

本科生课程论文成绩评定表

教师填写	2021 —— 2022 学年 第 2 学期 课程名称 现代回归分析	课程类别： 必修[] 选修[✓]
	要求提交论文的日期 2022 年 6 月 25 日	考试方式： 课程论文
学生填写	经济 学院 经济统计学 专业 2019 级	
	姓名 苏晓钰 学号 2019052121 内招[✓] 外招[] 姓名 李敏怡 学号 2019052020 内招[✓] 外招[] 论文题目 “民以食为天” ——基于 ARIMA 及随机森林变量筛选的 BP 神经网络的小麦价格预测模型	
任课教师意见	论文评语：	
	评定分数： 签章： 2022 年 月 日	
参考评价指标： 1. 选题是否结合实际，是否有创新性； 2. 立论是否鲜明，观点是否正确，是否反映培养的目标要求； 3. 文献收集的广泛程度，是否了解本课题的研究状况及最新的进展情况； 4. 是否较系统地掌握了本课程的基础知识和专业知识，是否理论联系实际； 5. 是否结构合理、层次分明，文字和图表是否规范，学风是否严谨。		

“民以食为天”——基于 ARIMA 及随机森林变量

筛选的 BP 神经网络的小麦价格预测模型

[摘要] 粮食对于我国而言是经济发展、社会稳定和国家安全的基础。粮食价格既是调节粮食产品供求关系的信号和手段，也关系到老百姓生活成本和农民收入水平，因此预测小麦的价格走势具有非常重要的现实意义。

本文使用 2009 年 1 月至 2021 年 9 月《中国农产品价格调查年鉴》中小麦月度现货价格时间序列数据作为预测变量，从国家经济政策、国内外经济环境、现货基本条件三个维度出发，建立了有关小麦价格预测的指标体系，并利用相关系数与随机森林重要性排序进行变量筛选，筛选出人民币实际有效汇率指数、道琼斯指数等六个指标作为输入变量进行以下步骤建模：（1）按照 ARIMA 模型的建模步骤建立经典的单变量时间序列模型，最终确定为一阶差分、滞后阶数为 1 期、2 期及 12 期的模型，即 $ARIMA((1,2,12),1,0)$ 。（2）为尽可能提高精度，利用更多市场信息构建模型，建立单变量 BP 神经网络模型和根据指标体系建立多变量 BP 神经网络并与 ARIMA 模型进行比较。（3）利用残差优化的“误差补偿”思想，发挥单一模型组合优势，选取对线性部分具有较高拟合度的 ARIMA 模型和对非线性部分解决效果较好的 BP 模型，构建 ARIMA- BP 神经网络组合预测模型，并考察组合模型的精度提升作用。

以 2021 年 5 月至 9 月的小麦价格作为测试集，得到 $ARIMA((1,2,12),1,0)$ 、单变量 BP 神经网络、多变量 BP 神经网络、ARIMA- BP 组合预测模型四种模型的平均预测相对误差分别为 1.89%、1.51%、0.56%、1.25%。结果显示：（1）ARIMA 模型和单变量 BP 模型的预测值相较真实值偏小；（2）单变量 BP 神经网络与 ARIMA 模型的总体预测效果相近，但多变量 BP 神经网络模型的预测效果得到显著提高；（3）ARIMA- BP 组合预测模型相较于单一模型的预测精度有所提高，但仍然比不上基于指标体系构建的多变量 BP 神经网络模型。

[关键词] 小麦现货价格预测；随机森林；BP 神经网络；组合模型

“Bread is the Staff of Life”——Research on Wheat Spot Price Prediction Model Based on ARIMA and BP Neural Network with Random Forest Variables Screening

Abstract : Food is the foundation of economic development, social stability and national security for our country. Grain price is not only a signal and means to adjust the supply and demand relationship of grain products, but also related to the living cost of ordinary people and the income level of farmers. Therefore, it is of great practical significance to predict the price trend of wheat.

This paper first selected variables from the three dimensions of national economic policy, domestic and foreign economic environment, and spot basic conditions, using the Pearson Correlation Coefficient and Random Forest importance ranking to screen variables and establishes an indicator system for wheat price prediction, Then we used the monthly time-series data of wheat spot price from January in 2009 to September in 2021 in the "China Agricultural Product Price Survey Yearbook" as the predictor variable, and six indicators such as the RMB Real Effective exchange rate index and the Dow Jones Index as input variables for the following model:(1) We established a classic univariate time series model according to the modeling steps of the ARIMA model, and finally determined the model with first-order difference and lag order of 1 period, 2 period and 12 period, namely $ARIMA((1,2,12),1,0)$. (2) In order to improve the accuracy as much as possible, we used more market information to build a new model, which is establishing a univariate BP Neural Network and build a multivariate BP Neural Network according to the index system mentioned above. We also compared its accuracy with the ARIMA model. (3) Considering the complex linear and nonlinear structure of wheat spot price, we considered building an ARIMA-BP Neural Network combination prediction model, and examine the effect of accuracy improvement of the combination model.

Taking the wheat price from May to September 2021 as the test set, we built four models including $ARIMA((1,2,12),1,0)$, univariate BP neural network, multivariate BP neural network and ARIMA-BP combined prediction model. The average prediction relative errors of these models are 1.89%, 1.51%, 0.56%, and 1.25%, respectively. The results show that: (1)The predicted values of the ARIMA model and the univariate BP model are smaller than

the real values; (2) the overall prediction effect of the univariate BP Neural Network and the ARIMA model is similar, but the prediction accuracy of the multivariate BP Neural Network has improvement greatly. (3) Compared with the single model, the prediction accuracy of the ARIMA-BP Neural Network combination model has been improved. However, it is still inferior to the multivariate BP Neural Network model constructed based on the index system.

Keywords: Wheat Spot Price Forecast; Random Forest; BP Neural Network; Combinatorial Model

目 录

1 绪论	1
1.1 研究背景	1
1.1.1 国家战略	1
1.1.2 国内市场	1
1.1.3 国际市场	2
1.1.4 外部影响	3
1.2 研究理论价值与实践意义	4
1.2.1 理论价值	4
1.2.2 实践意义	4
1.3 国内外研究现状	6
1.3.1 文献综述	6
1.3.2 文献述评	6
1.4 研究内容和研究目标	7
1.4.1 研究内容	7
1.4.2 研究目标	8
2 小麦现货价格影响因素分析	10
2.1 数据选取与来源	10
2.1.1 小麦现货价格	10
2.1.2 小麦价格指标体系的构建	11
2.2 影响因素分析	11
2.2.1 相关系数分析	11
2.2.2 随机森林特征重要性排序	12
2.3 小结	13
3 ARIMA 时间序列模型小麦价格预测	14
3.1 平稳性检验	14
3.1.1 时间序列平稳性判别	14
3.1.2 时间序列的平稳化操作	15
3.2 模型的优化选择与参数估计	16
3.3 ARIMA 模型的显著性（白噪声）检验	17
3.4 ARIMA 模型的预测效果	18

4 BP 神经网络模型小麦价格预测	20
4.1 BP 神经网络理论概述	20
4.1.1 BP 神经网络模型的构建	20
4.1.2 模型训练	21
4.2 单变量 BP 神经网络模型建模预测	21
4.3 多变量 BP 神经网络模型建模预测	23
5 ARIMA-BP 组合预测模型	27
6 结论	29
6.1 模型结果比较	29
6.2 总结	29
7 研究评价	31
7.1 特色与创新	31
7.2 不足与展望	31
参考文献	32
附录	35
附录 1: R 语言代码	35
附录 2: 原始数据	39

1 绪论

1.1 研究背景

1.1.1 国家战略

粮食是人类发展根基，粮食安全关乎国际民生，无论是粮食生产、储备还是粮食贸易历来备受各国重视，是国家发展的重中之重。中国，作为发展中国家中的人口大国，耕地资源低于世界 10%，是粮食生产大国，更是消费大国^[1]。粮食问题是关系全局的重大战略问题，解决粮食问题、发展社会经济，粮食价格是最主要的关注点之一。党的十九大报告中强调：要确保国家粮食安全，把中国人的饭碗牢牢端在自己手中。中央一号文件连续十二年关注“三农问题”，并提出“发展农产品期货交易”。2016 年，明确提出了探索建立农业补贴、涉农信贷、农产品期货与农业保险联动相关机制，发挥粮食期货稳定产销关系、控制价格变动风险的作用^[2]。

小麦，作为三大主粮之一，全球年均约 75%用于口粮消费，其余作饲料、工业等用途，而中国的小麦年产量和消费量都居世界前列。2006 年起，我国对小麦实行最低收购价政策，稳定主粮价格、改善供求关系以保障农民利益，从而提高农民生产积极性。2017 年中央一号文件提出“要坚持并完善稻谷、小麦最低收购价政策，合理调整最低收购水平，形成合理比价关系”^[3]。在近十年的国家政策扶持下，国产小麦产量提升、粮食库存充足，有效保障市场需求，小麦现货价格稳步上涨。

1.1.2 国内市场

中国小麦年总产量超 1 亿吨，年播种面积和产量均约占世界粮食常年种植面积和产量的四分之一，自给率达到 95%以上，在国民经济中占比巨大^[5]。2020 年小麦种植面积 23380 千公顷，2021 年小麦总产量达 13694.6 万吨，同比增加 269.6 万吨，增长率 2%。同时，随着人口增长和经济发展，小麦需求量逐年上涨，市场对于优质高等小麦和饲料级小麦的需求旺盛。

纵观 2007 年世界粮食危机以来，得益于临储和“托市”政策，我国小麦价格从 2009 年 1 月的 1790 元/吨上涨至 2013 年 9 月的 2519.29 元/吨，上涨幅度约 40.74%。2013 年以后，小麦价格波动明显且呈现出跌涨互现趋势^[6]。2020 至今，受外界多重因素影响，粮食稳定成为国家安全战略部署重要手段，国内小麦减

产、储备性需求激增导致价格明显上涨，到 2021 年 12 月 15 日小麦月度价格超过往年同期，达到了 2856.67 元/吨；今年第一季度，小麦每吨价格突破 3000 元大关。

而国内中筋小麦品种占比超 55%，是国产小麦重要来源。由于优质高筋、低小麦被排除在最低价补贴政策之外，普麦与优质麦价格差距小，农民普遍选择种植性价比更高的普麦，进口成为补充我国优质小麦缺口的重要途径^[6]。此前，由于国内小麦生产成本明显高于国际水平，小麦价格一直呈现国内高于国外到岸价的“倒挂”现象，2020 年美国墨西哥湾硬红冬麦配额内到岸税后价约每吨 1965 元，低于国内 220 元^[7]。

1.1.3 国际市场

小麦是国际农产品贸易中最为活跃、用途最广的农产品之一，其种植分布、消费及总贸易额占世界粮食作物第一，作为大宗粮食作物占世界贸易近 46%^[6]。国际上呈现出口国集中，进口国分散的局面，总体来看各地区进口来源国较为固定，其中美国是小麦第一大出口国。

随着人口增长以及人们对于优质小麦日益增长的消费需求，自 2001 年起，我国逐步从小麦净出口国转变为净进口国，主要进口优质小麦和饲料用小麦以调剂粮食结构。2021 年小麦进口量 977 万吨创新高，虽仅占我国总量约 3%，但仍为世界小麦进口大国。长期以来，美国、加拿大和澳大利亚进口小麦占我国进口量八成以上，依存度高，存在一定的粮食安全隐患；2018 年起，中美贸易摩擦不断，为中国小麦进口带来更多不确定性。而 2007 年和 2011 年两次粮食危机中，乌克兰和俄罗斯两个小麦出口大国采取了严厉的出口限制措施，进一步推高国际粮食价格，造成国际市场动荡^[8]。

随着一带一路倡议提出，沿线国家经济领域合作加深，由于俄罗斯和乌克兰小麦产量占全球 14%，出口总额近 30%，中国和俄罗斯、乌克兰、哈萨克斯坦的小麦贸易关系逐步建立，寻求多元的进口途径，降低对少部分国家的粮食依存度^[9]。本月，在国际小麦价格暴涨的形势下，以 3 月 3 日为例，国内小麦现货价达 3210.67 元/吨，进口美国小麦完税成本突破 3600 元/吨，价差倒挂每吨 400 元左右，进口美国小麦已无成本优势。期间，国内发布公告“允许俄罗斯全境小麦进口”，改善粮食储备结构。

1.1.4 外部影响

(1) 2020 新冠肺炎疫情袭来，WTO 将其定义为全球大流行病，当前仍在全球范围蔓延。最新《全球粮食危机报告》指出，新冠肺炎疫情大流行或导致全球遭受严重饥饿人口数量翻 1 倍，粮食安全成为关注焦点^[10]。疫情爆发初期，国内受疫情影响，交通封锁、停工停产等使粮食运输受限导致原料短缺、面粉生产加工停摆，加之人们在封锁隔离期恐慌性储备主粮，国内小麦价格短期内一度上涨。随着疫情转移到国外，2020 年 7 月以后，国际谷物价格进入了上升通道。同年 10 月 20 日，芝加哥期货交易所（CBOT）小麦主力合约报收比月初上涨了 11.2%^[7]。受疫情影响，各国封锁下进出口贸易受阻，外部供给环境恶化；谷物出口大国担忧粮食供应，相继限制粮食出口；同时发达国家采取量化宽松政策刺激经济推高大宗商品价格；在经济衰退下，投资者纷纷将目光转向粮食市场，多重作用下，国际小麦价格上涨。由于国内小麦现价与国际价格紧密关联，国际疫情引起的市场波动往往对我国小麦生产销售市场产生影响。

(2) 粮食作物产量和质量直接影响现期货价格。同一时期，非洲蝗灾、美欧小麦主产区遇寒潮，影响小麦优良率和作物生长状况，降低新麦上市预期^[7]。国内，华北地区秋收秋种期遭遇持续降雨，部分地区小麦受损质量下降、播种推迟增加成本，引发市场对小麦减产担忧，支撑小麦价格高位运行^[11]。

(3) 乌克兰作为“欧洲粮仓”，与俄罗斯小麦产量皆占世界前五。随着俄乌冲突升级，欧美对俄罗斯制裁，加剧疫情后复苏带来的通胀压力。一方面，国际市场尤其是依赖两国出口小麦的国家对于未来粮食供应状况充满忧虑，且黑海港口贸易陷入混乱，增大粮食运输阻力；另一方面，小麦期货作为金融产品，除了作为现货价格风向标，在特殊时期全球资本涌入市场避险、投机。半月内，伴随国际油价上涨，推动小麦价格更上一层楼，CBOT 小麦期货价格涨幅 40%，为 14 年最高^[12]。2 月 21~25 日，中国小麦进口成本水涨船高，普遍上涨 10% 以上^[13]。当下，中国开放俄罗斯小麦进口，节省海运成本，以较低价格保障国家小麦库存量，同时也借此机会加快进口小麦结构调整，降低对特定国家依存度。

总体而言，商品的供求及价格决定，即市场中的均衡分析是经济分析中最基本的问题之一，小麦作为中国乃至世界主粮，其现货价格变动以及市场动态备受各方关注。

1.2 研究理论价值与实践意义

1.2.1 理论价值

本课题构建 ARIMA 模型、单变量的 BP 神经网络模型、多变量的 BP 神经网络模型、ARIMA-BP 组合预测模型，克服了单变量预测模型的片面和预测准确率偏低的情况，同时对单一模型的预测结果进行了优化，具有一定的创新意义，丰富了农产品价格预测模型。

在数据的实证分析方面，本文采用了 2009 年 1 月至 2021 年 9 月的月度数据，其中包含了 18 年中美贸易战、19 年年末至今的全球新冠肺炎流行病等有可能使小麦价格产生大幅波动的关键时间节点，弥补了在国际冲突不断的后疫情时代下关于小麦期货价格预测实证研究的空白。

1.2.2 实践意义

（1）调节市场供求关系

价格受供求关系影响，又反作用于供求关系。过去 20 年，我国小麦产量稳步提升，基本满足国民消费需求，粮食库存充足，同时受政策影响价格涨幅较大。由于国内小麦生产从成本高，国家通过小麦最低收购价补贴、临储麦政策“托市”，维持了小麦供应量，从而保障消费需求。

在此期间，我国粮食期货市场日臻成熟，小麦期货价格成为市场价格“风向标”。作为大宗商品，小麦期货为农民和供应链下游的加工商、贸易商提供套期保值的功能。通过期货市场和现货市场的信息传递，有效反应市场当前和预期供求关系。合理的价格预测，一方面，指导农民调整小麦种植结构、合理安排售粮时间，保障粮食供给、控制供求不平衡带来的损失；另一方面，对于供应链下游，有助于生产商、销售商根据预期价格趋势进行采购、生产和存货管理^[2]。通过有效价格预测，减少卖方与市场的信息不对称，提高整条小麦供应链效率、保障生产者利益，满足市场需求，构建健康的主粮市场。

（2）规范国内小麦定价

2006 年实行小麦最低收购价以来，国产小麦价格增长近一倍。由于长期以来粗放型的生产方式导致我国小麦种植成本高于国际水平，为保障国家粮食安全，国家政策补贴在一定程度上扭曲小麦价格。2016 年对于粮价补贴政策的改革在一定程度上提升了市场对于小麦价格的调节能力，此后我国小麦价格一直

在小范围波动上涨，更加有效地反应市场供求关系。

此外，作为粮食进口大国，我国小麦贸易逆差大、买方市场势力相对较弱，对于粮食定价影响力低^[14]。相比于出口大国对于小麦种植的扶持力度，我国还有一定距离。尽管国内外价差有缩小趋势，但今年三月以前，进口小麦价格一直优于国内，国产小麦价格及质量竞争力较低。对小麦价格的有效预测，是国家加快推进小麦和稻谷定价机制、进行补贴政策和收储体制改革的重要依据；与此同时，帮助人们有效应对国际粮价波动带来的风险，促进国产小麦价格与国际化、标准化接轨，提升我国小麦产业国际影响力。

（3）保障国家粮食安全

后疫情时代，在小麦价格波动剧烈、市场风险加剧的整体经济大环境下，准确预测小麦现货价格有利于降低外部环境对于粮食供应、粮食价格的冲击，优化粮食储备。尤其是当下疫情和政治等不确定性因素加剧，小麦现货市场动态是外界环境的“晴雨表”。

作为金融期货市场的物质基础，小麦期货市场围绕现货价格波动。近期小麦期货价格的暴涨，在一定程度上反映了主粮，这类大宗商品，其避险属性突出；此外，众多投机者借助国际事件涌入小麦期货市场以期进行无风险套利，同时借机推高国际利率，收割高额回报^[15]。这一特殊时期，小麦价格的合理预测有助于稳定现货价格，缓解全球性恐慌造成的价格虚高，平衡期现货市场从而避免投机行为的出现。小麦作为国家主粮，其作用应该是满足人们口粮需求、确保社会生产有序，而非投机者哄抬物价获利的工具。因此，掌握未来价格趋势的对于小麦发挥粮食作用、国家粮食供应稳定具有重要意义。

1.3 国内外研究现状

1.3.1 文献综述

小麦价格受到供给、需求、政策导向、替代品价格等因素的影响,导致价格波动频繁。下列为国内外学者关于小麦价格的预测方法的探究。

国外学者的研究主要集中在小麦价格的预测方法上:Haile(2016)^[16]等通过采用新开发的多国作物日历特定季节分解模型估算了小麦全球总供应量,并相应地预测出其价格变化和价格波动。研究表明,产出价格波动与小麦供应呈负相关,价格风险降低了小麦的生产反应,从而抑制了价格激励效应。Pal(2018)^[17]等通过对原油价格和粮食价格进行交叉相关分析,证明了原油价格与世界粮食价格之间存在强烈的正向依存关系;Ahumada(2016)^[18]等利用向量自回归方法(VAR)建立小麦、玉米和大豆三者价格交叉影响的价格预测模型,结果表明三者互为替代品,价格存在此消彼长的关系。

国内关于小麦价格预测主要集中在影响因素上:如柳燕子(2018)^[19]基于2009至2017年全国小麦收购价格指数,针对价格序列季节性、周期性、非线性及多尺度性等特征,提出基于EEMD分解、ARIMA预测的多模型小麦价格预测方法,为小麦价格预测提供了新思路。李雪(2018)^[20]等将最低收购价政策作为虚拟变量引入GARCH模型,实证研究政策对平抑小麦期现货价格波动的作用,结果表明,政策对小麦期现货价格波动产生了明显的影响,能够显著降低小麦现货市场的波动程度,但政策对期货价格波动的作用方向则相反。张瑞娟和高芸(2016)^[21]通过VAR模型和格兰杰因果检验分析了小麦最低收购价格和国际小麦月度价格对我国小麦价格的影响,提出国际小麦价格与我国小麦价格变化没有因果关系,而小麦最低收购价是拉高国内小麦价格的主要因素。崔海莹(2022)^[22]等通过构建DCC-MGARCH模型对多个时间序列之间的动态相关性进行估计,分析新冠肺炎疫情冲击对于三大主粮价格的影响效应和作用机制,认为的国内粮价会受到国际传导作用,并通过“国际疫情恶化-国际粮食价格波动-国内粮食价格波动”的传导机制产生影响。

1.3.2 文献述评

总体而言,国内外学者关于小麦价格的研究主要从探寻价格影响因素、研究价格波动走势、开展小麦价格与其他重要物资价格影响机制等角度的相关研究。主要运用的模型有向量自回归方法(VAR)、定季节分解模型、BP神经网络

(BPNN)算法、ARIMA 预测等进行定量研究，尤以 VAR 模型使用最广，但前三种方法所需要的指标较多，且对于我国目前的统计发展现状而言，大部分指标仍难以大范围获取。

同时，现有文献主要采用定性的方法，研究新冠肺炎疫情对中国粮食价格的影响、习惯将小麦价格与其它粮食关联，但对中国粮食价格的影响效应及作用机制没有得到充分关注，从而忽略了小麦自身价格属性和变动情况。同时，在后疫情时代，粮食价格波动情况不同于以往，对于这一时期小麦现货价格的研究还处于未成熟阶段。笔者认为对于粮食价格预测的模型可能受外部因素影响而使准确率发生变化，采用疫情前数据进行研究的模型需适时改进。

表 1 不同预测模型间的优缺点比较

模型	优点	缺点
ARMA	模型简单方便	只可对平稳序列进行预测
AMIMA	模型简单，只需借助内生变量	只能捕捉线性关系
SVR	可解决非线性、高维问题	数据量大时训练速度较慢
BP 神经网络	准确度高，能充分逼近复杂的非线性关系	需要大量的参数，学习时间过长
灰色预测 (GM)	所需要的样本数据量小，且对复杂系统的预测效果较好	只适合指数增长和中短期的预测

1.4 研究内容和研究目标

1.4.1 研究内容

本项目基于《中国农产品价格调查年鉴》中的 2009 年 1 月 4 日至 2021 年 9 月 17 日共 3205 个样本我国小麦期货价格序列的日数据、以及国家统计局、中国人民银行、国际清算银行、国际货币基金组织关于国家经济政策、国内外经济环境、现货基本条件等多个影响小麦现货价格的因素，建立数学模型预测小麦期货价格。

第一部分：构成小麦期货价格的影响因素指标体系。

计算 Pearson 相关系数度量变量之间先行相关关系，变量之间的非线性相关程度通过随机森林特征重要性排序，最终筛选出少数具有影响力的因素，构

成预测小麦期货价格的指标体系。

第二部分：基于所构建的小麦期货价格的预测指标体系，建立模型进行预测。

- (1) 通过传统的 ARIMA 模型建立经典的单变量时间序列模型；
- (2) 建立预测市场现货价格的单变量 BP 神经网络模型；为了提升模型预测效果，根据由随机森林提取的特征构建影响因素指标体系，构建多变量 BP 神经网络模型；
- (3) 将小麦价格序列分解为线性趋势和非线性趋势，构建单变量的 ARIMA-BP 组合预测模型。

第三部分：评价单一预测模型与组合模型的预测效果，并进行最优选择。

通过计算得到测试集的误差、相对误差以及平均相对误差，并比较单一模型（ARIMA、单变量 BP 神经网络模型），组合模型（ARIMA-BP 组合预测模型），多变量模型（多变量 BP 神经网络模型）等预测模型下的指标，最终选出最适合的预测模型，并应用于我国小麦市场发展趋势预测。

1.4.2 研究目标

- (1) 了解国内小麦市场的粮食农业扶持政策、市场发展现状和了解小麦市场风险，深入挖掘国内小麦价格的重要影响因素。
- (2) 基于近十二年各个维度的数据，构建多个小麦期货价格预测模型。
- (3) 比较 ARIMA 模型、单变量 BP 神经网络模型、多变量 BP 神经网络模型和 ARIMA-BP 组合预测模型的预测效果，评价并选择出效果最好的预测模型。
- (4) 将本课题的预测模型应用到农业经济领域中，并利用该模型对未来农业布局、农产品供需关系的探讨和相关政策提出有价值的建议。

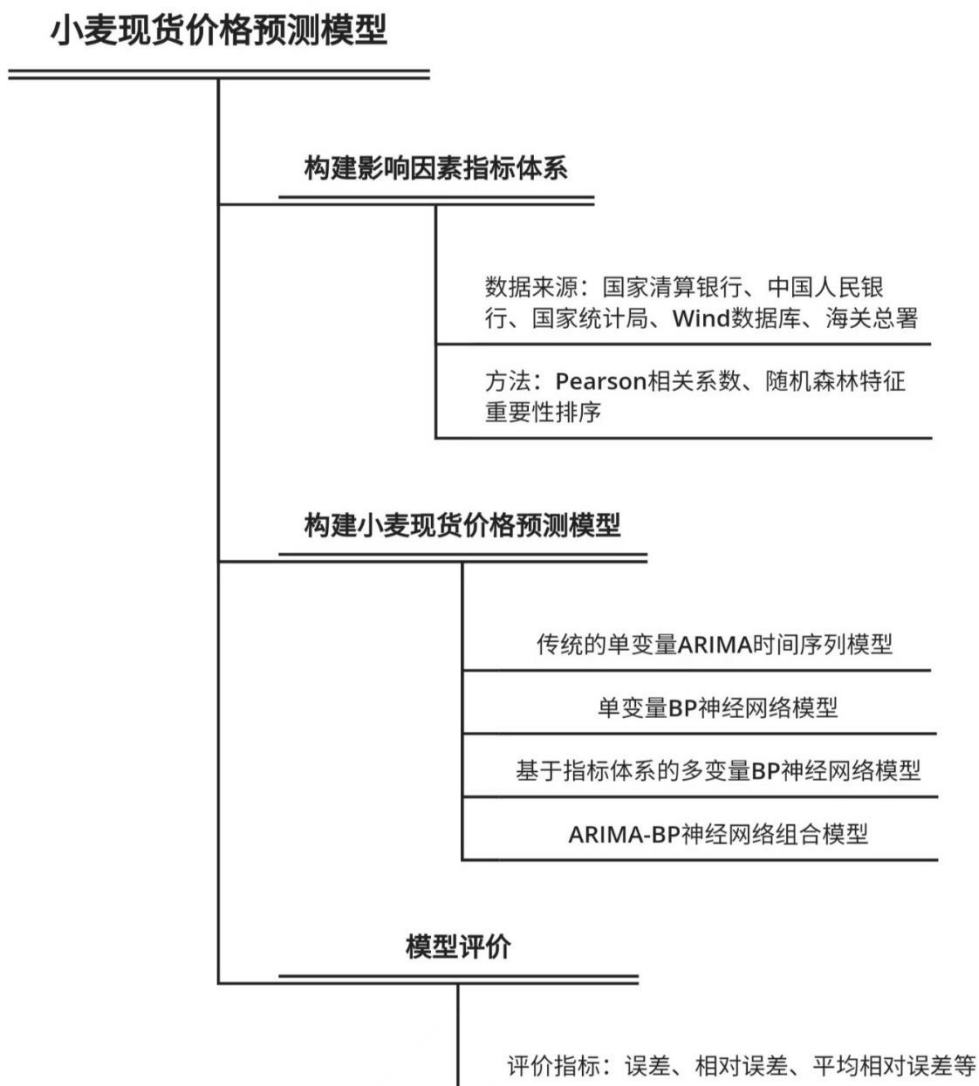


图 1 技术路线图

2 小麦现货价格影响因素分析

2.1 数据选取与来源

2.1.1 小麦现货价格

我国小麦现货价格序列是来自《中国农产品价格调查年鉴》当中的 2009 年 1 月 4 日至 2021 年 9 月 17 日共 3205 个样本日数据，其中有 15 个样本存在缺失值。由于考虑到小麦现货价格在各个月的价格几乎相同，为了删除冗余信息，本项目取从 2009 年 1 月至 2021 年 9 月的每个月月中（15 号）作为代表一个月平均价格水平的样本，其中某些月份缺失月中（15 日）的数据，将取离 15 日最近的一天数据作为一个平均价格水平的代表，最终获得共 156 个最终月样本。

小麦现货价格序列是通过每日记录所得的，在记录过程中难免会出现人为记录错误、记录仪器出现损坏等突发情况出现，导致价格序列部分数据缺失。而缺失的数据将会在后期与预测过程中对模型的构建和准确率造成影响，因此需要对缺失数据进行处理。本文采取传统插值法处理数据缺失值。观察序列发现，2012 年 7 月 16 日的数据存在缺失，由于 7 月 15 日数据同样存在缺失，因此不适合用插值法，因此对此进行丢弃，取 7 月 13 日数据作为月代表值。下列为小麦现货价格最终数据。

表 2 2009 年至 2021 年小麦现货价格每月数据

年份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
2009	1790.00	1817.00	1855.00	1862.00	1878.00	1836.00	1853.00	1866.00	1920.00	1960.00	1977.00	2021.00
2010	2006.00	2005.00	2014.00	2006.00	2011.00	2005.00	2016.00	2028.00	2050.00	2058.00	2113.00	2120.00
2011	2117.00	2116.00	2171.00	2153.00	2155.00	2157.00	2050.00	2044.00	2068.00	2133.00	2150.00	2139.00
2012	2123.00	2140.00	2145.71	2173.57	2189.29	2145.71	2102.86	2156.43	2255.71	2268.57	2319.29	2364.29
2013	2470.00	2472.14	2470.71	2466.43	2477.14	2396.43	2413.57	2448.57	2519.29	2568.57	2574.29	2561.43
2014	2552.14	2550.71	2555.00	2527.14	2515.00	2441.43	2487.14	2510.71	2573.57	2570.00	2552.86	2558.57
2015	2563.57	2555.71	2557.14	2555.71	2527.86	2366.43	2390.71	2367.86	2355.71	2282.14	2267.14	2317.86
2016	2323.57	2331.43	2341.43	2363.57	2396.43	2183.57	2254.29	2285.71	2330.71	2406.43	2487.14	2547.14
2017	2565.71	2585.56	2610.56	2622.22	2616.67	2354.44	2400.00	2418.89	2492.22	2492.22	2530.56	2547.78
2018	2545.56	2542.22	2543.33	2490.00	2461.67	2378.89	2408.33	2414.44	2440.56	2445.56	2477.22	2488.89
2019	2470.00	2460.00	2447.22	2432.22	2429.44	2291.11	2271.11	2289.44	2295.56	2324.44	2407.22	2392.22
2020	2390.00	2410.56	2415.00	2445.56	2436.67	2296.67	2350.00	2429.44	2435.56	2455.56	2486.67	2480.00
2021	2580.00	2581.11	2589.44	2555.56	2540.00	2522.22	2561.67	2595.56	2612.78			

2.1.2 小麦价格指标体系的构建

本研究主要参考战玉锋、刘放^[1]和张恒等^[2]的研究成果，从国家经济政策、国内外经济环境、现货基本条件三个维度出发，初步筛选出以下影响小麦期货价格的指标。

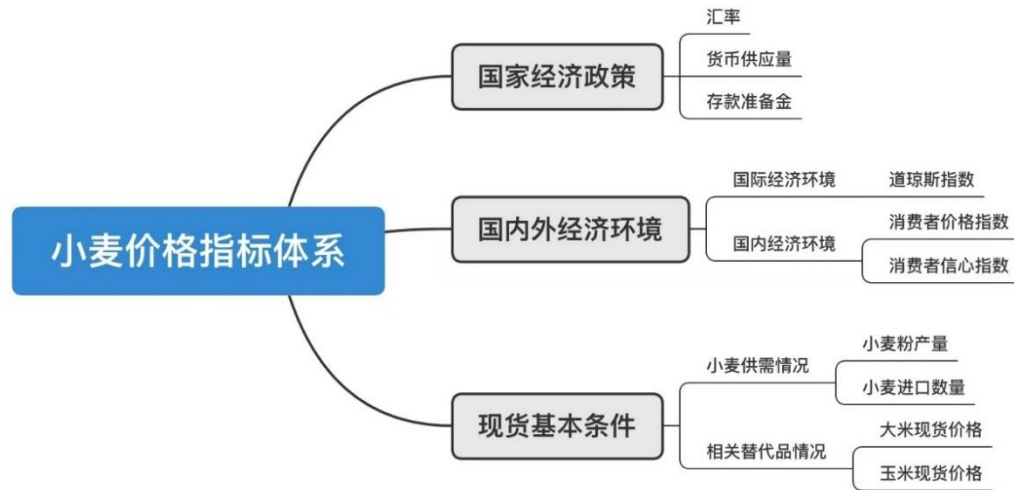


表 3 小麦价格指标体系及其数据来源

具体指标	单位	对应变量	数据来源
人民币实际有效汇率指数	2010 年=100	X1	国际清算银行
货币供应量 M2（同比）	%	X2	中国人民银行
人民币存款准备金率	%	X3	国家统计局
道琼斯指数		X4	Wind 数据库
消费者价格指数 CPI	%	X5	国家统计局
消费者信心指数		X6	国家统计局
小麦进口数量	吨	X7	海关总署
全球大米实际市场价格	美元/公吨	X8	国际货币基金组织
全球玉米实际市场价格	美元/公吨	X9	国际货币基金组织

2.2 影响因素分析

2.2.1 相关系数分析

加入影响因素进行预测时变量的选择是关键，将不同的影响因素引入模型进行预测可能会导致不同的预测精度^[25]。为筛选出合适的影响变量，首先计算出各影响变量 $x_1 \sim x_9$ 与小麦价格指标 y 的 Pearson 相关系数，明确变量间的相

关程度，绘制热力图如下：

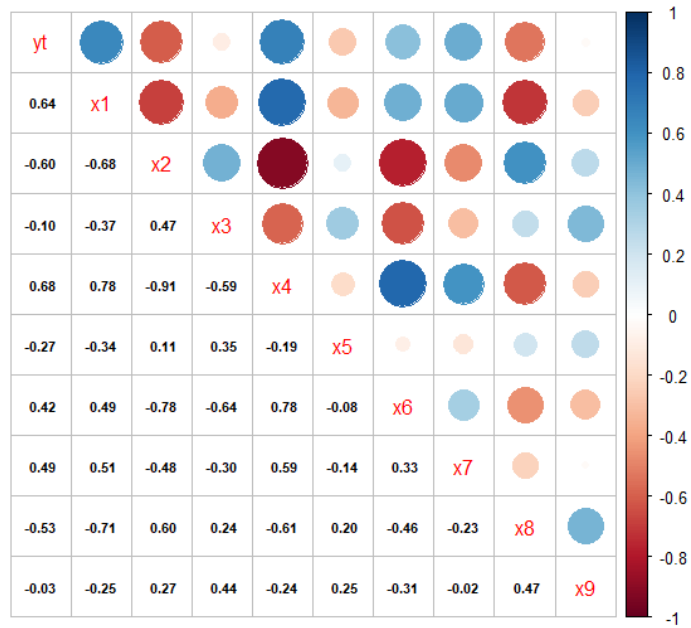


图 3 相关系数热力图

根据相关系数热力图，可以看出，除 x3，x9 以外，其他变量与小麦价格 yt 均呈现较高相关关系，且在相关系数检验中显著。

2.2.2 随机森林特征重要性排序

Pearson 相关系数衡量的是仅为变量间的线性相关程度，为度量变量间的非线性相关程度，以及考虑到变量间存在的高度相关性，本节采用随机森林算法通过将特征按照重要性排序对变量进行筛选。该算法对共线性问题不敏感，比较各因素的重要性大小最终筛选出对结果贡献较大的特征^[26]。

因此为得到对小麦价格贡献较大的变量，提高模型的预测能力及拟合速度，采用随机森林模型对各影响因素进行重要性排序，结果如下图所示：

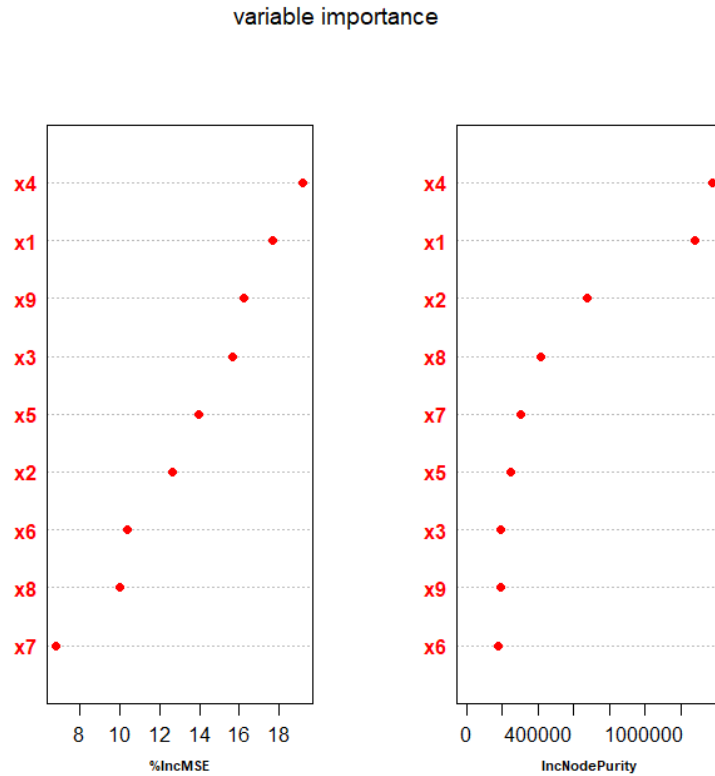


图 4 随机森林变量重要性排序图

“%IncMSE”即 increase in mean squared error 通过对每一个预测变量随机赋值，如果该预测变量更为重要，那么其值被随机替换后模型预测的误差会增大。因此，该值越大表示该变量的重要性越大。

“IncNodePurity”即 increase in node purity，通过残差平方和来度量，代表了每个变量对分类树每个节点上观测值的异质性的影响，从而比较变量的重要性。该值越大表示该变量的重要性越大。

2.3 小结

本章根据相关理论研究选取了 9 个与经济发展相关的指标，首先通过相关性分析，得到本文选取的指标除 x3、x9 外均与小麦价格相关程度较大的结论。然后通过随机森林特征根据每个变量对分类树每个节点上观测值的异质性的影响进行排序，发现重要性较低的变量为 x3、x9、x6。

结合两方面分析，本次研究最终选取除 x3、x6、x9 的其他 6 个对小麦价格影响程度较高的指标构成小麦价格预测指标体系，以便在后文中构建多变量模型预测小麦价格。

3 ARIMA 时间序列模型小麦价格预测

本次研究将 2009 年 1 月至 2021 年 4 月作为训练样本（共 148 个月样本），2021 年 5 月至 2021 年 9 月作为测试样本（共 5 个月样本）。

3.1 平稳性检验

3.1.1 时间序列平稳性判别

平稳序列有着均值、方差为常数，不受其他因素的影响而变化的特点。本文首先将通过绘制时间序列图观察序列的平稳性。

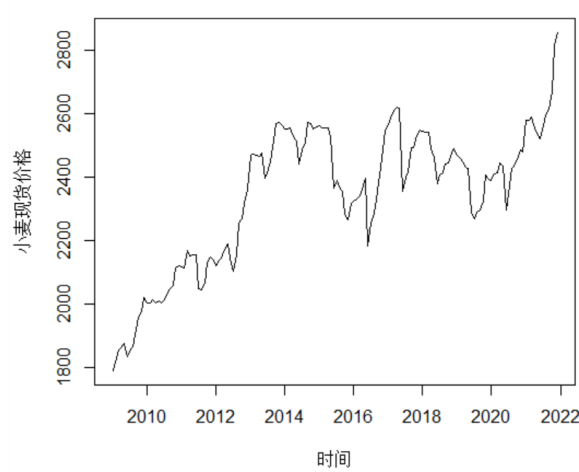


图 5 小麦现货价格原始数据时序图

从图 5 可以看出小麦现货价格呈现逐年递增的趋势，且具有季节性趋势，季节效应表现为 12 个月为一个周期，可以判断这个时间序列是非平稳的。由于季节效应的存在，选择模型时应考虑其效应的影响，首先排除非季节性的 ARMA(p, q)模型。进一步运用单位根检验法对小麦现货价格时序的平稳性进行验证。由表 4 可以看出，在显著性水平为 5% 下，ADF 检验的卡方统计量为 -2.3799，且 P-value 值为 0.4182（大于 0.05），因此判定原序列存在单位根，为非平稳序列，需要先进行差分处理。

表 4 小麦现货价格时间序列的单位根检验

	Dickey-Fuller	P-value
ADF test statistic	-2.3799	0.4182

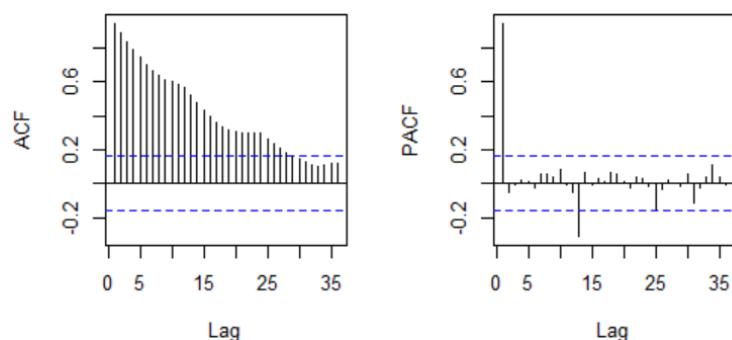


图 6 小麦现货价格原始数据自相关图与偏自相关图

表 5 ARMA 模型定阶原则

模型定阶	AR (p)	MR (q)	ARMA (p, q)
自相关系数 (ρ_k)	拖尾	P 阶截尾	拖尾
偏自相关系数 (ϕ_{kk})	Q 阶截尾	拖尾	拖尾

由小麦现货价格原始数据自相关图与偏自相关图可看出，ACF 图中显示是指数衰减趋势，即拖尾现象，并且衰减速度缓慢，一直没有衰减到接近 0 的数值，仍有大量值大于 0.5，说明序列 y_t 与之前的时刻有强烈的相关性，而且相关项的范围是比较大的，可进一步可判断为非平稳时间序列；而在 PACF 图呈现一阶截断，显示为截尾，因此运用 ARIMA(p, d, q)模型建模。

3.1.2 时间序列的平稳化操作

由于原时序数据为非平稳序列，为了消除线性趋势效应，我们取每一项与前一项的差：首先需要先对数据进行差分处理，将连续观察值之间做差，再对一阶差分后的数据做出时序图、ACF 与 PACF 图检验。

从上图可见一阶差分后的时序图在 0 附近波动，但仍然存在较为显著的波动起伏大的情况，因此仍然不够平稳；且 ACF 图的滞后 12 期、滞后 24 期、滞后 36 期处显示序列存在强自相关，因此可以推测存在一个 12 阶的周期性影响因素。为了消除趋势效应，我们首先通过再次进行差分做出时序图、ACF 与 PACF 图检验与单位根检验，判断差分后的序列是否平稳。

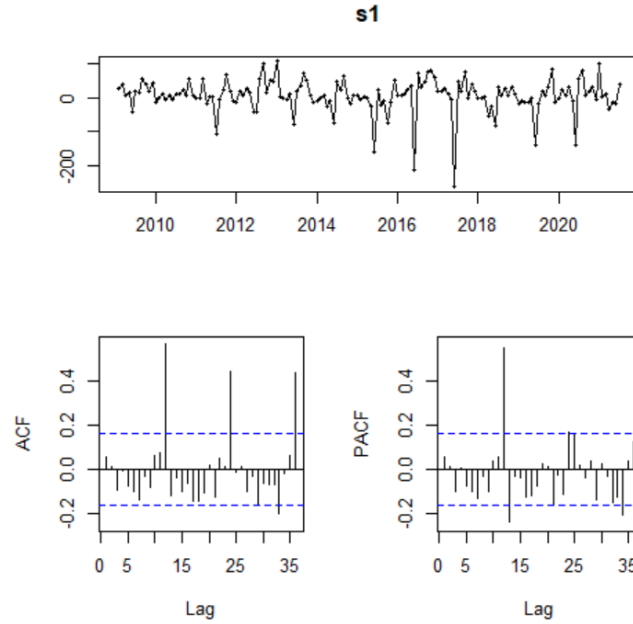


图 7 一阶差分 and 季节差分后的小麦现货价格时序图、ACF 图和 PACF 图

表 6 季节差分后的小麦现货价格单位根检验

	Dickey-Fuller	P-value
ADF test statistic	-5.6747	0.01

从上图可见一阶差分后的时序图在 0 附近波动，2015 年后的差分序列存在较为显著的波动起伏大的情况，由 ADF 检验可以看出，在 5% 的水平下，卡方统计量值为-5.6747，且 P 值大于 0.05，因此可以拒绝原假设，认为经过一阶差分的序列是平稳序列，符合 ARMA 时间序列模型的建立。由 ACF 图可以看出，在已经消除了单位根的情况下，滞后 12 期、滞后 24 期、滞后 36 期处仍然存在序列强自相关，因此考虑建立 ARIMA(12,1,0)。

3.2 模型的优化选择与参数估计

下列为假设的 ARIMA 模型，根据最小信息量准则可以得到最优的模型。

表 7 ARIMA 模型检验

	似然比统计量	AIC
ARIMA((1, 2, 12), 1, 0)	-754.35	1516.7
ARIMA((1, 2, 3, 12), 1, 0)	-754.18	1518.35
ARIMA((1, 2, 6, 12), 1, 0)	-754.28	1518.55
ARIMA((1, 2, 3, 12), 1, 0)	-750.95	1511.91

根据简练原则，选取最小 AIC 与似然比统计量对应的模型，可得

ARIMA((1,2,12),1,0)是最优的模型，以下是该模型的参数估计,可见系数在 1%的显著性水平下显著，因此模型参数是有意义的。

表 8 ARIMA((1,2,12),1,0)模型系数表

	ar1	ar2	ar12
Coeff	0.0224	-0.0111	0.5916
标准差	0.0650	0.0647	0.0685

由系数表可得最后估计出来的模型为：

$$y_t = 0.0224y_{t-1} - 0.0111y_{t-2} + 0.5914y_{t-12} + \varepsilon_t \quad \text{式 (1)}$$

3.3 ARIMA 模型的显著性（白噪声）检验

为检验模型的有效性，采取观察残差标准差图、Ljung-Box 法来检验模型的显著性。若上述各种检验方法不能拒绝原假设，则可以认为 ARIMA((1,2,12),1,0)模型的残差序列是白噪声序列，该模型有意义。

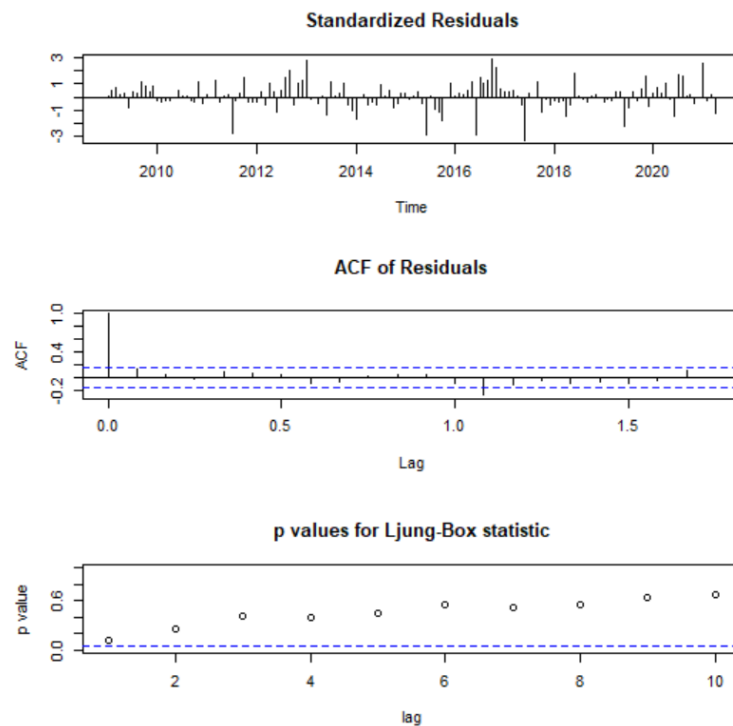


图 8 残差标准差、ACF 图、Ljung-Box 检验图

表 9 Ljung-Box 检验

	X-squared	p-value
model\$residuals	8.9772	0.7049

图 8 为标准化后所得到的残差图，可见残余误差都在零均值附近范围内波

动，具有均匀的方差，没有明显的异常值。残差标准差的自回归函数都为近似 0（两虚线内），Ljung-Box 检验图中随着滞后阶数的增大，p 值增加并总体在 0.5 上方，因此可以判断残差序列大致是白噪声序列。LB 检验的 p 值为 0.7049（大于 0.05），不能拒绝原假设 H_0 ，可以认为该序列为白噪声序列，拟合模型 ARIMA((1,2,12),1,0)为显著有效的模型。

3.4 ARIMA 模型的预测效果

利用训练好的模型对未来我国小麦的月度价格进行预测，得到 2021 年 5 月到 2021 年 9 月的月度价格预测值。

表 10 2021 年 5 月到 2021 年 9 月价格预测值

月份	真实值	预测值	误差	相对误差	平均相对误差
2021-05	2540.00	2549.449	9.45	0.37%	1.89%
2021-06	2522.22	2466.864	-55.36	-2.19%	
2021-07	2561.67	2496.633	-65.04	-2.54%	
2021-08	2595.56	2545.214	-50.35	-1.94%	
2021-09	2612.78	2549.592	-63.19	-2.42%	

运用均方根误差（RMSE）与平均绝对百分比误差（MAPE）两个模型预测的评价指标对小麦现货价格的预测结果进行评估。该 AIRIMA 模型的 RMSE 较小，平均绝对百分比误差 MAPE 仅为 1.0347%，结果显示该模型拟合效果较好。

表 11 ARIMA 模型检验

评价标准	结果
MAE	28.3136
RMSE	40.0983
MAPE	1.2230

根据以上 ARIMA 价格预测图以及预测结果可以看出，ARIMA 模型能较准确地预测小麦现货价格的时间序列整体趋势。

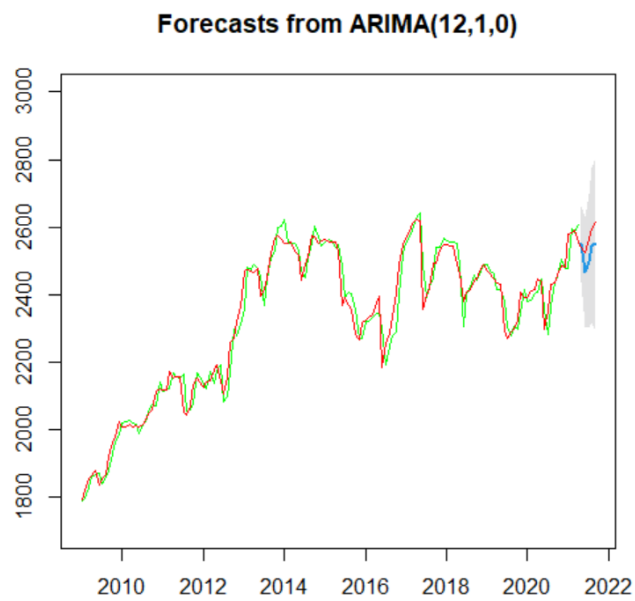


图 9 ARIMA 模型预测图

4 BP 神经网络模型小麦价格预测

根据小麦价格序列不平稳的问题，以下考虑使用神经网络模型。神经网络算法众多，其中 BP 神经网络因其网络结构简单直观的特点而被广泛应用。

4.1 BP 神经网络理论概述

BP 神经网络的概念由 Rumelhart 和 McClelland 在 20 世纪末提出，是利用误差逆传播方法在人工网络进行训练，被广泛应用于多因素、非线性和不确定的预测和评价问题。通过网络的变结构调节过程来学习训练样本，从而在不确定的环境中对复杂问题进行判断。

从结果上划分，BP 神经网络包括输入层，隐含层和输出层，隐含层的数量为 1 个或者多个，输入层和输出层数量皆为 1。层级之间的神经元相互连接，构成了神经网络基本结构。正向传播数据传递方向为输入层至隐含层至输出层。将样本数据代入网络进行训练和学习，使输出值不断接近期望值，不断调整权值和阈值，探索变量之间的关系，当误差达到可接受范围，学习过程就会结束；如果输出值误差超出预期范围，则进入误差反向传播，神经网络能够视误差的大小，由后至前逐层进行权重修正，从而降低误差，并重复上述过程，直至误差降低至预期范围。

4.1.1 BP 神经网络模型的构建

（1）网络层次确定

具有隐藏层的 BP 神经网络，所有闭合区间的连续函数去可以通过学习来进行模拟，寻找到数据之间的函数关系，从而提高映射的准确度，然而，这也会使得学习的周期加长，并且可能出现过拟合的现象。基于此，本章选择建立 2 层的 BP 神经网络模型。

（2）网络各层级点数确定

输入层和输出层节点数通常根据具体问题来决定。输入层节点的个数往往与输入的维数相同；输出层的变量为因变量，本文旨在对小麦现货价格进行预测，因此本文构建的神经网络模型的输出节点数量为 1。

BP 神经网络的隐藏层节点数的选取是一个比较复杂的问题，目前尚且没有统一的方法，往往根据经验和多次的试验来确定，不存在一个理想的解析式表示，因此本文参考目前常用的三个公式选择隐藏层节点数。

$$\sum_{i=0}^n C_{n_i}^i > k \quad \text{式 (2)}$$

其中, k 为样本数, n_i 为隐藏层节点数, n 为输入层节点数。如果 $i > n_1$, $C_{n_i}^i = 0$

$$n_1 = \sqrt{n + m} + a \quad \text{式 (3)}$$

其中 m 代表输出节点数, n 代表输入层节点数, a 为 $[1, 10]$ 之间的常数。

$$n_1 = \log_2 n \quad \text{式 (4)}$$

其中 n 为输入层节点数。

因此本文隐藏层节点数在 $[3, 14]$ 之间。

4.1.2 模型训练

本次研究的 BP 神经网络模型的实证部分基于 R 语言中 `neuralnet` 程序包完成。首先将数据进行归一化处理, 由于本文变量均为极大化指标, 归一化公式为:

$$Z_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \quad \text{式 (5)}$$

进行 BP 神经网络模型预测后, 对得到的训练集预测值进行反归一化以得到真正的预测数据。

4.2 单变量 BP 神经网络模型建模预测

经试验将隐藏层数定为 2, 两个隐藏层的神经元个数分别为 5 和 4, 目标收敛误差设置为 0.01, 使用 MSE(均方误差)作为损失函数, \tanh 作为激活函数, 最大迭代次数为 10000 次, 对归一化后的训练集进行建立 2 层的前馈神经网络模型。

本次研究将数据分成 2 部分, 2009 年 1 月至 2021 年 4 月的数据作为训练数据来训练神经网络模型, 2021 年 5 月至 2021 年 9 月的数据作为测试集用来检验神经网络模型。参考相关文献及前文 ACF 检验中推测小麦价格很可能存在一个滞后 12 期的影响因素 (即小麦价格受到一年前同样月份的价格影响), 本次研究采取将小麦价格滞后 1、2、3、12 期作为自变量进行单变量 BP 神经网络模型的建模预测。运行代码, 结果如下:

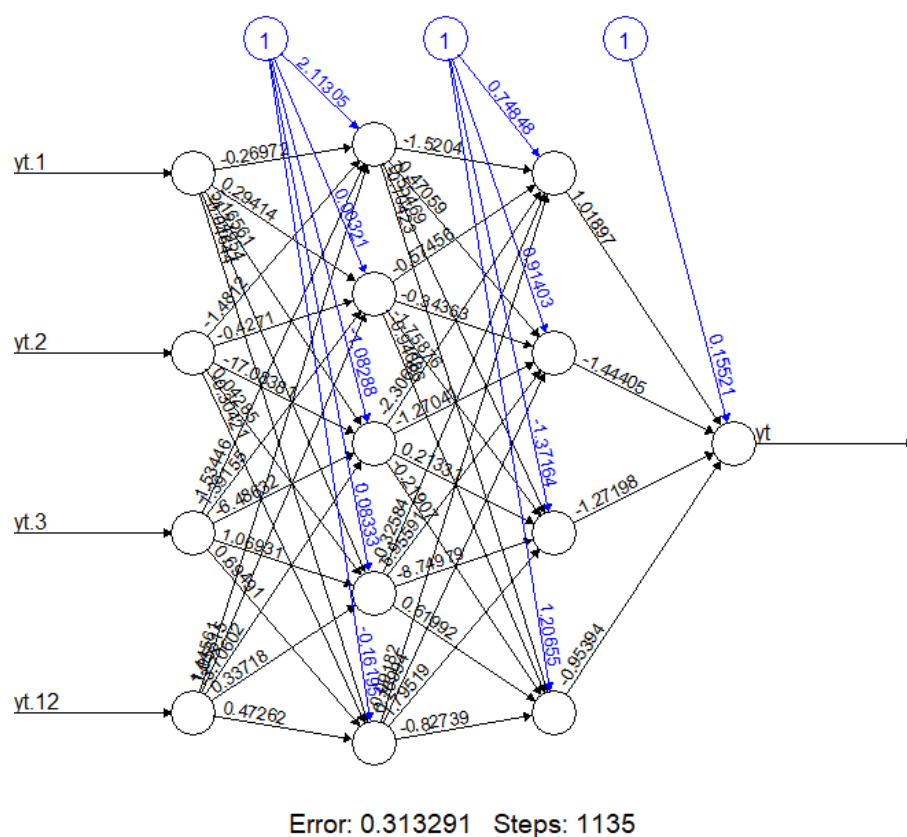


图 10 单变量 BP 神经网络模型

均方根误差可以衡量模型的预测性能。对于单变量 BP 神经网络模型，训练集的均方根误差为 45.5，绘制训练集的拟合情况如下图所示：

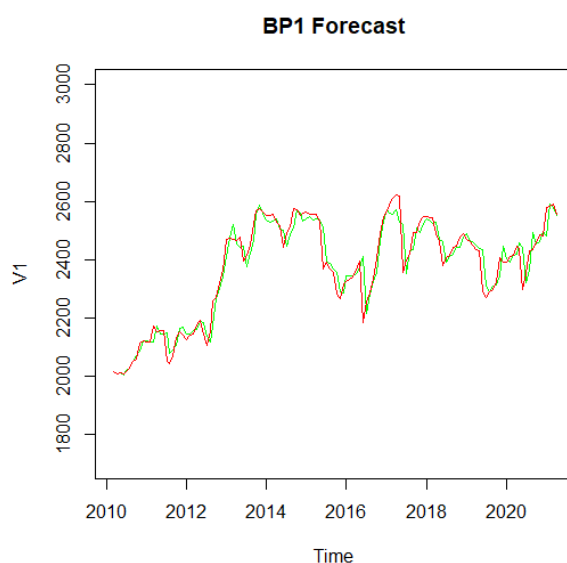


图 11 单变量 BP 神经网络模型训练集预测拟合情况

根据训练集拟合情况图，可以看出数据拟合程度较好。ARIMA 模型的训练集均方根误差为 28.36，与 ARIMA 模型相比，单变量 BP 模型对训练集的拟合效果不如 ARIMA 模型。

用测试集对训练集拟合的模型进行验证，得到测试集的均方根误差为 40.9，比测试集略大。其详细预测结果见下表：

表 12 单变量 BP 神经网络模型测试集预测效果

	真实值	预测值	误差	相对误差	平均相对误差
2021-05	2540.00	2494	-46.00	-1.81%	1.51 %
2021-06	2522.22	2502	-20.22	-0.80%	
2021-07	2561.67	2507	-54.67	-2.13%	
2021-08	2595.56	2551	-44.56	-1.72%	
2021-09	2612.78	2584	-28.78	-1.10%	

结果显示，单变量 BP 神经网络的预测值与实际值相比偏低，且误差逐渐变大。ARIMA 模型的未来 5 期预测值平均相对误差为 1.89%，从预测精度来看，单变量 BP 模型预测结果比 ARIMA 模型预测结果更好。

4.3 多变量 BP 神经网络模型建模预测

ARIMA 模型在外推时不方便加入外生变量，即没有考虑结构变动带来的影响，如政策因素、新冠疫情等突发情况等，而我国小麦现货价格是政策、市场等多方面因素综合影响下的结果，因此用 ARIMA 模型可能导致未来的预测效果不好。因此此处考虑建立多变量的 BP 神经网络模型。

参考多变量时间序列的 BP 神经网络模型的建模方式，本节将第二章得到的 6 个影响因素引入，建立多变量的 BP 神经网络模型对小麦价格进行预测。经试验将隐藏层数定为 2，两个隐藏层的神经元个数分别为 5 和 4，目标收敛误差设置为 0.01，使用 MSE(均方误差)作为损失函数，tanh 作为激活函数，最大迭代次数为 10000 次，对归一化后的训练集进行建立 2 层的前馈神经网络模型。

本次研究将数据分成 2 部分，2009 年 1 月至 2021 年 4 月的数据作为训练数据来训练神经网络模型，2021 年 5 月至 2021 年 9 月的数据作为测试集用来检验神经网络模型。参考相关文献及前文 ACF 检验中推测小麦价格很可能存在一个滞后 12 期的影响因素（即小麦价格受到一年前同样月份的价格影响），本次研究采取尝试分别将小麦价格滞后 1、2、3、12 期作为自变量加入进行多变

量 BP 神经网络模型的建模并取其中均方根误差最小的一个模型进行小麦价格的预测。

表 13 不同滞后期数的多变量 BP 模型预测效果对比

	均方根误差（训练集）	均方根误差（测试集）
滞后 1、12 期	37.8	34.2
滞后 1、2、12 期	34.2	16.7
滞后 1、2、3、12 期	39.4	98.8

根据上表可以看出，整体而言将小麦价格滞后 1、2、12 期加入自变量时，其训练集和测试集的模型均方根误差均为最小，预测效果最优。而将因变量滞后 1、2、3、12 期加入自变量时，其训练集均方根误差虽然也较小，但测试集预测效果呈现出不稳定的现象，向外预测的效果不佳。

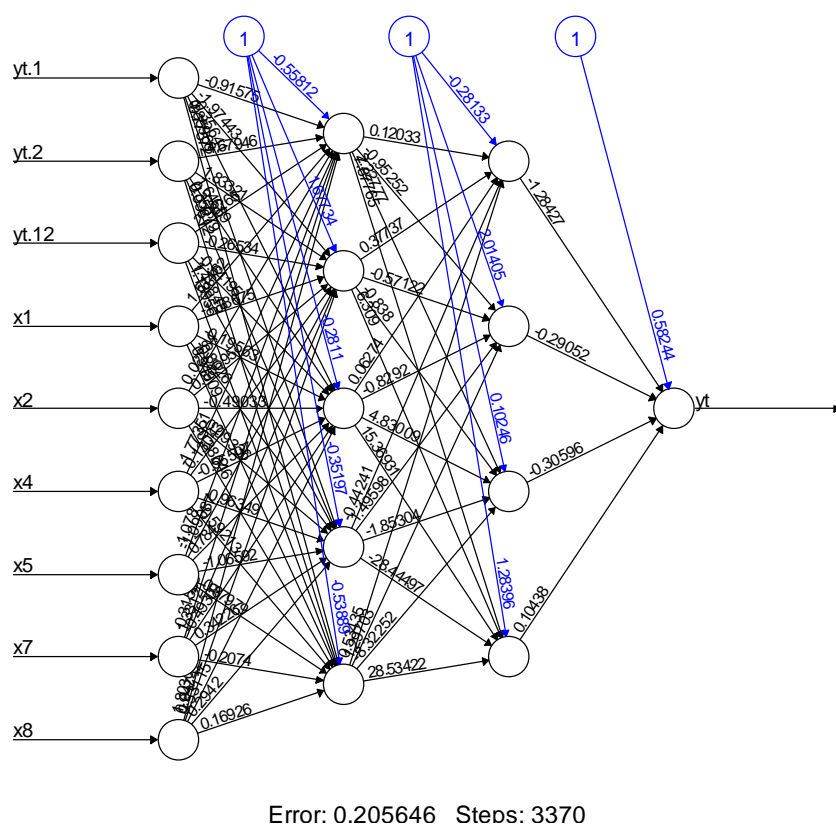


图 12 多变量 BP 神经网络模型

绘制训练集的拟合情况如下图所示：

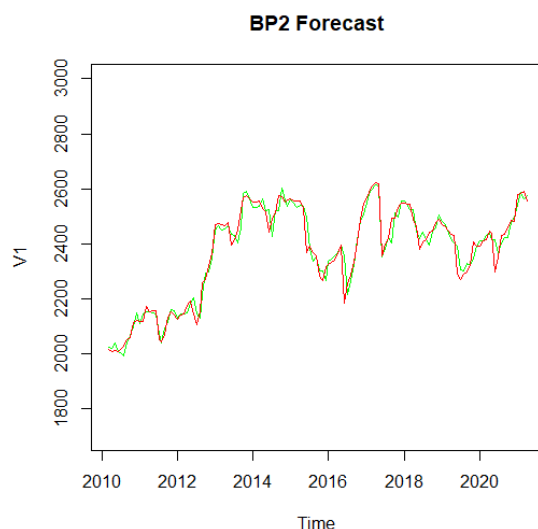


图 13 多变量 BP 神经网络模型训练集预测拟合情况

用测试集对训练集拟合的模型进行验证，预测结果见下表：

表 14 多变量 BP 神经网络模型测试集预测效果

	真实值	预测值	误差	相对误差	平均相对误差
2021-05	2540	2531	-9	-0.35%	0.56%
2021-06	2522	2498	-24	-0.95%	
2021-07	2562	2565	3	0.12%	
2021-08	2596	2621	25	0.96%	
2021-09	2613	2602	-11	-0.42%	

相比单变量 BP 神经网络模型，引入了影响因素后，多变量 BP 神经网络模型的预测效果明显提升，平均相对误差从 1.51% 降为 0.56%，且修正了传统 ARIMA 模型和单变量 BP 模型预测值较真实值偏低的特点。从结果来看，BP 神经网络在处理非线性序列上有着独特的优势。整体而言，多变量 BP 模型比 ARIMA 模型和单变量 BP 模型的预测效果更好，这启示我们在考虑问题时需要从整体出发，对影响问题的因素进行整体的探究，更有利于我们对问题进行全面分析和推断。

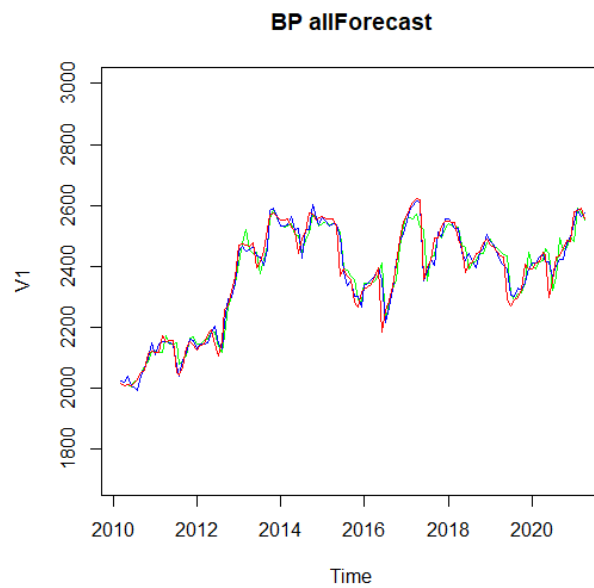


图 14 BP 神经网络拟合结果

(红线：真实值 蓝线：多变量 BP 拟合值 绿线：单变量 BP 拟合值)

5 ARIMA-BP 组合预测模型

残差优化是一种“误差补偿”的思想，第一种模型得到的预测值与真实值的误差输入到第二种模型进行残差优化，使两种单一模型充分发挥各自优势，实现优势互补^{[27][28]}。通过分析可知，小麦价格包含线性和非线性部分，本文选取对线性部分具有较高拟合度的 ARIMA 模型和对非线性部分解决效果较好的 BP 模型，并将其组合进行 ARIMA-BP 残差优化组合模型预测分析，将小麦价格中的线性与非线性部分进行细化，充分提取数据信息，并应用于小麦价格预测研究，进行模型对比分析以探究方法的可行性。

ARIMA-BP 组合模型的构建思路如下^[29]：将小麦价格序列分解为线性趋势和非线性趋势后，首先运用适用于预测线性趋势的 ARIMA 模型预测小麦价格线性部分，得到预测值 1；将真实值与预测值 1 作差得到残差值，将前 3 期的残差值输入到 BP 神经网络模型中对当期残差值进行预测，得到预测值 2，即为非线性趋势部分；将预测值 1 与预测值 2 加和，得到最终 ARIMA-BP 模型的预测值。

(1) 线性趋势预测。由第三章构建的 ARIMA 模型，得到 2009 年 1 月至 2021 年 4 月的原始小麦价格序列与 ARIMA 预测值的残差值如下图：

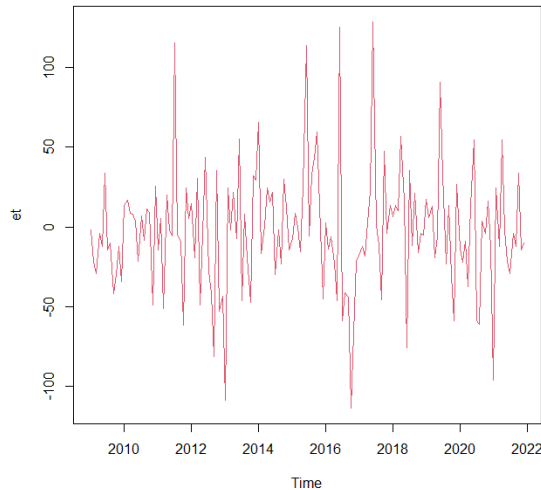


图 15 残差序列图

由残差序列图可以看出，残差值大的波动往往跟随着大的波动，小的波动往往跟随着小的波动，因此构造残差序列的自回归方程有利于帮助我们修正原 ARIMA 模型与真实值之间的差距。

(2) 非线性趋势预测。非线性趋势预测通过构建 BP 神经网络对(1)中的残差值预测实现。首先对残差部分作归一化处理,然后以滞后 1、2、3 期的残差值作为输入预测当期残差,得到 2009 年 1 月至 2021 年 4 月的残差数据的训练样本。

经试验将隐藏层数定为 3,每个隐藏层的神经元个数为 5,目标收敛误差设置为 0.01,使用 MSE(均方误差)作为损失函数, tanh 作为激活函数,最大迭代次数为 10000 次,对归一化后的训练集进行建立 3 层的前馈神经网络模型。

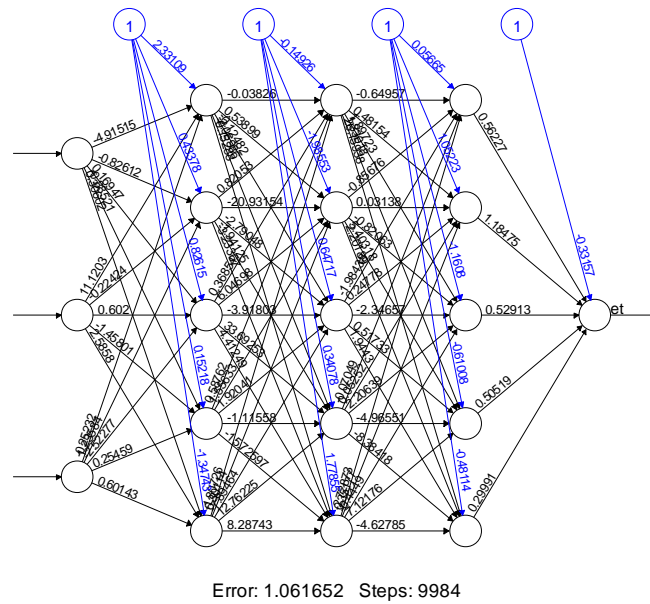


图 16 残差序列 BP 神经网络拟合模型

对得到的预测值进行反归一化处理,得到残差序列的真正预测值。将(1)和(2)中得到的预测值加和,即为 2009 年 1 月至 2021 年 4 月训练集最终的预测值,使用 2021 年 5 月至 2021 年 9 月作为测试集数据进行验证,结果见下表:

表 15 ARIMA-BP 组合模型测试集预测效果

	真实值	ARIMA	预测残差	ARIMA-BP	误差	相对误差	平均相对误差
2021-05	2540.00	2538.918	-1.18	2540.10	0.10	0.00%	1.25%
2021-06	2522.22	2457.344	-31.90	2489.24	-32.98	-1.31%	
2021-07	2561.67	2503.054	-84.00	2587.05	25.38	0.99%	
2021-08	2595.56	2541.991	-11.70	2553.69	-41.87	-1.61%	
2021-09	2612.78	2539.56	-11.70	2551.26	-61.52	-2.35%	

相较于使用单一模型 ARIMA 模型进行预测,各年份误差、相对误差均有所下降,整体的预测效果更优,这主要应该这是由于 ARIMA 模型预测结果往往较真实值偏小,而 ARIMA-BP 模型对此进行了一定程度的修正。

6 结论

6.1 模型结果比较

将本次研究建立的模型进行比较，结果如下表所示：

表 16 预测效果的模型结果比较

	真实值	ARIMA	单变量 BP	多变量 BP	ARIMA-BP
2021-05	2540	2549	2494	2531	2540
2021-06	2522	2467	2502	2498	2489
2021-07	2562	2497	2507	2565	2587
2021-08	2596	2545	2551	2621	2554
2021-09	2613	2549	2584	2602	2551
平均相对误差		1.89%	1.51%	0.56%	1.25%

对比 4 个模型结果，发现：

（1）对于训练集样本，单变量 BP 模型不如 ARIMA 模型拟合效果好；对于测试集，单变量 BP 模型的外推预测效果比 ARIMA 精度稍高，这体现了 BP 神经网络在处理非线性序列的外推预测上有着独特的优势。

（2）考虑到加入影响因素能够有效提升模型的预测精度，根据第二章构建的小麦价格指标体系，构建多变量 BP 神经网络模型，经过验证，加入影响因素后小麦价格测试集的平均相对误差均降低，模型能够最为有效的预测小麦价格。

（3）本次研究尝试将文献提出的 ARIMA-BP 组合预测模型运用于小麦价格序列的建模预测。发现 ARIMA-BP 组合预测模型对于同样只使用了一个时间序列（即小麦现货价格时间序列）进行预测建模的单变量序列的 ARIMA 模型和单变量 BP 神经网络模型，其模型预测精度较高，稳定性更好。但 ARIMA-BP 组合预测模型的预测效果不及多变量 BP 神经网络模型。

6.2 总结

本次研究针对小麦价格序列构建了 3 个单一序列模型。首先构建了 2 个关于单一时间序列（小麦价格序列）的单一模型：ARIMA 模型与单变量 BP 模型。结合以往文献发现组合模型相比单一模型预测精度更高且更为稳定，构建单一序列的组合模型：ARIMA-BP 组合预测模型。

选取小麦价格作为因变量，将小麦价格的相关影响因素经相关性分析、随机森林特征重要性排序后筛选出 6 个贡献最大的指标，构建小麦价格指标体系。

将小麦价格指标体系中的指标序列和小麦序列的滞后值作为输入层构建多变量 BP 神经网络模型，发现将小麦序列滞后 1、2、12 期时模型拟合和预测效果最优。

总体而言，ARIMA 模型与单变量 BP 模型预测和拟合精度相差不大，ARIMA-BP 模型优于 ARIMA 模型和单变量 BP 模型，多变量 BP 模型最优。因此，在进行小麦价格预测时，我们可以考虑以下情况：

（1）当只存在单一时间序列即小麦价格序列时，对比起 ARIMA 模型和单变量 BP 模型，优先考虑使用 ARIMA-BP 组合模型进行小麦价格的外推预测。

（2）当外部统计体系比较健全，对于想要预测的时序影响因素有一定认识时，我们应该优先考虑构建关于想要探究的因变量的影响因素指标体系，将指标体系序列纳入我们的模型预测中。对比单一序列模型，此时优先考虑多变量模型如多变量 BP 模型。这也说明当外部信息越充分，辅助信息利用越多时，预测效果往往越好。

7 研究评价

7.1 特色与创新

(1) 本文建立了有关小麦价格预测的指标体系, 并利用相关系数与随机森林进行变量筛选, 综合考虑了变量间的线性和非线性相关程度, 筛选出对小麦现货价格变化贡献较大的特征, 有利于提高模型的预测能力与拟合速度。

(2) 本项目考虑到小麦现货价格同时具有线性与非线性特征, 选择了适用于线性预测、短期预测精度较高的 ARIMA 模型拟合趋势, 以及选择在处理非线性时间序列数据方面性能较好、泛化能力较强的 BP 神经网络模型对残差建模, 建立的 ARIMA-BP 组合预测模型相比于单一预测模型具有提高精度的优势。

(3) 在建立单变量和多变量 BP 神经网络模型时, 综合考虑了小麦价格时间序列本身的自相关图确定滞后阶数, 发现小麦价格很可能存在一个滞后 12 期的影响因素 (即小麦价格受到一年前同样月份的价格影响), 从而帮助我们确定 BP 神经网络模型的输入层节点数, 对于构建 BP 神经网络有一定的指导意义。

(4) 本文根据加法结构, 分别建立 ARIMA 模型、单变量 BP 神经网络模型、多变量 BP 神经网络模型以及 ARIMA-BP 神经网络组合模型进行预测精度的比较, 并得出了最优模型, 丰富了有关市场价格预测的理论研究。

7.2 不足与展望

(1) 由于时间仓促, 本文拟定的小麦价格指标体系可能不够完善, 希望后续研究中能进一步完善指标体系。

(2) BP 神经网络模型需要大量的样本作为训练样本, 对国家统计工作要求较高。由于我国目前统计发展仍落后于发达国家, 许多经济数据都不能做到以月、日为步长公布, 希望将来统计部门能进一步完善统计指标的统计数据, 帮助未来的研究者有效提高预测精度。

(3) 由于统计的时效性, 本次研究的许多指标数据由于缺乏 2022 年的数据, 不能针对 2022 年 1 月至今的俄乌冲突小麦价格进行有效的预测分析, 希望如果能进行后续的研究, 能有效完善这一情况。

(4) 在后续数据充足的条件下, 考虑使用更丰富的模型对小麦价格进行预测研究, 如 VAR 模型、XGBoost 模型和 LSTM 模型等。

参考文献

- [1]朱晶,张庆萍.中国利用俄罗斯、乌克兰和哈萨克斯坦小麦市场分析[J].农业经济问题,2014,35(04):42-50+111.DOI:10.13246/j.cnki.iae.2014.04.007.
- [2]杨惠珍,韦敬楠,张立中.我国粮食期货市场价格发现功能的实证分析——以玉米和小麦市场为例[J].价格月刊,2017(05):19-23.DOI:10.14076/j.issn.1006-2025.2017.05.04.
- [3]刘平,方旖旎.中国小麦进口结构及竞争力比较研究[J].市场周刊,2021,34(08):144-146.
- [4]高云,孙一铮,郭新宇,矫健.国内外小麦价格相关性及其影响因素研究[J].价格理论与实践,2018(05):71-74.DOI:10.19851/j.cnki.cn11-1010/f.2018.05.018.
- [5]林春霞.国际粮价将呈高位震荡趋势[N].中国经济时报,2021-09-29(001).DOI:10.28427/n.cnki.njjsb.2021.002172.
- [6]赵荣,曹洁,朱婷婷.浅议中国小麦的国际贸易及竞争力提升[J].滁州学院学报,2021,23(01):14-19+41.
- [7]朱聪,曲春红,王永春,赵伟.新一轮国际粮食价格上涨:原因及对中国市场的影响[J/OL].中国农业资源与区划:1-13[2022-03-08].<http://kns.cnki.net/kcms/detail/11.3513.S.20211130.1225.010.html>.
- [8]钟钰,陈希,牛坤玉.粮食出口限制政策的实施效果与我国应对——来自部分小麦出口国的证据[J].经济纵横,2021(08):29-39+137.DOI:10.16528/j.cnki.22-1054/f.202108029.
- [9]潘寅茹.乌克兰为何被称为“欧洲粮仓”?玉米小麦出口均占全球10%以上[N].第一财经日报,2022-03-04(010).
- [10]曹慧.新冠肺炎疫情对国内外小麦产业的冲击:趋势、问题与建议[J].世界农业,2021(01):4-10+27.DOI:10.13856/j.cnki.1097/s.2021.01.001.
- [11]申洪源.2021年中国小麦市场分析[N].粮油市场报,2022-02-15(B02).DOI:10.28553/n.cnki.nlysc.2022.000245.
- [12]邵海鹏.小麦也“疯狂”:进口首次突破配额,价格站上历史高位[N].第一财经日报,2022-03-02(A06).
- [13]邵海鹏.CBOT小麦期货连续涨停,国内外小麦价格倒挂影响几何[N].第一财经日报,2022-03-04(A10).

- [14]孙致陆. 贸易开放背景下国际小麦贸易市场势力实证分析[J]. 华中农业大学学报 (社 会 科 学 版), 2019(04):1-14+169. DOI:10.13300/j.cnki.hnwkxb.2019.04.001.
- [15]陈植. 芝商所小麦玉米期货迭创新高 对冲基金押宝“新避险投资”[N]. 21 世纪经济报道, 2022-03-04(005).
- [16]Haile M G, Kalkuhl M. Braun J V Worldwide Acreage and Yield Response to International Price Change and Volatility: A D ynamic Panel Data Analysis for Wheat, Rice, Corn, and Soybeans M//Food Price Volatility and Its Implications for Food Secu rity and Policy. Springer International Publishing, 2016
- [17]Pal D, Mitra S K Interdependence between crude oil and world food prices: A detrended cross correlation analysis. Phys ica A Statistical Mechanics&Its Applications, 2018
- [18]Ahumada H, Cornejo M Forecasting food prices: The case of corn, soybeans and wheatly] International Journal of Foreca sting. 2016.
- [19]柳燕子. 基于多模型分析的小麦价格预测研究[J]. 山西农经, 2018(04):148. DOI:10.16675/j.cnki.cn14-1065/f.2018.04.103.
- [20]李雪, 韩一军, 付文阁. 最低收购价政策对小麦市场价格波动影响的实证分析[J]. 华 中 农 业 大 学 学 报 (社 会 科 学 版), 2018(02):1-7+154. DOI:10.13300/j.cnki.hnwkxb.2018.02.001.
- [21]张瑞娟, 高芸. 国内外小麦价格联动关系研究[J]. 价格理论与实践 , 2016(07):112-115. DOI:10.19851/j.cnki.cn11-1010/f.2016.07.029.
- [22]新冠肺炎疫情对中国粮食价格的影响效应及作用机制. 崔海莹
- [23]战玉锋, 刘放. 我国小麦期货价格影响因素实证分析[J]. 沈阳工业大学学报(社会科学版), 2012, 5(3): 223-231.
- [24]张恒, 赵宇洋, 安起光. 基于 BEKK-GARCH 模型的小麦期现货市场价格波动溢出性分析[J]. 粮食科技与经济, 2022, 47(1): 40-44.
- [25]张鑫, 吴海涛, 曹雪虹. Hadoop 环境下基于随机森林的特征选择算法[J]. 计算机技术与发展, 2018, 28(07):88-92.
- [26]顾鑫. 基于 ARIMA-LSTM-XGBoost 组合模型的中国经济发展预测[D]. 青岛大学, 2021.

- [27] GAO W J, LI Z, MIN X, et al. Application system response time pre - diction based on ARIMA-LSTM composite mode [J] . Computer & Digital Engineering, 2021, 49 (5) : 880-885. 高文俊, 李志, 闵星, 等 . 基于 ARIMA-LSTM 组合模型的应用系统响应时间预测 [J] . 计算机与数字工程, 2021, 49 (5) : 880-885.
- [28] XU D D, JIANG Z X. End-to-end speech recognition based on deeply optimized residuals convolutional neural network [J] . Application Research of Computers, 2020, 37 (S2) : 139-141. 徐冬冬, 蒋志翔 . 基于深度优化残差卷积神经网络的端到端语音识别 [J] . 计算机应用研究, 2020, 37 (S2) : 139-141.
- [29] Du Y. Predicting China GDP Based on ARIMA-BP Neural Network Model[A]. Information Engineering Research Institute(USA) 、 Asia Pacific Human-Computer Interaction Research Center(Hong Kong). Proceedings of 2018 IEA 3rd International Conference on Frontier of Computer Science and Information Engineering(ICFCI 2018) [C]. Information Engineering Research Institute(USA) 、 Asia Pacific Human-Computer Interaction Research Center(HongKong): 智能信息技术应用学会, 2018:6.

附录

附录 1: R 语言代码

#相关系数图

```
library(PerformanceAnalytics)
```

```
library(corrplot)
```

```
x=read.table('clipboard',header=T);x
```

```
chart.Correlation(x, histogram=TRUE, pch=19)
```

```
corrplot.mixed(cor(x,method="pearson"),lower.col="black", number.cex= .7,mar =  
c(2,2,3,2))
```

#随机森林特征重要性排序

```
wine = read.table("clipboard",header=T);wine
```

#将数据集分为训练集和测试集,比例为 7:3

```
train_sub = sample(nrow(wine),7/10*nrow(wine))
```

```
train_data = wine[train_sub,]
```

```
test_data = wine[-train_sub,]
```

```
library(pROC) #绘制 ROC 曲线
```

```
library(randomForest)
```

#随机森林

```
wine_randomforest <- randomForest(yt~x1+x2+x3+x4+x5+x6+x7+x8+x9,  
                                   data = train_data,  
                                   ntree =500,  
                                   mtry=3,  
                                   importance=TRUE ,  
                                   proximity=TRUE)
```

#查看变量的重要性

```
wine_randomforest$importance
```

```
varImpPlot(wine_randomforest, main = "variable importance",font = 2,lwd = 5,col =  
"red",col.axis = "blue",font.axis = 2,cex = 1,cex.axis = 5,pch = 16,cex.lab = 0.7,font.lab  
= 2)
```

#arima 模型

```
library(tseries)
```

```
library(forecast)
library(fUnitRoots)
library(fGarch)
y1=read.table("clipboard")
head(y1)
na.omit(y1)
y11<-ts(y1,start=c(2009,1),end=c(2021,12),frequency=12)
plot(y11,type="l",xlab="时间",ylab="小麦现货价格")
ts.plot(diff(y11))
acf(y11,main="自相关图")#拖尾
pacf(y11,main="偏自相关图")#拖尾
trainy11<-ts(as.vector(y11[1:148]),frequency=12,start=c(2009,1))
#看拆掉之后的训练集图形
plot(trainy11,type="l",xlab="时间",ylab="小麦现货价格")
acf(trainy11,main="自相关图") #画出自相关图
pacf(trainy11,main="偏自相关图") #画出偏自相关图
tsdisplay(trainy11)
adf.test(trainy11)#接受原假设，存在单位根
s1<-diff(trainy11,1)
plot(s1,type="l",xlab="时间",ylab="一阶差分结果")
acf(s1,main="自相关图") #画出自相关图
pacf(s1,main="偏自相关图") #画出偏自相关图
#进一步使用 adf 检验，看一下是否存在单位根（验证平稳性）
adf.test(s1)#拒绝原假设，序列平稳
tsdisplay(s1)

a1=arima(trainy11,order=c(2,1,1),method="ML")
a2=arima(trainy11,order=c(12,1,1),fixed=c(NA,NA,0,0,0,0,0,0,0,0,0,NA,NA),transform
.pars = FALSE)#AIC=907.17
res2<-a2$residuals^2
tsdiag(a2)
Box.test(a2$residuals, type="Ljung-Box",lag=12)#接受 h0
Box.test(res2, type="Ljung-Box",lag=12)#拒绝 H0
#然后使用 tsdiag 看一下各自的结果，图中表明残差标准差基本都在[-1,1]之间，
#残差的自回归都为 0（两虚线内），Ljung-Box 检验的 p 值都在 0.05 之上，结果
不错。
```

```
qqnorm(a2$residuals)
qqline(a2$residuals)

fore2<-forecast(a2,h=5,level=c(99.5))
plot(fore2,ylim=c(1700,3000))
lines(fore2$fitted,col="green")
lines(y11,col="red")

#神经网络
library(neuralnet)
options(digits=3)
x=read.table("clipboard",header=T);x
maxs=apply(x,2,max);maxs
maxs[1]
mins=apply(x,2,min);mins
mins[1]
m=(maxs-mins)[1];m
x1=as.data.frame(scale(x,center=mins,scale=maxs-mins));x1
attach(x1)
train.cse=x1[1:146,];train.cse
test.cse=x1[147:151,];test.cse

#滞后 1 期
cse.net=neuralnet(yt~yt.1+ yt.12+x1+x2+x4+x5+x7+x8,data=train.cse,hidden=c(5,4),
threshold=0.01,act.fct="tanh",linear.output=T)
#滞后 2 期
cse.net=neuralnet(yt~yt.1+yt.2+
yt.12+x1+x2+x4+x5+x7+x8,data=train.cse,hidden=c(5,4),
threshold=0.01,act.fct="tanh",linear.output=T)
#滞后 3 期
cse.net=neuralnet(yt~yt.1+yt.2+yt.3+
yt.12+x1+x2+x4+x5+x7+x8,data=train.cse,hidden=c(5,4),
threshold=0.01,act.fct="tanh",linear.output=T)
#neuralnet 包的误差函数默认为 sse，没有 mse，本次研究只有 1 个输出神经元，
可以理解为 mse
#单变量预测
```

```
cse.net=neuralnet(yt~yt.1+yt.2+yt.3+yt.12,data=train.cse,hidden=c(5,4),
threshold=0.01,act.fct="tanh",linear.output=T)
ythat=cse.net$net.result[[1]]*m+mins[1];ythat
plot(cse.net)

rmse=function(x,y){sqrt(mean((x-y)^2))}
tyt=train.cse$yt*m+mins[1];tyt
rmse(ythat,tyt)

#测试集
ythat2=compute(cse.net,test.cse[,2:10])$net.result*m+mins[1];ythat2
tyt2=test.cse$yt*m+mins[1];tyt2
rmse(ythat2,tyt2)

#将数据结果写入
write.table(ythat[,1], file = "C:\\Users\\mujin1922\\Desktop\\Rresult1", append =
FALSE, quote = TRUE, sep = " ",
eol = "\n", na = "NA", dec = ".", row.names = TRUE,
col.names = TRUE, qmethod = c("escape", "double"),fileEncoding = "")
```

附录 2：原始数据

	yt	x1	x2	x3	x4	x5	x6	x7	x8	x9
2009-01	1790.00	104.82	18.79	15.50	8000.86	101.00	101.30	3247.00	634.00	163.40
2009-02	1817.00	107.62	20.48	15.50	7062.93	98.40	101.00	92033.00	625.25	164.58
2009-03	1855.00	107.41	25.51	15.50	7608.92	98.80	100.30	28063.00	577.25	168.80
2009-04	1862.00	104.93	25.89	15.50	8168.12	98.50	100.50	35876.00	540.75	180.07
2009-05	1878.00	101.56	25.74	15.50	8500.33	98.60	101.20	71562.00	548.60	179.39
2009-06	1836.00	99.58	28.46	15.50	8447.00	98.30	101.00	192905.00	623.00	151.80
2009-07	1853.00	99.18	28.42	15.50	9171.61	98.20	102.10	85771.00	576.25	152.04
2009-08	1866.00	98.40	28.53	15.50	9496.28	98.80	102.70	8436.00	593.67	150.59
2009-09	1920.00	97.47	29.31	15.50	9712.28	99.20	102.80	74838.00	566.25	167.28
2009-10	1960.00	95.93	29.42	15.50	9712.73	99.50	103.20	13241.00	566.25	171.69
2009-11	1977.00	95.72	29.74	15.50	10344.84	100.60	103.30	94683.00	606.00	164.61
2009-12	2021.00	97.26	27.68	15.50	10428.05	101.90	103.90	203476.00	598.00	167.34
2010-01	2006.00	97.98	25.98	16.00	10067.33	101.50	104.70	86638.00	584.75	161.91
2010-02	2005.00	100.45	25.52	16.50	10325.26	102.70	104.20	95851.00	540.13	158.93
2010-03	2014.00	99.00	22.50	16.50	10856.63	102.40	107.90	192283.00	502.23	157.27
2010-04	2006.00	98.95	21.48	16.50	11008.61	102.80	106.60	132287.00	472.48	163.24
2010-05	2011.00	101.43	21.00	17.00	10136.63	103.10	108.00	191412.00	458.55	152.87
2010-06	2005.00	101.91	18.46	17.00	9774.02	102.90	108.50	191419.00	470.68	163.92
2010-07	2016.00	100.81	17.60	17.00	10465.94	103.30	107.80	133085.00	486.86	175.60
2010-08	2028.00	99.77	19.20	17.00	10014.72	103.50	107.30	90221.00	519.95	205.84
2010-09	2050.00	99.76	18.96	17.00	10788.05	103.60	104.10	91340.00	532.38	235.70
2010-10	2058.00	98.30	19.30	17.00	11118.49	104.40	103.80	12237.00	543.14	238.24
2010-11	2113.00	100.19	19.50	18.00	11006.02	105.10	102.90	7657.00	536.78	250.63
2010-12	2120.00	101.46	19.72	18.50	11577.51	104.60	100.40	6294.00	528.38	265.29
2011-01	2117.00	101.81	17.20	19.00	11891.93	104.90	99.90	60063.00	532.80	293.40
2011-02	2116.00	102.36	15.70	19.50	12226.34	104.94	99.60	116923.00	508.96	290.43
2011-03	2171.00	100.78	16.60	20.00	12319.73	105.38	107.60	75949.00	500.57	318.94
2011-04	2153.00	99.65	15.30	20.50	12810.54	105.34	106.60	11042.00	500.55	308.58
2011-05	2155.00	99.81	15.10	21.00	12569.79	105.52	105.80	10893.00	518.09	310.54
2011-06	2157.00	100.12	15.90	21.50	12414.34	106.36	108.10	280648.00	546.19	300.83
2011-07	2050.00	100.43	14.70	21.50	12143.24	106.45	105.60	108482.00	577.30	310.31
2011-08	2044.00	101.25	13.50	21.50	11613.53	106.15	105.00	128664.00	615.55	296.30
2011-09	2068.00	104.47	13.00	21.50	10913.38	106.07	103.40	177664.00	602.14	274.85
2011-10	2133.00	105.76	12.90	21.50	11955.01	105.50	100.50	60649.00	609.09	274.28
2011-11	2150.00	106.45	12.70	21.50	12045.68	104.23	97.00	8633.00	580.91	258.52
2011-12	2139.00	108.08	13.60	21.00	12217.56	104.07	100.50	218443.00	541.06	272.92
2012-01	2123.00	109.58	12.40	21.00	12632.91	104.50	103.90	213976.00	547.48	279.51
2012-02	2140.00	108.11	13.00	20.50	12952.07	103.20	105.00	373195.00	577.05	280.72
2012-03	2145.71	108.36	13.40	20.50	13212.04	103.60	100.00	537103.00	585.95	274.22
2012-04	2173.57	108.20	12.80	20.50	13213.63	103.40	103.00	293963.00	612.43	268.88
2012-05	2189.29	109.09	13.20	20.00	12393.45	103.00	104.20	557627.00	606.14	267.33
2012-06	2145.71	108.92	13.60	20.00	12880.09	102.20	99.30	217779.00	578.36	333.00
2012-07	2102.86	108.83	13.90	20.00	13008.68	101.80	98.20	252649.00	582.87	332.23
2012-08	2156.43	108.39	13.50	20.00	13090.84	102.00	99.40	252154.00	590.50	320.71
2012-09	2255.71	107.56	14.80	20.00	13437.13	101.90	100.80	524906.00	584.74	321.38
2012-10	2268.57	107.81	14.10	20.00	13096.46	101.70	106.10	351450.00	590.73	321.60
2012-11	2319.29	109.70	13.90	20.00	13025.58	102.00	105.10	118941.00	565.52	308.75
2012-12	2364.29	110.04	13.80	20.00	13104.14	102.50	103.70	7293.00	573.39	303.36
2013-01	2470.00	111.58	15.90	20.00	13860.58	102.03	104.50	181388.00	574.07	302.61
2013-02	2472.14	113.43	15.20	20.00	14054.49	103.22	108.20	223876.00	564.52	309.58
2013-03	2470.71	113.77	15.70	20.00	14578.54	102.07	102.60	289987.00	553.73	279.76
2013-04	2466.43	114.89	16.10	20.00	14839.80	102.39	103.70	204264.00	552.09	295.37
2013-05	2477.14	115.87	15.80	20.00	15115.57	102.10	99.00	284646.00	546.25	297.11
2013-06	2396.43	115.92	14.00	20.00	14909.60	102.67	97.00	221425.00	538.26	279.01
2013-07	2413.57	116.90	14.50	20.00	15499.54	102.67	97.20	311137.00	509.77	234.94
2013-08	2448.57	116.97	14.70	20.00	14810.31	102.57	97.80	421056.00	470.00	207.49
2013-09	2519.29	117.30	14.20	20.00	15129.67	103.05	99.80	719874.00	453.26	201.78
2013-10	2568.57	116.05	14.30	20.00	15545.75	103.21	102.90	1310449.00	448.81	199.18
2013-11	2574.29	117.47	14.20	20.00	16086.41	103.02	98.90	972141.00	447.55	197.55
2013-12	2561.43	118.54	13.60	20.00	16576.66	102.50	102.30	395215.00	441.04	198.77
2014-01	2552.14	120.80	13.20	20.00	15698.85	102.49	101.10	731935.57	447.00	209.36
2014-02	2550.71	120.71	13.30	20.00	16321.71	101.95	103.10	583985.08	434.14	222.39
2014-03	2555.00	117.15	12.10	20.00	16457.66	102.38	107.90	541479.87	409.94	222.43
2014-04	2527.14	114.52	13.20	20.00	16580.84	101.80	104.80	375355.19	403.59	216.98
2014-05	2515.00	113.89	13.40	20.00	16717.17	102.48	102.30	226461.69	414.67	202.63
2014-06	2441.43	113.89	14.70	20.00	16826.60	102.34	104.70	114271.02	428.00	182.29
2014-07	2487.14	114.53	13.50	20.00	16563.30	102.29	104.40	120423.58	440.38	176.48
2014-08	2510.71	116.46	12.80	20.00	17098.45	101.99	103.80	36223.17	436.50	163.13
2014-09	2573.57	119.24	12.90	20.00	17042.90	101.63	105.40	134612.88	432.70	163.38
2014-10	2570.00	121.13	12.60	20.00	17390.52	101.60	103.40	41426.64	419.00	178.71
2014-11	2552.86	123.80	12.30	20.00	17828.24	101.44	105.50	45572.78	410.74	178.71
2014-12	2558.57	125.29	12.20	20.00	17823.07	101.51	105.80	53208.73	409.68	174.79
2015-01	2563.57	127.37	10.80	20.00	17164.95	100.76	105.70	78881.78	409.50	173.76
2015-02	2555.71	129.30	12.50	19.50	18132.70	101.43	109.80	134158.30	400.73	174.29
2015-03	2557.14	130.34	11.60	19.50	17776.12	101.38	107.10	154667.25	392.41	172.15
2015-04	2555.71	129.57	10.10	18.50	17840.52	101.51	107.60	245241.12	382.43	166.35

2015-05	2527.86	127.89	10.80	18.50	18010.68	101.23	109.90	368868.52	370.55	166.81
2015-06	2366.43	128.68	11.80	18.50	17619.51	101.39	105.50	431837.16	387.74	179.61
2015-07	2390.71	130.69	13.30	18.50	17689.86	101.65	104.48	302986.93	375.57	162.74
2015-08	2367.86	130.10	13.30	18.50	16528.03	101.96	104.00	344444.15	358.77	166.08
2015-09	2355.71	130.03	13.10	18.00	16284.70	101.60	105.60	209033.45	360.45	171.45
2015-10	2282.14	128.85	13.50	17.50	17663.54	101.27	103.80	318181.06	358.38	166.09
2015-11	2267.14	130.87	13.70	17.50	17719.92	101.49	104.10	199606.12	354.35	164.00
2015-12	2317.86	130.21	13.30	17.50	17425.03	101.60	103.70	218955.53	359.48	161.10
2016-01	2323.57	129.72	14.00	17.50	16466.30	101.80	104.00	240000.00	373.19	159.72
2016-02	2331.43	130.94	13.30	17.50	16516.50	102.30	104.40	110000.00	370.48	159.23
2016-03	2341.43	128.72	13.40	17.00	17685.09	102.30	100.00	210000.00	376.05	164.47
2016-04	2363.57	126.33	12.80	17.00	17773.64	102.33	101.00	300000.00	408.82	169.00
2016-05	2396.43	125.33	11.80	17.00	17787.20	102.04	99.80	440000.00	420.68	179.91
2016-06	2183.57	123.32	11.80	17.00	17929.99	101.88	102.90	480000.00	456.24	161.82
2016-07	2254.29	122.09	10.20	17.00	18432.24	101.77	106.80	310000.00	414.22	150.22
2016-08	2285.71	121.19	11.40	17.00	18400.88	101.34	105.60	350000.00	384.00	148.51
2016-09	2330.71	121.61	11.50	17.00	18308.15	101.92	104.60	440000.00	367.67	152.62
2016-10	2406.43	121.36	11.60	17.00	18142.42	102.10	107.20	270000.00	360.82	151.33
2016-11	2487.14	122.04	11.40	17.00	19123.58	102.25	108.60	30000.00	367.55	152.73
2016-12	2547.14	123.12	11.30	17.00	19762.60	102.08	108.40	230000.00	372.55	160.01
2017-01	2565.71	123.64	10.70	17.00	19864.09	102.55	109.20	340000.00	368.50	162.96
2017-02	2585.56	122.44	10.40	17.00	20812.24	100.80	112.60	230000.00	367.78	158.98
2017-03	2610.56	120.79	10.10	17.00	20663.22	100.90	111.00	510000.00	374.50	156.54
2017-04	2622.22	119.76	9.80	17.00	20940.51	101.20	113.40	600000.00	402.91	158.60
2017-05	2616.67	118.86	9.10	17.00	21008.65	101.50	112.00	510000.00	444.68	157.97
2017-06	2354.44	119.20	9.10	17.00	21349.63	101.50	113.30	460000.00	419.95	157.61
2017-07	2400.00	119.34	8.90	17.00	21891.12	101.40	114.60	240000.00	405.65	148.54
2017-08	2418.89	120.07	8.60	17.00	21948.10	101.80	114.70	290000.00	413.14	147.32
2017-09	2492.22	121.69	9.00	17.00	22405.09	101.60	118.60	440000.00	407.23	148.72
2017-10	2492.22	122.03	8.90	17.00	23377.24	101.90	123.90	410000.00	404.05	148.79
2017-11	2530.56	121.93	9.10	17.00	24272.35	101.70	121.30	180000.00	407.95	149.09
2017-12	2547.78	122.14	8.10	17.00	24719.22	101.80	122.60	210000.00	422.30	155.92
2018-01	2545.56	123.28	8.60	17.00	26149.39	101.50	122.30	250000.00	417.85	163.46
2018-02	2542.22	126.20	8.80	17.00	25029.20	102.90	124.00	80000.00	409.77	172.08
2018-03	2543.33	124.32	8.20	17.00	24103.11	102.10	122.30	310000.00	432.38	175.60
2018-04	2490.00	124.61	8.30	16.00	24163.15	101.80	122.90	370000.00	433.22	179.18
2018-05	2461.67	125.36	8.30	16.00	24415.84	101.80	122.90	630000.00	413.14	165.14
2018-06	2378.89	124.52	8.00	16.00	24271.41	101.90	118.20	310000.00	378.68	156.55
2018-07	2408.33	120.92	8.50	15.50	25415.19	102.10	119.70	140000.00	383.43	162.52
2018-08	2414.44	119.84	8.20	15.50	25964.82	102.30	118.60	140000.00	383.00	154.95
2018-09	2440.56	120.53	8.30	15.50	26458.31	102.50	118.50	280000.00	392.04	160.37
2018-10	2445.56	120.03	8.00	14.50	25115.76	102.50	119.10	220000.00	390.82	160.88
2018-11	2477.22	119.98	8.00	14.50	25538.46	102.20	122.10	140000.00	380.33	167.49
2018-12	2488.89	121.09	8.10	14.50	23327.46	101.90	123.00	240000.00	383.74	166.85
2019-01	2470.00	122.10	8.40	13.50	24999.67	101.70	123.70	610000.00	390.55	169.66
2019-02	2460.00	124.53	8.00	13.50	25916.00	101.50	126.00	140000.00	383.14	166.32
2019-03	2447.22	124.38	8.60	13.50	25928.68	102.30	124.10	290000.00	397.41	161.65
2019-04	2432.22	124.01	8.50	13.50	26592.91	102.50	125.30	220000.00	390.30	171.18
2019-05	2429.44	122.19	8.50	13.50	24815.04	102.70	123.40	210000.00	396.25	195.17
2019-06	2291.11	120.43	8.50	13.50	26599.96	102.70	125.90	310000.00	395.96	189.49
2019-07	2271.11	121.21	8.10	13.50	26864.27	102.80	124.40	220000.00	414.68	163.63
2019-08	2289.44	119.84	8.20	13.50	26403.28	102.80	122.40	120000.00	415.43	157.37
2019-09	2295.56	120.00	8.40	13.00	26916.83	103.00	124.10	160000.00	399.13	167.30
2019-10	2324.44	120.89	8.40	13.00	27046.23	103.80	124.30	280000.00	394.57	166.36
2019-11	2407.22	122.63	8.20	13.00	28051.41	104.50	124.60	450000.00	397.00	167.06
2019-12	2392.22	122.25	8.70	13.00	28538.44	104.50	126.60	500000.00	428.52	171.81
2020-01	2390.00	124.80	8.40	12.50	28256.03	105.40	126.40	340000.00	430.25	168.74
2020-02	2410.56	126.45	8.80	12.50	25409.36	105.20	118.90	340000.00	471.68	162.18
2020-03	2415.00	126.20	10.10	12.50	21917.16	104.30	122.20	590000.00	542.82	147.01
2020-04	2445.56	125.74	11.10	12.50	24345.72	103.30	116.40	360000.00	498.86	143.99
2020-05	2436.67	123.61	11.10	12.50	25383.11	102.40	115.80	810000.00	503.05	148.04
2020-06	2296.67	121.50	11.10	12.50	25812.88	102.50	112.60	910000.00	464.48	152.59
2020-07	2350.00	122.48	10.70	12.50	26428.32	102.70	117.20	930000.00	484.43	149.39
2020-08	2429.44	122.79	10.40	12.50	28430.05	102.40	116.40	700000.00	491.91	166.60
2020-09	2435.56	124.92	10.90	12.50	27781.70	101.70	120.50	1070000.00	458.86	186.77
2020-10	2455.56	125.76	10.50	12.50	26501.60	100.50	121.70	630000.00	459.86	190.63
2020-11	2486.67	126.25	10.70	12.50	29638.64	99.50	124.00	800000.00	499.35	199.18
2020-12	2480.00	126.35	10.10	12.50	30606.48	100.20	122.10	880000.00	518.48	231.75
2021-01	2580.00	127.65	9.40	12.50	29982.62	99.70	122.80	1500000.00	535.75	245.55
2021-02	2581.11	129.57	10.10	12.50	30932.37	99.80	127.00	990000.00	502.35	245.76
2021-03	2589.44	128.80	9.40	12.50	33874.85	100.40	122.20	440000.00	470.27	268.62
2021-04	2555.56	127.20	8.10	12.50	34529.45	100.90	121.50	900000.00	477.19	304.30
2021-05	2540.00	126.98	8.30	12.50	34502.51	101.30	121.80	790000.00	454.86	292.52
2021-06	2522.22	126.86	8.60	12.50	34935.47	101.10	122.80	750000.00	414.86	278.45
2021-07	2561.67	127.72	8.30	12.00	35360.73	101.00	117.80	880000.00	386.77	257.02
2021-08	2595.56	127.74	8.20	12.00	33843.92	100.80	117.50	710000.00	387.00	235.82
2021-09	2612.78	127.80	8.30	12.00	35819.56	100.70	121.20	640000.00	388.90	239.60