtas.20200905.17>
 - [3] Nyoni and Thabani. (2019) ARIMA Modeling and Forecasting of Consumer Price Inde (CPI) in Germany. MPRA Paper, 1-13.
 - [4] Wedding, D.K. and Cios, K.J. (1996) Time Series Forecasting by Combining RBF Networks, Certainty Factors, and the Box-Jenkins Model. *Neurocomputing*, **10**, 149-168. [https://doi.org/10.1016/0925-2312\(95\)00021-6](https://doi.org/10.1016/0925-2312(95)00021-6)
 - [5] 朱颜杰, 樊顺厚, 雷怀英. 基于 SARIMA 模型的居民消费价格指数的预测[J]. 统计与决策, 2014, 30(4): 37-39.
 - [6] 蔡敏. 基于 SARIMA 和 BP 神经网络模型的江苏省 CPI 预测分析[D]: [硕士学位论文]. 桂林: 广西师范大学, 2022.
 - [7] 黄树花. ARIMA 与 BP 模型在我国居民消费价格指数的实证分析[D]: [硕士学位论文]. 昆明: 云南大学, 2015.
 - [8] 姚金海, 邹家骏. CPI 预测的 SVM-ARIMA 模型构建与数值模拟[J]. 统计与决策, 2022, 38(21): 48-52.
 - [9] 闫泓水. 基于 EEMD-SARIMA-LSTM 混合模型对 CPI 指数预测[D]: [硕士学位论文]. 济南: 山东大学, 2020.
 - [10] 郑雪平, 水庆象. 基于分段线性 Lagrange 插值 Grey-Markov 模型的 CPI 预测[J]. 统计与决策, 2020, 36(14): 29-32.
 - [11] 潘静, 张颖, 刘璐. 基于 ARIMA 模型与 GM(1,1)模型的居民消费价格指数预测对比分析[J]. 统计与决策, 2017(20): 110-112.
 - [12] 许明燕. 基于 ARIMA 模型和 BP 神经网络模型的江苏省 GDP 预测分析[D]: [硕士学位论文]. 济南: 山东大学, 2020.
 - [13] 彭乃驰, 党婷. 基于小波分析的 BP-SARIMA 模型的 CPI 预测[J]. 统计与决策, 2018, 34(16): 225.
 - [14] 唐玉辉. 基于百度搜索指数的居民消费价格指数预测研究[D]: [硕士学位论文]. 烟台: 山东工商学院, 2022.
 - [15] 应虹存. 基于网络搜索数据的 CPI 组合预测研究[D]: [硕士学位论文]. 北京: 北方工业大学, 2023.
 - [16] 韩春蕾, 高婉君. 我国月度 CPI 的组合预测及分析[J]. 统计与决策, 2014(1): 11-13.
 - [17] 朱文燕. 组合预测模型在 CPI 预测中的应用[D]: [硕士学位论文]. 广州: 暨南大学, 2017.
 - [18] 毛钰嘉. 基于时间序列的组合预测模型研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2020.