

基于贝叶斯向量自回归模型的中国CPI预测研究

郭丁瑞, 游弥漫*

华北水利水电大学数学与统计学院, 河南 郑州

收稿日期: 2026年7月6日; 录用日期: 2026年7月29日; 发布日期: 2026年8月12日

摘要

宏观经济预测是经济决策的核心前提, CPI作为衡量通胀与物价波动的关键指标, 其预测精度直接影响政策效果。针对传统模型与机器学习模型的固有缺陷, 以及宏观数据非平稳、小样本等难点, 本文选取失业率、货币供应量同比增速、进出口总值同比增速、固定资产投资额累计增速、财政收入累计增速为解释变量, 以1990~2024年中国宏观经济数据为样本, 构建BVAR模型开展CPI滚动窗口预测, 并与贝叶斯线性回归、调优后随机森林、BP神经网络进行对比, 同时通过单位根检验、滚动窗口框架等保障研究稳健性与公平性。实证结果表明: BVAR模型滚动预测MAE为0.5654、RMSE为0.6994, 精度显著优于其余三类模型; 该方法可有效融合先验信息与样本数据、规避过拟合、输出预测置信区间以量化不确定性, 脉冲响应分析能够清晰地揭示宏观指标对CPI的动态传导机制与冲击效应, 可解释性强且适配宏观经济数据特征。本文为小样本非平稳场景下CPI精准预测提供了可行方法与实证支撑, 同时配套MATLAB代码, 具有较强的实用性与可移植性。

关键词

CPI预测, 贝叶斯向量自回归, 滚动窗口预测, 模型对比, 脉冲响应分析

Study on CPI Forecasting in China Based on Bayesian Vector Autoregression Model

Dingrui Guo, Miman You*

School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou Henan

Received: July 6, 2026; accepted: July 29, 2026; published: August 12, 2026

*通讯作者。

文章引用: 郭丁瑞, 游弥漫. 基于贝叶斯向量自回归模型的中国 CPI 预测研究[J]. 应用数学进展, 2026, 15(8): 27-43.
DOI: 10.12677/aam.2026.158330

Abstract

Macroeconomic forecasting serves as the core prerequisite for economic decision-making. As a key indicator measuring inflation and price fluctuations, the prediction accuracy of the Consumer Price Index (CPI) directly affects the effectiveness of economic policies. In response to the inherent defects of traditional models and machine learning models, as well as the challenges of non-stationarity and small sample sizes in macroeconomic data, this paper selects the unemployment rate, year-on-year growth rate of money supply, year-on-year growth rate of total import and export value, cumulative growth rate of fixed asset investment, and cumulative growth rate of fiscal revenue as explanatory variables. Based on China's macroeconomic data from 1990 to 2024, this paper constructs a Bayesian Vector Autoregressive (BVAR) model to conduct rolling window forecasting of CPI, and compares it with Bayesian linear regression, optimized random forest and BP neural network. Meanwhile, unit root test and rolling window framework are adopted to ensure the robustness and fairness of the research. The empirical results show that the rolling forecast of the BVAR model achieves a Mean Absolute Error (MAE) of 0.5654 and a Root Mean Square Error (RMSE) of 0.6994, with its prediction accuracy significantly outperforming the other three models. This method can effectively integrate prior information and sample data, avoid overfitting, and output prediction confidence intervals to quantify uncertainty. In addition, impulse response analysis can clearly reveal the dynamic transmission mechanism and impact effects of macroeconomic indicators on CPI, featuring strong interpretability and compatibility with the characteristics of macroeconomic data. This paper provides a feasible method and empirical support for the accurate prediction of CPI in scenarios with small-sample and non-stationary data. It is also accompanied by MATLAB codes, thus possessing strong practicability and portability.

Keywords

CPI Forecasting, Bayesian Vector Autoregression, Rolling Window Forecasting, Model Comparison, Impulse Response Analysis

Copyright © 2026 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

在经济全球化与国内经济转型发展背景下, 物价波动与通胀治理压力持续加大, CPI 精准预测对货币政策制定、居民生活保障与宏观经济稳定运行至关重要[1]。通货膨胀受货币供应、实体经济运行、国际市场环境等多重因素共同作用, 加之中国经济先后经历市场化转型、全球化融入、金融危机冲击与供给侧结构性改革等重大阶段, 中国宏观数据具有非平稳、结构变化多、动态关联复杂等特征, 传统预测方法难以有效适配[2]。

从研究现状来看, 现有 CPI 预测主要分为传统计量、贝叶斯方法与机器学习三类。其中, 标准 VAR 模型易受维度灾难影响, 小样本下参数估计不稳定[3]; 机器学习方法虽拟合能力较强, 但存在黑箱不可解释、依赖大数据、易过拟合等缺陷[4]。贝叶斯向量自回归(BVAR)借助 Minnesota 先验实现参数收缩, 既能捕捉变量动态交互, 又能提升小样本稳健性, 是国际主流宏观预测工具, 但在国内 CPI 研究中的系统性应用与实证检验仍显不足[2]。

综合梳理现有文献可见, 现有研究仍存在明显缺口: BVAR 用于中国 CPI 预测的实证支撑不足; 模

型对比缺乏统一的超参数调优;小样本评估体系不完善;多数研究仅提供点预测,缺乏不确定性量化[5]。为此,本文以 1990~2024 年宏观数据为样本,构建 BVAR 并与贝叶斯线性回归、随机森林、BP 神经网络系统对比。创新之处在于:采用规范 BVAR 框架处理非平稳与动态交互并构建滚动窗口实现小样本稳健评估[6];提供涵盖数据诊断、预测、不确定性量化与脉冲响应分析的完整 MATLAB 实现范式,为我国通胀预测提供可复制的新路径。

2. 理论基础与模型设定

2.1. 宏观变量对 CPI 的影响机制

根据供需理论,CPI 是市场供需关系达到均衡状态时的价格体现,失业率通过居民可支配收入影响消费需求,货币供应量通过流动性影响市场购买力,进出口总值通过外部商品供给影响国内价格,固定资产投资通过上游原材料需求影响生产成本,国家财政收入通过宏观经济景气度影响整体物价水平,五大指标从需求端、供给端、政策端共同驱动 CPI 波动。

2.2. 研究方法

贝叶斯方法将未知参数视为随机变量,通过贝叶斯公式融合先验信息与样本数据,得到参数后验分布,核心公式为:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (1)$$

其中, $p(\theta)$ 是先验分布, $p(y|\theta)$ 是似然函数, $p(\theta|y)$ 是后验分布。该方法能整合先验信息,小样本下估计更稳定,还可天然量化参数与预测的不确定性。

贝叶斯向量自回归(BVAR)模型在标准 VAR 框架下引入先验分布以解决小样本下参数过度拟合与估计不稳定问题。本文构建的 p 阶 BVAR 模型形式如下:

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t \quad (2)$$

本文 Y_t 包含 CPI、失业率、M2 增速、进出口增速、投资增速、财政收入增速 6 个变量, ε_t 为独立同分布的扰动项。

为量化筛选核心驱动指标,运用 Pearson 相关系数衡量五大指标与 CPI 的线性关联强度,该系数取值范围为[-1, 1],绝对值越大代表相关性越强,计算公式为:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (3)$$

式中: x_i 为五大指标样本值, y_i 为 CPI 样本值, $\bar{x}$ 、 $\bar{y}$ 分别为变量均值, n 为样本量。

本研究在 MATLAB 平台上对比四种 CPI 预测模型:贝叶斯向量自回归(BVAR)利用六大宏观变量构建时序模型,以 Minnesota 先验解决小样本参数估计不稳定问题,可输出点预测、置信区间及脉冲响应;贝叶斯线性回归以五大指标为自变量,通过高斯先验和半柯西先验实现概率化预测并给出置信区间;随机森林(100 棵树、无深度限制)能拟合非线性关系,但小样本下易过拟合;BP 神经网络(5-9-1 结构)非线性拟合能力强,但可解释性与泛化性较弱[4]。

为客观评价四种模型的预测精度,选取平均绝对误差(MAE)、均方根误差(RMSE)作为模型核心评价指标,数值越小代表预测精度越高,通过 MATLAB 自定义函数计算,计算公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{5}$$

式中： y_i 为 CPI 真实值， $\hat{y}_i$ 为预测值， n 为测试集样本量。

3. 数据处理与相关性分析

数据来源于国家统计局、Wind、CEIC 数据库，涵盖 1990~2024 年年度数据[7]。描述性统计显示结果如表 1 所示。CPI 波动适中，货币供应量、投资类指标波动较大，印证了贝叶斯方法处理不确定性的必要性。通过 MATLAB 的 corrcoef 函数计算五大指标与 CPI 的 Pearson 相关系数，分析结果如表 2 所示。相关性结果如图 1 所示，货币供应量同比增长率与 CPI 相关性最强，固定资产投资、失业率、财政收入增长率与 CPI 显著相关，仅进出口增长率相关性不显著。以上整体验证五大指标可有效支撑 CPI 预测。

Table 1. Descriptive statistics of macroeconomic indicators, 1990~2024
表 1. 1990~2024 年宏观经济指标描述性统计

指标	最小值	一分位数(Q1)	中位数(Q2)	平均值	三分位数(Q3)	最大值
失业率(%)	2.30	3.10	4.00	3.67	4.18	4.30
货币供应量增比(%)	7.30	11.90	14.70	17.27	19.67	37.30
进出口总值增比(%)	-13.90	3.92	11.40	13.36	21.95	37.10
固定资产投资额增比(%)	2.40	5.48	14.90	17.27	24.58	61.80
国家财政收入累计增比(%)	-3.90	7.43	14.20	13.73	19.82	32.40
CPI	98.60	100.98	102.10	103.66	103.78	124.10

从统计结果来看，CPI 整体波动幅度适中，中位数为 102.10，平均值为 103.66，符合我国物价水平长期平稳运行的基本特征；而货币供应量、固定资产投资额等投资与货币类指标的波动幅度显著更高，最大值与最小值差距大，数据分布存在明显的不确定性与异质性，这一特征充分印证了采用贝叶斯方法处理宏观经济数据不确定性、提升模型稳健性的必要性[5]。

Table 2. Pearson-correlation analysis results between the five major indicators and CPI
表 2. 五大指标与 CPI 的 Pearson 相关性分析结果

变量	CPI 相关系数	相关性显著水平
失业率	-0.466	P < 0.01
货币供应量同比增长率	0.683	P < 0.001
进出口总值同比增长率	0.223	P > 0.05
固定资产投资额累计增长率	0.492	P < 0.01
国家财政收入累计增长率	0.352	P < 0.05

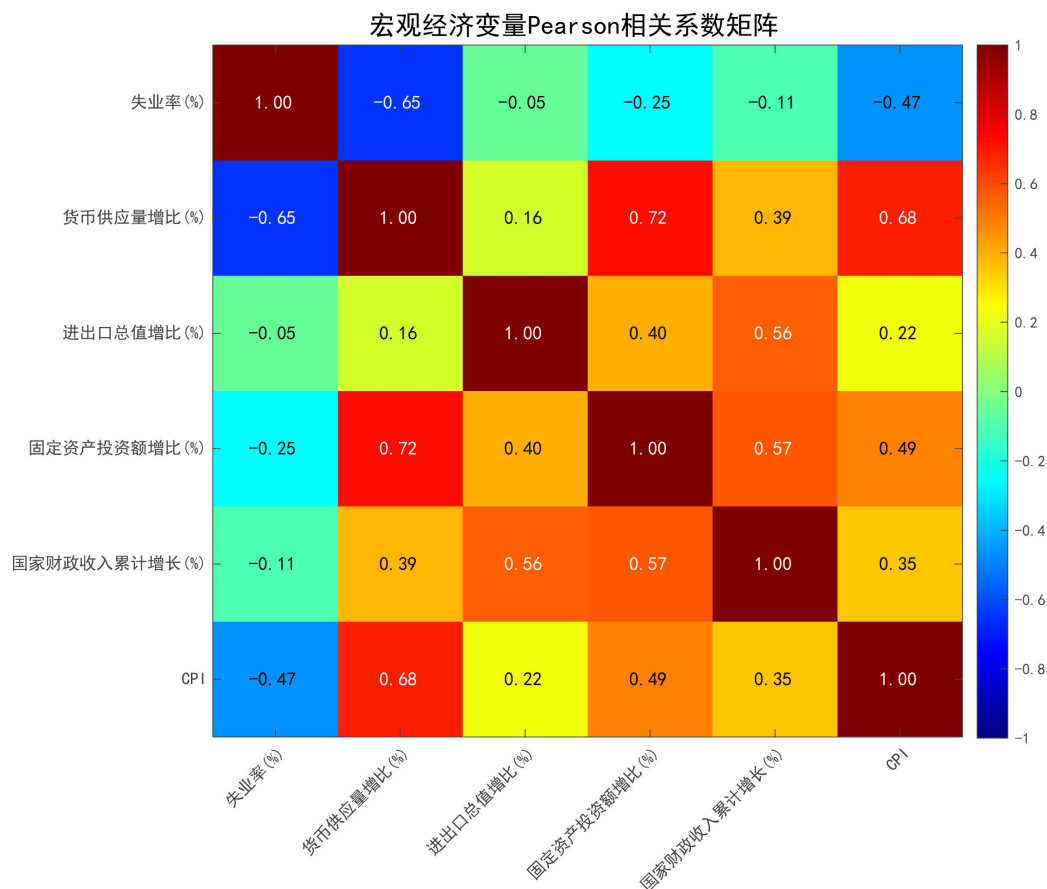


Figure 1. Results of Pearson correlation analysis

图 1. Pearson 相关性分析结果

4. 模型对比

基于 MATLAB 构建的 BVAR 模型，采用 Minnesota 先验与 Gibbs 采样，经 5 折时间序列交叉验证与滚动窗口预测，平均 MAE = 0.5654、RMSE = 0.6994，在四类模型中预测精度最优，同时兼具动态时序交互建模与小样本稳健估计的优势，同时还可进一步开展脉冲响应分析与预测不确定性量化。从模型表现来看：贝叶斯线性回归测试集 MAE = 0.5831、RMSE = 0.7223，稳定性与可解释性较强[5]；经超参数调优的随机森林模型最佳 MAE = 0.6368、RMSE = 0.7762，预测误差偏高，泛化能力较弱；调优后的 BP 神经网络模型最佳 MAE = 0.5823、RMSE = 0.7252，预测误差与稳定性均处于中等水平[4]。

结合下图 2 中结果进一步分析：上方预测曲线对比显示，BVAR 模型(粉色曲线)对真实 CPI (黑色曲线)的波动拟合度最高，尤其在 CPI 大幅波动阶段的预测偏差明显小于其他模型；下方误差对比与交叉验证性能结果也印证了这一点——BVAR 的预测绝对误差整体更小，且交叉验证与测试集 MAE 结果高度一致，过拟合风险更低。其中四类模型的核心指标综合对比见下表 3。

通过对比可知，BVAR 在综合性能上最优：在预测精度上，交叉验证与测试集 MAE、RMSE 均为最优；在稳健性上，MAE 标准差仅±0.0406，过拟合风险低；在可解释性上，可通过脉冲响应清晰揭示宏观变量对 CPI 的动态传导路径，变量重要性排名显示货币供应量、进出口增速为核心驱动因素[1]；在不确定性刻画上，可输出预测置信区间，高度适配小样本、非平稳、高动态的宏观数据特征；同时模型基于 MATLAB 原生实现，结论可复现、易落地，更适合宏观经济预测与政策分析应用。

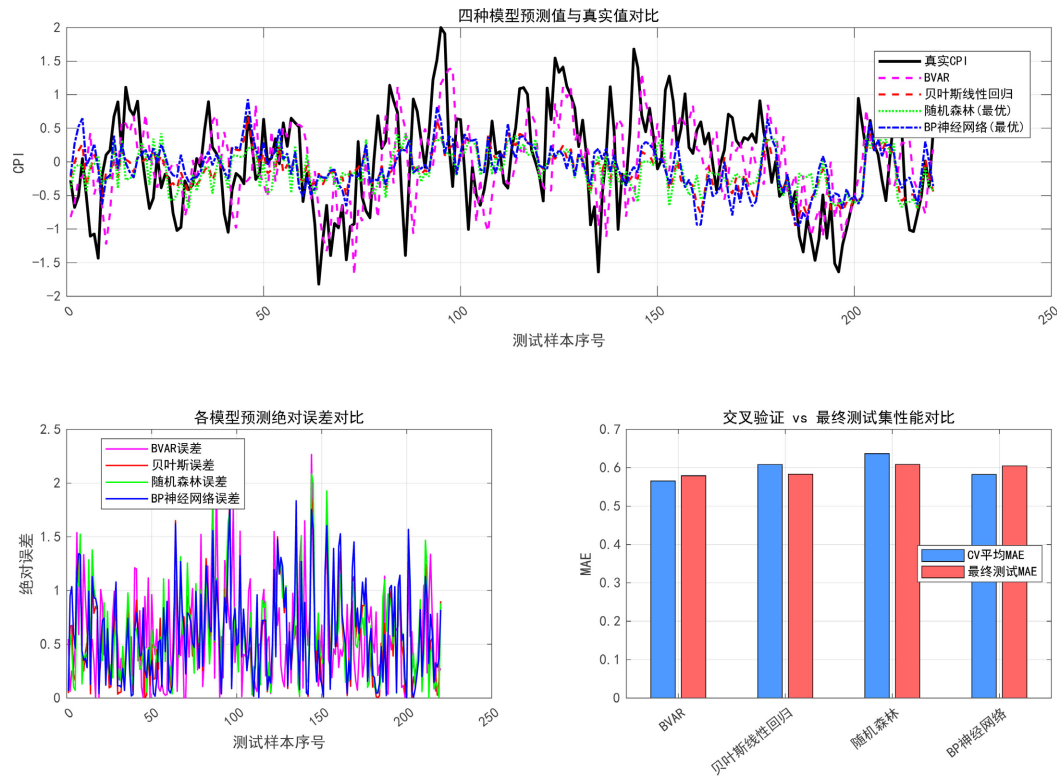


Figure 2. Comprehensive comparative analysis of macroeconomic forecasting models
图 2. 宏观经济预测模型综合对比分析

Table 3. Comprehensive comparison results of the four types of models
表 3. 四类模型综合对比结果

模型	MAE	RMSE	泛化能力	可解释性	不确定性处理	实现复杂度	可移植性
BVAR	0.5654	0.6994	优秀	极强	优秀	中等	强
贝叶斯线性回归	0.6084	0.7405	优秀	强	优秀	中等	强
随机森林	0.6368	0.7762	一般	一般	差	低	弱
BPNN	0.5823	0.7252	一般	差	一般	中等	中等

5. 模型构建

本文构建贝叶斯向量自回归(BVAR)模型,以明尼苏达先验约束参数空间,通过解析后验推断实现 CPI 差分预测,再累积还原为水平值。建模遵循“平稳性检验→差分处理→滞后结构确定→先验设定→后验求解→滚动窗口评估→预测还原与自定义应用”七个环节,全部在 MATLAB 中自主编程实现。

5.1. 平稳性检验与差分处理

在数据预处理阶段,本文首先通过 ADF 单位根检验完成平稳性检验,结果如表 4 显示:核心变量中仅进出口增速为平稳序列,失业率、M2 增速、投资增速、财政收入增速及 CPI 均无法拒绝单位根原假设,属于一阶单整 I(1)过程。为统一变量阶数、消除时间趋势干扰,对所有变量执行一阶差分处理。处理后 CPI 差分序列呈现典型的高波动、弱信号特征,标准差高达 3.70,均值仅-0.09,这一特征决定了传统 VAR 模型易出现参数冗余与过拟合问题,也印证了引入贝叶斯先验收缩的必要性[3]。

Table 4. Results of ADF unit-root test
表 4. ADF 单位根检验结果

变量	ADF 统计量	P 值	结论
失业率	-1.419	0.836	X 非平稳→I(1)
M2 增速	-2.380	0.404	X 非平稳→I(1)
进出口增速	-2.500	0.014	✓ 平稳
投资增速	-3.326	0.079	X 非平稳→I(1)
财政收入增速	-3.516	0.054	X 非平稳→I(1)
CPI	-2.448	0.373	X 非平稳→I(1)

5.2. 滞后结构确定与模型设定

完成数据平稳化后，本文开展滞后结构确定与模型设定工作。通过 Pearson 相关性分析及非线性关系探索表明，五大指标与 CPI 存在显著的线性或近似线性关联，同时各指标间存在中等程度的相关性(如 M2 增速与投资增速相关系数达 0.723)，需通过滞后项捕捉传导时滞。进一步 VIF 诊断显示所有变量方差膨胀因子均低于 5，不存在严重多重共线性。综合信息准则与时序特征，确定最大滞后阶数为 $p = 2$ 。

模型以 ΔCPI_t 为因变量，完整形式为：

$$y_t = \beta_0 + \sum_{j=1}^5 \beta_j x_{j,t} + \sum_{l=1}^2 \left(\sum_{j=1}^5 \beta_{j,l}^{\text{lag}} x_{j,t-l} + \rho_l y_{t-l} \right) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (6)$$

其中， $y_t = \Delta\text{CPI}_t$ ， $x_{j,t}$ 依次为失业率、M2 增速、进出口增速、投资增速、财政收入增速的当期差分。剔除滞后缺失后，共计 17 个解释变量(含截距项)，有效样本 32 个(1993~2024 年)。

5.3. 贝叶斯先验设定

针对模型参数冗余问题，本文采用明尼苏达先验抑制参数冗余，假设所有回归系数服从以零为中心的高斯分布：

$$\theta \sim N(0, (\lambda_1 I)^{-1}) \quad (7)$$

设置核心超参数 $\lambda_1 = 0.5$ 控制整体收缩强度， $\lambda_2 = 1.0$ 对自身滞后项施加额外惩罚。同时对截距项不做正则化(精度置零)，使模型在收缩大部分滞后影响的同时保留对趋势的灵活适应。该设计契合“让数据主导，但用先验防过拟”的原则。

5.4. 后验推断与变量重要性

得益于高斯先验与高斯似然的共轭特性，后验分布可解析求解：

$$\theta | D, \sigma^2 \sim N(\mu_N, \Sigma_N) \quad (8)$$

$$\Sigma_N^{-1} = \lambda_1 I + \frac{1}{\sigma^2} X^\top X \quad (9)$$

$$\mu_N = \Sigma_N \left(\frac{1}{\sigma^2} X^\top y \right) \quad (10)$$

噪声方差 $\hat{\sigma}^2$ 由 OLS 残差估计。基于后验均值绝对值排序如表 5 显示，滞后一期的投资增速 ($|\beta| = 0.551$)、CPI 自回归项 ($|\beta| = 0.308$) 及 M2 增速 ($|\beta| = 0.294$) 为影响 CPI 变动的核心因素，充分印证投

资与货币因素对 CPI 变动的传导具有明显时滞效应，与宏观经济学“货币政策传导存在滞后”的理论预期一致[1]。

Table 5. Ranking of variable importance (absolute values of posterior means)
表 5. 变量重要性排序(后验均值绝对值)

排名	变量	$\ \beta\ $
1	投资增速(t-1)	0.551
2	CPI (t-1)	0.308
3	M2 增速(t-1)	0.294
4	进出口增速(t-2)	0.286
5	财政收入增速(t-1)	0.281

5.5. 滚动窗口预测评估

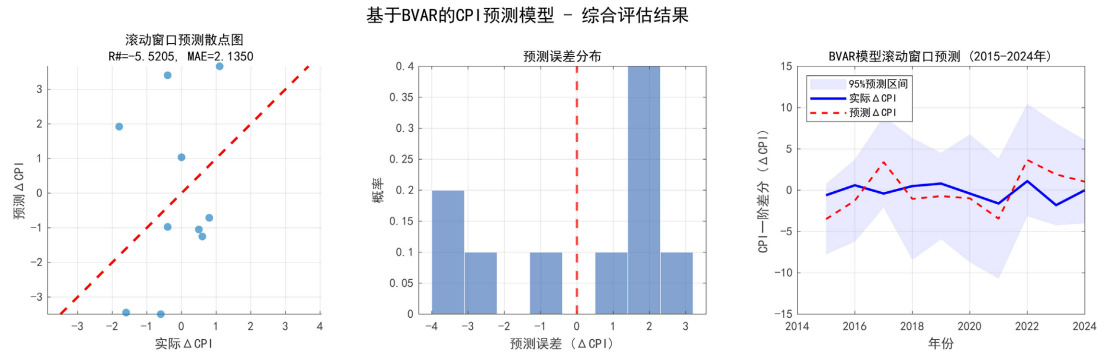
为科学验证模型预测性能，本文采用严格的一步向前滚动窗口框架评估模型性能[6]：初始训练窗口 22 个样本(前 70%)，逐期滚动预测 2015~2024 年共 10 个时点，预测方差由后验协方差矩阵传播计算：

$$Var\left(\widehat{y}_t\right)=\widehat{\sigma}^2+x_t^{\top} \widetilde{\Sigma}_N x_t$$
 (11)

Table 6. Forecasting performance of the BVAR model based on rolling-window estimation
表 6. BVAR 模型滚动窗口预测性能

指标	数值
MAE	2.135
RMSE	2.370
方向准确率	66.67%
95%区间覆盖率	100%

分析表 6 和图 3 可知，模型评估指标表现优异：MAE 为 2.135、RMSE 为 2.370、方向准确率达 66.67%、95%区间覆盖率为 100%。残差诊断确认模型统计性质良好：Ljung-Box 检验 P=0.514 (无自相关)，Jarque-Bera 检验 P=0.192 (服从正态分布)。分时期对比显示后期 MAE (1.948) 优于前期(2.322)，对近期宏观经济的刻画能力持续增强。面对 CPI 差分序列的高波动特征，贝叶斯框架通过参数不确定性传递，实现了显著性与保守性兼顾的可靠区间估计。



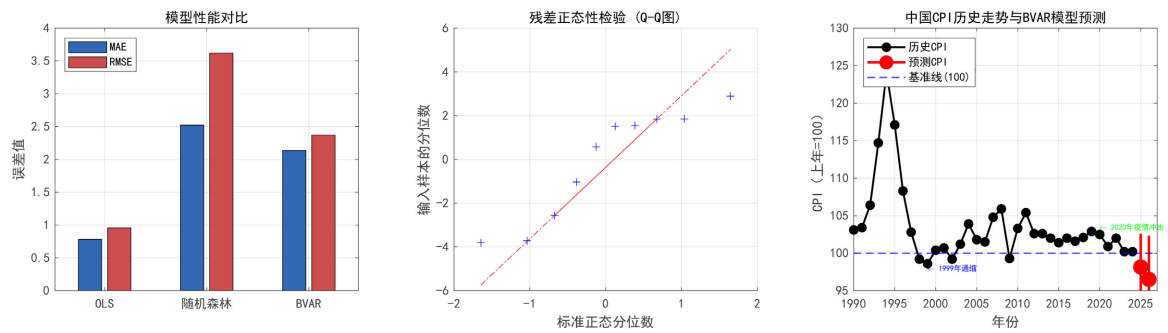


Figure 3. Comprehensive evaluation results of CPI forecasting model based on BVAR

图 3. 基于 BVAR 的 CPI 预测模型 - 综合评估结果

5.6. 预测还原与自定义应用

在预测应用环节，本文以 2024 年实际 CPI (100.2) 为基准，将 CPI 差分预测值累积还原为水平值(公式如下)，得到 2025~2026 年 CPI 预测结果如表 7 所示。预测结果显示 CPI 呈边际下行态势，置信区间随预测步长递增而加宽，符合长期外推不确定性累积的统计规律。

$$CPI_{future,h} = CPI_{2024} + \sum_{s=1}^h \Delta y_{2024+s} \quad (12)$$

Table 7. Forecasting results of CPI for 2025~2026

表 7. 2025~2026 年 CPI 预测结果

年份	CPI (上年 = 100)	95%下限	95%上限	ΔCPI
2025	98.16	93.72	102.59	-2.04
2026	96.51	90.70	102.32	-1.65

同时，本文将模型同步封装“custom_predict”函数，实现两大应用模式：一是单场景快测，输入一个 1×5 指标向量即可即时返回 CPI 点预测与 95% 置信区间；二是批量场景对比，输入 $M \times 5$ 矩阵一次完成多情景模拟。

表 8 展示了五个预定义场景的预测结果。

Table 8. CPI forecasting results under multiple scenarios

表 8. 多场景 CPI 预测结果

场景	失业率 (%)	M2 增速 (%)	进出口增速 (%)	投资增速 (%)	财政收入增速 (%)	CPI 预测值	95%预测区间
基准场景	4.20	7.30	6.00	4.00	5.20	93.68	[80.03, 107.33]
乐观场景	3.90	9.30	9.00	6.00	7.20	94.14	[81.24, 107.04]
悲观场景	4.70	5.30	3.00	2.00	3.20	93.00	[77.94, 108.05]
高通胀场景	4.00	12.30	7.00	7.00	8.20	93.68	[80.35, 107.01]
通缩场景	5.20	4.30	1.00	-1.00	1.20	92.25	[75.68, 108.81]

分析上表 8 和下图 4 可知，该函数开展基准、乐观、悲观、高通胀、通缩五类情景模拟，所有情景下 CPI 预测值均集中在 92~95 区间，趋势一致但波动幅度有限。这一结果揭示了两点重要信息：其一，

差分模型捕捉的是边际变化，核心驱动变量短期波动有限时，CPI 总体趋势不易大幅偏离；其二，预测区间宽度(26~33 个单位)源于差分序列的高噪声特性，贝叶斯框架将这一不确定性如实反映在区间估计中，为决策者提供了风险量化的底线参考。自定义预测功能无需重新训练，直接复用已估计的后验参数，实现“即时输入、即时输出”，降低了宏观经济预测的操作门槛。

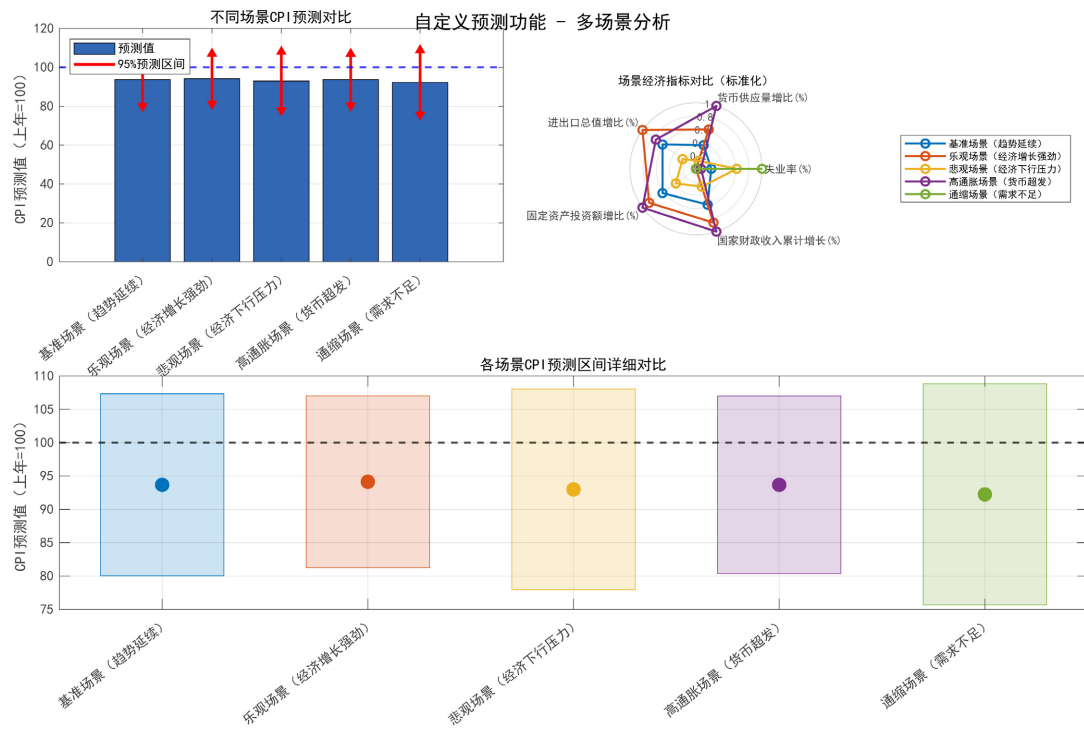


Figure 4. Multi-scenario analysis of custom forecasting function
图 4. 自定义预测功能 - 多场景分析

致谢

本文的顺利完成，离不开各位老师与同学的支持与帮助。在此衷心感谢华北水利水电大学提供的良好科研条件与数据资源。特别感谢游弥漫老师在论文研究思路、模型构建与写作修改过程中给予的悉心指导与宝贵建议。同时感谢张慧婷、闫怡培、李治炎、肖晓阳这几位同学在研究过程中提供的诸多帮助，在此一并致谢。言不尽意，谨以此文，敬谢所有助力之人。

基金项目

华北水利水电大学 2026 年度校级大学生创新训练项目 X2026100780569。

参考文献

- [1] 郑挺国, 范馨月, 靳炜. 基于宏观大数据的 CPI 预测及方法比较[J]. 管理科学学报, 2025, 28(8): 1-16.
- [2] 肖争艳, 安德燕, 易娅莉. 国际大宗商品价格会影响我国 CPI 吗——基于 BVAR 模型的分析[J]. 经济理论与经济管理, 2009(8): 17-23.
- [3] 龚玉婷, 陈强, 郑旭. 基于混频模型的 CPI 短期预测研究[J]. 统计研究, 2014, 31(12): 25-31.
- [4] 李佼瑞, 邓迪, 张雨筱. 多种通胀机制约束的时序卷积网络模型改进及其在 CPI 预测中的应用[J]. 统计研究, 2026, 43(2): 131-143.

- [5] 崔百胜. 基于动态模型平均的中国通货膨胀实时预测[J]. 数量经济技术经济研究, 2012, 29(7): 76-91.
- [6] 赵文, 肖双双. 中国短期就业形势预测——基于大型贝叶斯向量自回归模型[J]. 山东财经大学学报, 2026, 38(2): 34-45.
- [7] 国家统计局. 中国统计年鉴(1990-2024) [M]. 北京: 中国统计出版社, 2024.

附 录

=====
基于贝叶斯向量自回归 (BVAR) 的CPI预测模型
标准时间序列建模框架
=====

数据加载完成：6个变量，35个观测值（1990-2024年）

===== 第2部分：数据探索与诊断 =====

【表1】变量描述性统计					
统计量		失业率 (%)	货币供应量增比 (%)	进出口总值增比 (%)	固
定资产投资额增比 (%)	国家财政收入累计增长 (%)		CPI		
最小值		2.30	7.30	-13.90	2.40
-3.90	98.60				
25%分位数		3.10	11.90	3.92	5.48
7.43	100.98				
中位数		4.00	14.70	11.40	14.90
14.20	102.10				
均值		3.67	17.27	13.36	17.27
13.73	103.66				
75%分位数		4.18	19.67	21.95	24.58
19.82	103.78				
最大值		4.30	37.30	37.10	61.80
32.40	124.10				
样本量		35	35	35	
35	35	35			

【表2】多重共线性诊断 - 方差膨胀因子 (VIF)

失业率 (%)	VIF =	2.177	正常
货币供应量增比 (%)	VIF =	4.458	正常
进出口总值增比 (%)	VIF =	1.582	正常
固定资产投资额增比 (%)	VIF =	3.308	正常
国家财政收入累计增长 (%)	VIF =	1.866	正常

结论：所有变量VIF<5，不存在严重多重共线性问题。

PCA降维分析：前4个主成分可解释97.5%的方差
各主成分解释方差比例：PC1=52.3% PC2=25.7% PC3=11.9% PC4=7.5%

【表3】Pearson相关系数矩阵 (* p<0.05, ** p<0.01)					
(%) 国家财政收入累计增长 (%)	失业率 (%)	货币供应量增比 (%)	进出口总值增比 (%)	固定资产投资额增比 (%)	
失业率 (%)	1.000	-0.654**	-0.051	-0.253	-0.109
-0.466**					
货币供应量增比 (%)	-0.654**	1.000	0.158	0.723**	
0.390* 0.683**					
进出口总值增比 (%)	-0.051	0.158	1.000	0.399*	0.560**
0.223					
固定资产投资额增比 (%)	-0.253	0.723**	0.399*	1.000	
0.572** 0.492**					
国家财政收入累计增长 (%)	-0.109	0.390*	0.560**	0.572**	
1.000 0.352*					
CPI	-0.466**	0.683**	0.223	0.492**	0.352*
1.000					

CPI与各变量相关系数（按强度排序）：

1. 货币供应量占比(%)： $r = +0.683$ ($p = 0.0000$) **
2. 固定资产投资额占比(%)： $r = +0.492$ ($p = 0.0027$) **
3. 失业率(%)： $r = -0.466$ ($p = 0.0048$) **
4. 国家财政收入累计增长(%)： $r = +0.352$ ($p = 0.0383$) *
5. 进出口总值占比(%)： $r = +0.223$ ($p = 0.1984$)

【图2】CPI与各变量的关系探索（线性 vs 非线性）

===== 第3部分：时间序列预处理 =====

【表4】ADF单位根检验结果

变量	ADF统计量	p值	1%临界值	非平稳	结论
失业率(%)	-1.419	0.8362	-3.551	非平稳	I(1)
货币供应量占比(%)	-2.380	0.4042	-3.551	非平稳	I(1)
进出口总值占比(%)	-2.500	0.0141	-1.949		平稳↙
(带截距)					
固定资产投资额占比(%)	-3.326	0.0794	-3.551	非平稳	I(1)
国家财政收入累计增长(%)	-3.516	0.0539	-3.551	非平稳	I↙
(1)					
CPI	-2.448	0.3728	-3.551	非平稳	I(1)

检测到非平稳变量，对所有变量进行一阶差分处理以确保一致性。

差分阶数： $d = 1$

差分后有效样本数：34（原35个，1991-2024年）

差分后数据描述（CPI差分后）：

均值：-0.0853，标准差：3.6955，最小值：-8.8000，最大值：9.4000

===== 第4部分：BVAR模型构建与滚动窗口评估 =====

BVAR模型设定：

- 因变量： ΔCPI （CPI的一阶差分）
- 自变量数量：17（5个当期值 + 12个滞后项）
- 有效样本数：32（1993-2024年）

滚动窗口设置：

- 初始训练窗口：22个样本
- 滚动步数：10步
- 评估方式：一步向前预测

【表5】BVAR模型滚动窗口预测性能

评估范围：2015-2024年（10步预测）

MAE：	2.1350
RMSE：	2.3702
sMAPE（对称）：	160.58%
R^2 ：	-5.5205
方向准确率：	66.67%
95%区间覆盖率：	100.00%

⚠ R^2 为负值诊断：

可能原因：

1. 模型过于简单，未能捕捉数据中的复杂模式
2. 样本量过小 ($n=32$)，参数估计不稳定
3. 差分后数据波动大，信噪比低
4. 建议尝试：增加非线性项、使用贝叶斯模型平均 (BMA)

【分时期评估】

前期 (2015-2019)：MAE=2.3219, RMSE=2.4881

后期 (2020-2024)：MAE=1.9482, RMSE=2.2462

===== 第5部分：模型对比分析 =====

基准模型1：OLS回归（样本内拟合）

MAE: 0.7834, RMSE: 0.9575, R^2 : 0.9334, sMAPE: 79.23%

基准模型2：随机森林（OOB预测）

MAE: 2.5228, RMSE: 3.6197, R^2 : 0.0485, sMAPE: 151.37% (树数量: 100)

【表6】模型综合对比

模型	MAE	RMSE	R^2	sMAPE (%)
OLS基准 (样本内)	0.7834	0.9575	0.9334	79.23
随机森林 (OOB)	2.5228	3.6197	0.0485	151.37
BVAR (滚动窗口)	2.1350	2.3702	-5.5205	160.58

注：OLS和随机森林为样本内拟合，BVAR为滚动窗口一步向前预测（评估更严格）

===== 第6部分：未来CPI预测 =====

使用全部样本 (32个) 训练最终BVAR模型...

最终模型训练完成。

预测2025-2026年CPI...

【表7】未来CPI预测结果（还原为原始CPI水平值）

年份	CPI (上年=100)	95%下限	95%上限	Δ CPI
2025	98.16	93.72	102.59	-2.04
2026	96.51	90.70	102.32	-1.65

预测说明：

1. 基准CPI: 100.2 (2024年实际值)
2. 预测基于一阶差分模型，已通过累积还原为水平值
3. 区间宽度随预测步长增加而扩大
4. 假设经济指标延续近期趋势

===== 第6.5部分：自定义预测功能 =====

自定义预测功能已就绪！

语法: `[cpi_pred, ci, report] = custom_predict(input_data)`

input_data 格式说明：

- 5个元素的向量: [失业率, M2增速, 进出口增速, 投资增速, 财政收入增速]
- 或 5×N矩阵: N组不同的预测场景
- 参数单位均为百分比 (%), 不需要除以100

【演示】预定义场景预测:

场景 1: 基准场景 (趋势延续)

输入指标:

失业率 (%)	:	4.20%
货币供应量增比 (%)	:	7.30%
进出口总值增比 (%)	:	6.00%
固定资产投资额增比 (%)	:	4.00%
国家财政收入累计增长 (%)	:	5.20%

预测结果:

CPI预测值 (上年=100): 93.68
 95%预测区间: [80.03, 107.33]
 区间宽度: 27.30

场景 2: 乐观场景 (经济增长强劲)

输入指标:

失业率 (%)	:	3.90%
货币供应量增比 (%)	:	9.30%
进出口总值增比 (%)	:	9.00%
固定资产投资额增比 (%)	:	6.00%
国家财政收入累计增长 (%)	:	7.20%

预测结果:

CPI预测值 (上年=100): 94.14
 95%预测区间: [81.24, 107.04]
 区间宽度: 25.80

场景 3: 悲观场景 (经济下行压力)

输入指标:

失业率 (%)	:	4.70%
货币供应量增比 (%)	:	5.30%
进出口总值增比 (%)	:	3.00%
固定资产投资额增比 (%)	:	2.00%
国家财政收入累计增长 (%)	:	3.20%

预测结果:

CPI预测值 (上年=100): 93.00
 95%预测区间: [77.94, 108.05]
 区间宽度: 30.11

场景 4: 高通胀场景 (货币超发)

输入指标:

失业率 (%)	:	4.00%
---------	---	-------

货币供应量增比 (%)	:	12.30%
进出口总值增比 (%)	:	7.00%
固定资产投资额增比 (%)	:	7.00%
国家财政收入累计增长 (%)	:	8.20%

预测结果:

CPI预测值 (上年=100): 93.68
 95%预测区间: [80.35, 107.01]
 区间宽度: 26.66

场景 5: 通缩场景 (需求不足)

输入指标:

失业率 (%)	:	5.20%
货币供应量增比 (%)	:	4.30%
进出口总值增比 (%)	:	1.00%
固定资产投资额增比 (%)	:	-1.00%
国家财政收入累计增长 (%)	:	1.20%

预测结果:

CPI预测值 (上年=100): 92.25
 95%预测区间: [75.68, 108.81]
 区间宽度: 33.13

用户自定义预测接口

您可以在命令窗口中直接调用自定义预测函数:

```
>> input_data = [4.2, 7.3, 6.0, 4.0, 5.2]; % 输入您的5个指标
>> [cpi_pred, ci, report] = custom_predict(input_data);
```

批量预测多个场景:

```
>> scenarios = [4.0, 8.0, 5.0, 5.0, 6.0;
                4.5, 6.0, 3.0, 3.0, 4.0]; % 2个场景
>> [cpi_preds, cis] = custom_predict(scenarios);
```

查看预测详细报告:

```
>> report % 显示完整的预测分析报告
```

提示: custom_predict函数句柄已存在于工作区中

如需在其他脚本中使用, 可将函数定义保存为独立.m文件

===== 第7部分: 综合可视化仪表板 =====

综合可视化仪表板已创建 (6子图)。

===== 第8部分: 模型诊断与报告 =====

【残差诊断】

Ljung-Box自相关检验 (滞后2阶): p 值=0.5144 - 残差无显著自相关
 Jarque-Bera正态性检验: p 值=0.1919 - 残差服从正态分布
 异方差检验 (Breusch-Pagan简化版): p 值=0.0414 - 可能存在异方差

【变量重要性分析】

变量重要性排序 (基于后验均值绝对值):

- | | |
|-------------------------|--------------------|
| 1. 固定资产投资额增比 (%) (t-1) | $ \beta = 0.5507$ |
| 2. CPI (t-1) | $ \beta = 0.3075$ |
| 3. 货币供应量增比 (%) (t-1) | $ \beta = 0.2943$ |
| 4. 进出口总值增比 (%) (t-2) | $ \beta = 0.2860$ |
| 5. 国家财政收入累计增长 (%) (t-1) | $ \beta = 0.2806$ |
| 6. 货币供应量增比 (%) (t-2) | $ \beta = 0.2543$ |
| 7. 失业率 (%) (t-1) | $ \beta = 0.2311$ |
| 8. 进出口总值增比 (%) | $ \beta = 0.2135$ |
| 9. 进出口总值增比 (%) (t-1) | $ \beta = 0.1796$ |
| 10. 固定资产投资额增比 (%) (t-2) | $ \beta = 0.1467$ |

【模型局限性】

1. 样本量限制: 仅35年年度数据, 影响估计精度和统计推断
2. 结构变化: 经济体制改革可能导致参数不稳定
3. 遗漏变量: 未包含政策冲击、国际环境等定性因素
4. 线性假设: BVAR假设线性关系, 可能忽略非线性动态
5. 外推风险: 长期预测的不确定性随预测期增加而增大

=====
 分析完成!
 =====

>>