

基于ARIMA模型的万科公司股票收盘价研究

山西财经大学 石琳枫

摘要: 为研究预测房地产股票短期股价,选取万科公司2019年1月15日到2020年2月10日股票收盘价共254个样本数据为研究对象,首先利用ADF平稳性检验来判断时间序列是否平稳,然后,对前242个数据用R语言建立ARIMA模型,并基于该模型对后12个数据进行预测结果的检验。结果表明,构建的ARIMA模型能较准确地预测万科A的短期股价,预测结果可以为投资者和政府部门的决策作为参考。

关键词: 股票收盘价; 万科A; 平稳性检验; ARIMA模型; 模型预测

中图分类号: F830.91

文献标识码: A

DOI: 10.12245/j.issn.2096-6776.2021.03.16

1 研究现状

在许多西方发达资本主义国家,学者对经济预测进行了深入的研究,已经构建了较为成熟的理论和体系。J.H.Stock(1996)在时间序列的经济预测方面取得了重大进展。时间序列预测是基于对经济过程的理解,然后利用历史经验规律,做出未来的预测。经济预测已经广泛应用于各领域,其中就包括制定货币和财政政策、国家和地方预算等。经济预测的关键要素包括:选择适合当前问题的预测模型,评估和传递与预测相关的不确定性,并防范模型的不稳定性^[1]。LIU和MORLEY(2009)认为传统计量经济模型存在产生严重偏差的可能。基于该问题,学者们尝试不断地改进传统计量模型,其中就包括基于GARCH的股票价格预测模型^[2]。Lam等人(2013)选取2011年6月至2012年10月收集的每日库存记录用于模型建设,用神经网络方法针对大同煤矿公司的短期股价趋势进行了预测,研究结果表明,该模型可以预测股票价格的走势^[3]。

目前国内学者对股票价格预测已经进行了较深入的研究。孔华强(2006)为研究深证100和上证180的波动性,运用EGARCH(1,1)-M模型对所研究指数进行了拟合^[4]。江涛(2010)基于GARCH(1,1)模型度量和分析了股票市场的风险问题,认为市场风险是影响市场收益率的一个主要因素^[5]。李嘉松(2017)以沪深300指数为例,使用Eviews软

件建立ARIMA模型,指出造成其实际值和预测值之间有差异的原因,为投资者提供沪深300指数的预测方法,来作为未来进行股票投资和金融衍生产品投资的参考^[6]。王惠星等(2017)建立时间序列模型对上证50指数数据进行定量分析,在分析的基础上开展对指数未来趋势的预测,结果发现该模型对小样本投资组合的市场走势具有较好的模拟效果^[7]。张颖超等(2019)在建立ARIMA(4,1,4)模型的基础上,尝试利用该模型预测股价的变动趋势,结果表明,预测效果与预测的时间范围有关,在短期内预测精确度较高,而在长期预测效果欠佳^[8]。

2 实证研究

2.1 数据采集

本文选取深圳证券交易所的万科A股票收盘价作为研究对象,研究期间为2019年1月15日到2020年2月10日共254个样本数据。本文用R语言做统计分析,将前242个数据进行建模,最后12个数据用于预测对比,检验模型预测的准确性。数据来源为雅虎财经网站。

2.2 数据预处理

将原始序列记为X,做其序列图,可以看出,在研究期间万科股票收盘价在震荡中有一个上升的趋势,导致该序列不平稳。本文通过差分对其进行平稳化处理。用统计软

件做出一阶差分后的时序图(如图1所示),可以从图中看出新的序列在零的周围波动,基本平稳,偶尔会出现一些极大和极小值。从而初步判断一阶差分后的序列是平稳时间序列,要对万科公司股票收盘价建立 $d=1$ 的ARIMA模型,再基于该模型预测将来万科公司股票价格。

通过上述操作,即利用一阶差分后的序列图,仅能够对序列的平稳性作出初步判断。所以接下来需要进行单位根检验,深入判断序列的平稳性。若存在单位根,则无法通过检验,不能进一步建立模型。对本文的时间序列进行ADF平稳性检验。若时间序列是非平稳的,可以用差分法消除单位根,得到平稳序列。

在建模之前有必要通过单位根检验来验证时间序列的平稳性,若不存在单位根则时间序列是平稳的。在此我们所用检验方法为R语言tseries包中的adf.test函数。如表1所示,ADF检验结果表示原始时间序列是不稳定的。一阶差分后的万科公司股票收益率数据是稳定的,因此需要拟合一阶差分模型。至此,上述内容是我们对文本数据进行的基本分析,并且已经对数据进行了平稳化处理。接下来是本文模型构建与具体分析过程。

表1 ADF平稳性检验结果

变量	ADF 值	P 值	结论
close	-2.8283	0.2274	不稳定
dclose	-5.5917	0.01	稳定

2.3 模型定阶

根据上述分析,利用一阶差分可以将非平稳的单序列数据处理成平稳的时间序列,所以我们选择ARIMA(p, d, q)模型对数据进行建模分析。首先需要对模型进行定阶(如图2所示)。为了确定模型的阶数,接下来需要估计ARIMA模型里的 p, d 和 q 三个参数。从前面可以得到选择一阶差分的数据,则 $d=1$ 。对于图中的万科公司股票收盘价一阶差分序列dclose,在滞后1阶、3阶、6阶时有比较大的自相关系数,并且在滞后1阶、3阶、6阶时,其偏自相关系数也比较显著。

AIC准则可以用来确定模型的阶数,AIC值越大,表示模型拟合度越低。我们分别计算ARIMA(1,1,1)、ARIMA(3,1,1)、ARIMA(6,1,1)、ARIMA(1,1,3)、ARIMA(3,1,3)、ARIMA(6,1,3)、ARIMA(1,1,6)、ARIMA(3,1,6),以及ARIMA(6,1,6)模型的AIC值,结果分别为2408.47、2409.44、2410.63、2409.3、2407.42、2408.79、2411.73、2409.17、2412.98。

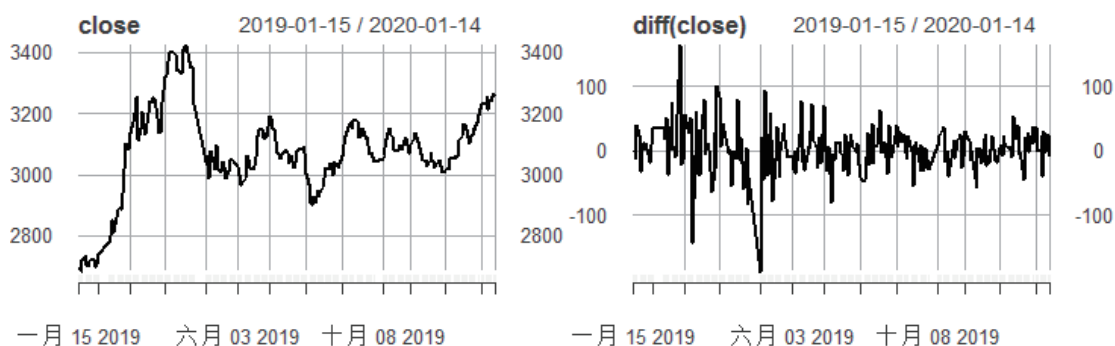


图1 万科公司股票收盘价及其一阶差分图

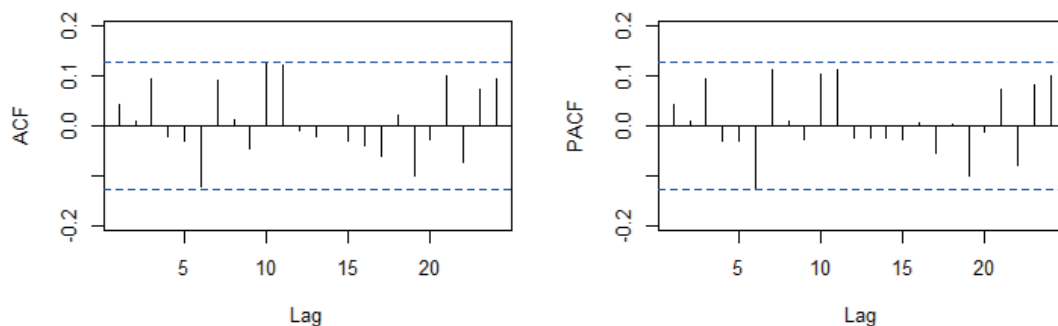


图2 确定模型的阶数

由此可知,AIC值最小的模型为ARIMA(3,1,3)。

2.4 模型参数估计

在此,我们利用拟合度最优的ARIMA(3,1,3)模型对数据进行拟合,然后估计模型参数,并且通过计算t比,发现所有的系数均显著,结果如表2所示。

表2 ARIMA(3,1,3)模型的参数估计结果表

	ar1	ar2	ar3	ma1	ma2	ma3
估计值	-0.7015	-0.9025	-0.7150	0.730	0.9133	0.8339
标准差	0.1788	0.0368	0.1731	0.151	0.0284	0.1507
t比	-3.9227	-24.5076	-4.1314	4.8354	32.1629	32.1629

根据表2的参数估计结果,最后得到如下股票价格的ARIMA(3,1,3)模型:

$$X_t = 0.7015X_{t-1} + 0.9025X_{t-2} + 0.715X_{t-3} + \varepsilon_t + 0.73\varepsilon_{t-1} + 0.0133\varepsilon_{t-2} + 0.8339\varepsilon_{t-3}$$

2.5 残差检验

模型是否可行,还需要对模型残差部分进行分析,检验模型残差部分的相关性是否被充分提取,在R语言中使用函数tsdiag来检验。结果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),模型的残差部分不存在相关性,说明模型拟合得很好。Ljung-Box检验的p值都在0.05之上,检验通过,可用于万科公司股票收盘价的预测。

2.6 结果预测

选定ARIMA(3,1,3)后,用R语言中的predict函数可以进行预测。在此,我们选择万科公司股票12天的实际收盘价与预测的收盘价进行比较,来检测模型预测的准确性和有效性。比较结果从2020年1月15日到2020年2月6日的12天里,其股票收盘价实际值分别为3237.81、3221.1、

3222.53、3243.82、3198.08、3207.12、3118.92、2878.14、2916.57、2953.01、3003.81和3013.74,预测值分别为3262.145、3268.785、3255.143、3253.904、3262.338、3267.295、3257.092、3253.745、3261.757、3266.452、3258.320和3254.058。

从以上数据中可以看到,本文拟合的ARIMA(3,1,3)模型能够较好地预测万科A短期的收盘价格,相对误差较小。

3 结语

本文以万科企业股份有限公司的股票收盘价为研究对象,通过实证研究,利用ARIMA模型对万科公司的收盘价序列进行了拟合,并进行了误差分析。通过ADF平稳性检验得出原始时间序列不平稳,进行一阶差分后的时间序列是平稳的,可以进一步用R语言拟合ARIMA时间序列模型,通过一阶差分平稳时序的ACF和PACF图确定模型可能的阶数,并利用AIC准则选取了合适的ARIMA模型,并且该模型通过了残差检验,说明模型的残差部分的相关性已经被充分提取,可以用于万科公司股票收盘价的预测。

从实证结果来看,在2020年1月15日到2020年2月10日的日期预测中,相对误差最高为13.05%,最小的误差为0.75%,且预测结果都要比实际值大。但整体来看,ARIMA模型能够较好地拟合和预测股票的收盘价格,可以应用于房地产业股票预测。另外,对于一月份数据的预测结果相对误差总体比二月份的小,这可能是由于受新冠疫情影响,全国房地产行业的开发及销售规模有所下降导致,而本文拟合的ARIMA模型对该信息的吸收不充分,对于万科公司的股票价格依然是不断上升的预测结果,这一点经济复苏以后该模型的预测效果可能会更好一些。

参考文献

- [1] J.H.Stock.Timeseries:Economic Forecasting,1996:15721-15724.
- [2] Liu W,Morley B.Volatility forecasting in the Hang Seng Index using the GARCH approach [J].Asia-Pacific Financial Markets,2009,16(01).
- [3] Cao Wen,Han Ying,Lam Sarah S.IIE Annual Conference.Proceedings,Short-Term Stock Price Trend Prediction of an Emerging Market Using Neural Networks,2013.
- [4] 孔华强.金融市场波动率模型及实证研究[D].北京:首都经济贸易大学,2006.
- [5] 江涛.基于GARCH与半参数法VaR模型的证券市场风险的度量和分析:来自中国上海股票市场的经验证据[J].金融研究,2010(06):103-111.
- [6] 李嘉松.基于ARIMA模型的沪深300指数预测及误差因素分析[J].赤峰学院学报,2017(33).
- [7] 王惠星,林嘉喜.基于ARIMA模型的上证50指数的分析及预测[J].时代金融,2017(24):143-144.
- [8] 张颖超,孙英隽.基于ARIMA模型的上证指数分析与预测的实证研究[J].经济研究导刊,2019(11):131-135.