

新股上市价格的影响因素研究分析

刘灯阁, 黄希芬*

云南师范大学数学学院, 云南 昆明

收稿日期: 2022年7月25日; 录用日期: 2022年8月5日; 发布日期: 2022年8月17日

摘要

新股的发行价格对市场有着重要意义, 研究新股的发行价格可以为新股定价提供依据。本文选取我国2018年A股上市的50只股票作为研究对象, 从定性分析和定量分析两方面对新股定价问题进行研究, 将定性分析的四大类影响因素量化后转化为七个定量因素, 并由此建立多元线性模型。模型表明新股发行价格可由发行市盈率、每股收益、发行前每股净资产来解释。最后用模型进行了样本内和样本外预测, 模型结果较好。

关键词

新股上市, 定性分析, 多元线性模型, 模型检验与预测

Research and Analysis on the Influencing Factors of the Listing Price of New Shares

Dengge Liu, Xifen Huang*

School of Mathematics, Yunnan Normal University, Kunming Yunnan

Received: Jul. 25th, 2022; accepted: Aug. 5th, 2022; published: Aug. 17th, 2022

Abstract

The issue price of new shares is of great significance to the market. Studying the issue price of new shares can provide a basis for the pricing of new shares. This paper selects 50 stocks listed in China A-sharemarket in 2018 as the research object, we study the issue of the price of new shares from qualitative analysis and quantitative analysis, then quantify the four categories of influencing factors in qualitative analysis and convert them into seven quantitative factors, and then establish a multivariate linear model. The model shows that the issue price of new shares can be explained by the issue price-earnings ratio, earnings per share, and net assets per share before the

*通讯作者。

issue. Finally, the model is used for in-sample and out-of-sample prediction, and the model results are better.

Keywords

New Stock Listing, Qualitative Analysis, Multiple Linear Models, Model Testing and Prediction

Copyright © 2022 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

资金盈余单位和资金紧缺单位之间, 通过直接融资迅速集中大量资金, 使得资金盈余单位的盈余资金得到增值, 资金紧缺单位筹集到资金, 即股份公司通过面向社会发行股票的过程及意义, 它为资本市场的有效运行提供了坚实的基础, 在此过程中新股的定价问题尤为重要。对发行人来说, 合理的价格决定企业是否能够达到融资目的, 以保证企业筹集到的资金能够满足生产所需; 对投资人来说, 制定市场能够接受的价格, 可以有效降低风险, 保证投资者利益。

张效梅[1]对中国股市的现状做了分析, 提出了保持我国经济的增长速度不下滑、少下滑等有效措施的建议。陶冶[2]阐述了新股定价过高或过低都将直接影响股票市场资源配置的有效性理论, 从而揭示了合理的价格对市场的重要性。Ibboston [3]通过实证研究发现, 发达的股票市场上供应是充足的, 以及新股低价发行的合理性。张炜[4]运用实证研究的方法探讨我国新股定价机制的合理性, 讨论了我国定价机制发展中出现的问题, 并提出了部分建议以供参考。成惠芝[5]认为新股上市首日价格的可量化影响因素有发行价、换手率、发行市盈率等; 不可量化的因素有市场因素、投资者因素等。胡志强[6]探讨了不同影响因素与 IPO 价格表现之间的关系并通过系列的实证分析得出结论。童艳和刘煜辉[7]通过理论分析和实证检验, 以 1996~2007 年在沪深上市的 1182 只新股为样本, 对中国资本市场 IPO 高抑价的制度基础和具体的形成原因进行了系统的分析。孔玉生和王忠杰[8]以 2000 年 1 月至 2002 年 6 月在上海证券交易所上市的 184 家新股(A 股)为样本, 建立了一个新股上市首日定价的预测模型, 以期对投资者的投资决策行为提供有益的帮助。杨文涛和陈杰[9]从整体的层面出发对利率水平和市场整体平均市盈率做了实证分析。

张效梅和陶冶对新股定价的问题进行定性分析, 并没有给出适当的模型; 张炜、成惠芝等人, 对新股定价问题进行了定性分析后, 选择了相应的影响因素建立多元线性模型, 但选取的因变量不够全面, 对数据的描述也不够详细; 孔玉生和王忠杰使用的数据没有随机或是平均地选取其他交易所的股票, 只能反映上证上市股票的情况, 在参考时不能获得更多有利的信息。综上, 我们对选取的研究对象的具体数据进行了搜集、筛选和分析。在现有的研究基础上, 先对问题进行定性分析, 用理论知识结合相应的数据, 分析影响新股定价的不可量化因素, 为之后的定量分析提供基础; 再对问题进行定量分析, 用 EViews 对问题进行定量分析, 以建立一个合理的多元线性模型, 用统计学方法解析各个因变量对自变量的影响程度并得出结论。

本文通过对发行新股公司所属行业、财务状况、市场因素和投资者因素进行分析并得出初步结论。之后再对问题进行定量分析: 首先建立多元线性模型; 再对模型进行检验; 最后对模型进行预测, 并对预测效果做了检验, 模型预测效果可靠。

2. 定性分析

2.1. 定性变量分析

我们选择企业所属行业、每股收益、市场利率、投资者因素四个因素[10]进行分析,从而为定量分析提供新思路。企业所属的行业与其股票价格的变化是息息相关的,企业所属行业的发展前景会影响投资者对公司发展前景的预期。每股收益是投资者或其他信息使用者据以评价企业的盈利能力、成长潜力等进而做出相关经济决策的重要的财务指标之一。每股收益越高,理论上股票价格应该是越高的。市场利率是指由资金市场上供求关系决定的利率,市场利率下降,企业的融资成本降低,使得净利润上升,最终企业的市值上涨;市场利率下降,企业投资于固定收益的资金减少,投资于股市的资金增加,使得市场的总体成交量上升。投资者因素,它对股市交易的影响是不可忽视的。为了研究这一影响,我们将其与市盈率结合起来,我们就市盈率来研究投资者因素对新股价格的影响,在股票相对估值方法中,价格倍数会随着投资者情绪的变化而变化。如果投资者保持乐观态度,收益乘数就较高,市场价格也较高。通过对上述不可量化影响因素的分析,初步确定了模型研究中需要用到的解释变量,我们再对变量进行定义并进行定量分析。

2.2. 数据的相关描述

在 2018 年度中,总共发行了 102 只股票(数据来自东方财富网数据中心——新股申购与中签查询),由于有些企业的过往业绩数据无法得到,我们剔除了数据搜集不完整的一些股票后,选取数据完整的 50 只股票作为此次的研究对象。选取的 50 只股票中,涉及到的行业十分广泛,我们将其划分为九个类别进行归类,统计占比如表 1 所示。

Table 1. Industry share
表 1. 行业占比

行业	比重
计算机、通信和其他电子设备制造业	12%
专用设备制造业	10%
汽车制造业	10%
资本市场服务	6%
软件和信息技术服务业	4%
电气机械和器材制造业	4%
化学原料及化学制品制造业	4%
医药制造业	4%
其他行业	46%

此外,选取的 50 只股票发行时间为 2018 年 1 月 17 日至 2018 年 12 月 27 日,且均来自 A 股(即人民币普通股股票)。表 2 给出它们的发股类型和上市版块。

Table 2. Types of issuance and listing sections
表 2. 发股类型和上市版块

发股类型	比重	上市版块	比重
主板	58%	上证交易所	58%
中小板	18%	深证交易所	42%
创业板	24%		

在确定了数据样本之后, 我们选取如下的影响因素进行定量分析: Y 为新股发行价格, 单位为元; X 为股票的总发行数量, 单位为股; X_1 为对股票的总发行数量取对数后的值, 由于总发行股数数值较大, 为了便于计算, 我们对 X 取对数, 即 $X_1 = LN\left(\frac{X}{1000}\right)$; X_2 为发行市盈率(%), 即股票价格/每股收益; X_3 为市场利率(%), 此处使用的是 SHIBOR, 对照每只新股的发行时间选择出的利率; X_4 为行业前景, 为虚拟变量, 1 表示占比大的前三个行业, 0 表示其他行业; X_5 为每股收益(选择上年度 12 月的数据为样本), 即税后利润与总股数的比值, 反映企业的经营成果; X_6 为股票发行前每股净资产(选择上年度 12 月的数据为样本), 企业完成改制后评估出的资本总额和总股数的比值, 反映每一股所拥有的资产现值; X_7 为每股现金流, 即(营业带来的净现金流量 - 优先股股利)/流通在外的普通股股数, 是评估每股收益“含金量”的重要指标。

3. 建模分析

3.1. 模型建立

将每股发行价格 Y 作为因变量, 对定义的总发行数量对数化后的数值 X_1 、发行市盈率 X_2 、市场利率 X_3 、行业前景 X_4 、每股收益 X_5 、发行前每股净资产 X_6 和每股现金流 X_7 作为自变量进行建模, 对 2018 年新股的定价问题建模分析, 从而得出影响新股价格的因素。

首先我们通过 EViews 绘制出的散点图来确定因变量与各个自变量之间的线性关系。在 EViews 中导入数据后绘制出现因变量和各个自变量的散点图如下所示。

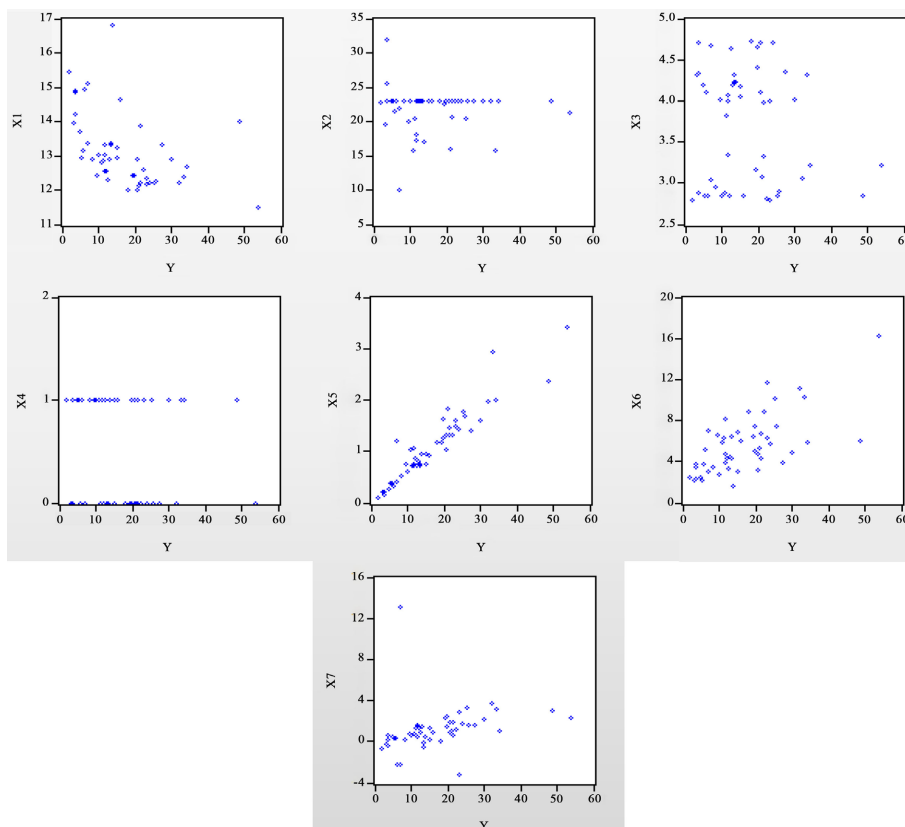


Figure 1. Scatter plot of IPO price and its influencing factors

图 1. 新股发行价格及其影响因素的散点图

从图 1 中可看出, X_4 和 X_5 与 Y 之间存在十分显著的线性关系; X_1 、 X_5 、 X_6 、 X_7 与 Y 之间存在较明显的线性关系; X_2 和 X_3 与 Y 之间没有明显的线性关系。需要行一步的研究, 来判断新股上市价格与各个影响因素之间是否存在关系, 为了对其进行进一步判断, 更加直观地得出结论, 我们可以将模型设定为:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 \quad (1)$$

其中, β_0 为常数项, β_1, \dots, β_7 为变量的系数, 以便于对模型进行参数估计和检验。

3.2. 模型检验

在初步确定模型的形式后, 通过最小二乘回归(WLS)估计模型的参数, 并对模型进行检验。利用 WLS 对模型(1)进行回归分析(表 3), 模拟多元回归方程。

Table 3. Simulate regression equation output results
表 3. 模拟回归方程输出结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-22.6975	8.6593	-2.6212	0.0121
X_1	0.4999	0.4773	1.0474	0.3009
X_2	0.7689	0.1582	4.8612	0.0000
X_3	0.0026	0.6235	0.0040	0.9968
X_4	-0.5248	0.8823	-0.5948	0.5552
X_5	18.0949	1.0234	17.6806	0.0000
X_6	-0.5210	0.2316	-2.2500	0.0297
X_7	-0.2133	0.2421	-0.8811	0.3833
R-squared	0.9420	Mean dependent var		16.9814
Adjusted R-squared	0.9324	S.D. dependent var		11.0344
S.E. of regression	2.8696	Akaike info criterion		5.0918
Sum squared resid	345.8446	Schwarz criterion		5.3978
Log likelihood	-119.2961	Hannan-Quinn criter		5.2083
F-statistic	97.5059	Durbin-Watson stat		1.9851
Prob (F-statistic)	0.0000			

此时 $R^2 = 0.9420$, $\bar{R}^2 = 0.9323$, 可决系数较高, F 值为 97.5059, 较为显著, 但是解释变量 X_1 、 X_3 、 X_4 和 X_7 的 t 检验没有通过, 这表明可能存在严重的多重共线性。根据可用的检验多重共线性的方法, 我们选择通过相关系数矩阵的计算确定是否存在多重共线性。

3.3. 多重共线性

多重共线性会由于模型的解释变量之间存在高度的相关关系而使建立出的模型估计量与经济意义不相符, 而后进行的显著性检验会失去意义; 更有可能导致我们的预测失去作用, 由于多重共线性使得方差变大, 会使预测区间变大。所以为了模型的有效性, 我们要对模型进行多重共线性的检验。

一般来说, 相关系数矩阵中相关系数大于 0.7 左右我们就认为各变量之间存在多重共线性。根据表

4 中的数据: X_5 和 X_6 之间的相关系数较大, 证明确实存在多重共线性。接下来我们采用逐步回归法(向前回归和向后回归), 去解决各变量之间存在的多重共线性。

可利用逐步回归的方法消除多重共线性, 选择向前法回归[Forward], 回归结果如表 5 所示:

Table 4. Explanatory variable correlation coefficient matrix

表 4. 解释变量相关系数矩阵

		Correlation					
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1.000	-0.0163	0.0244	0.0926	-0.5063	-0.4999	-0.0013
X_2	-0.0163	1.000	-0.1674	-0.0827	-0.2496	-0.1713	-0.5069
X_3	0.0244	-0.1674	1.000	-0.2109	-0.0911	-0.1012	0.1344
X_4	0.0926	-0.0827	-0.2109	1.000	-0.0136	-0.1223	-0.1654
X_5	-0.5063	-0.2496	-0.0911	-0.0136	1.000	0.7644	0.3879
X_6	-0.4999	-0.1713	-0.1012	-0.1223	0.7644	1.000	0.2700
X_7	-0.0013	-0.5069	0.1344	-0.1654	0.3879	0.2700	1.000

Table 5. Forward regression method regression results

表 5. 向前回归法回归结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X_5	16.1018	0.4869	33.0685	0.000
X_7	-0.6490	0.2543	-2.5523	0.0139
R-squared	0.8959	Mean dependent var		16.9814
Adjusted R-squared	0.8938	S.D. dependent var		11.0344
S.E. of regression	3.5967	Akaike info criterion		5.4371
Sum squared resid	620.9668	Schwarz criterion		5.5136
Log likelihood	-133.9283	Hannan-Quinn criter		5.4663
Durbin-Watson stat	1.9732			

此时 $\alpha = 0.05$, 模型为 $Y = 16.108X_5 - 0.6490X_7$ 。

根据回归结果可以看出, $R^2 = 0.8959, \bar{R}^2 = 0.8938$, 可决系数和修正后的值较大, $t = (33.0685), (-2.5523)$, 即 t 检验通过, 此时的拟合效果较好。回归结果中只有 X_5 和 X_7 留下了, 向前法下的逐步回归剩下的解释变量较少。我们需要综合考虑模型的拟合效果, 探究是否有其他更好的拟合模型, 所以接下来采用向后法回归[Backward], 回归结果如表 6 所示:

Table 6. Backward regression method regression results

表 6. 向后回归法回归结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X_5	17.5904	0.9092	19.3132	0.0000
X_6	-0.5385	0.2171	-2.4803	0.0168

Continued

X_2	0.8165	0.1321	6.1807	0.0000
C	-17.0056	3.1556	-5.3890	0.0000
R-squared	0.9397	Mean dependent var		16.9814
Adjusted R-squared	0.9358	S.D. dependent var		11.0344
S.E. of regression	2.7963	Akaike info criterion		4.9711
Sum squared resid	359.6778	Schwarz criterion		5.1240
Log likelihood	-120.2766	Hannan-Quinn criter		5.0293
F-statistic	239.0084	Durbin-Watson stat		1.9149
Prob (F-statistic)	0.0000			

此时 $\alpha = 0.05$ ，模型为 $Y = -17.0056 + 0.8165X_2 + 17.5604X_5 - 0.5385X_6$ 。

根据回归结果可以看出， $R^2 = 0.9397, \bar{R}^2 = 0.9358$ ，可决系数和修正后的值较大， $t = (-5.3890), (6.1807), (6.1807), (19.3132), (-2.4803)$ ，即 t 检验通过，又 $F = 239.0084$ 远远大于临界值；综合上述，此时的拟合效果较好，相较向前回归的效果，此时回归后留下的解释变量较多，所以我们选择向后法回归的结果来进行建模。

消除了多重共线性之后，模型的拟合结果还没有很好，为了保证模型参数的估计量具有良好的统计性质，我们还应该对模型进行异方差检验。异方差产生的主要有原因：1) 来源于截面数据；2) 来源于测量误差、模型中被省略的一些因素对被解释变量的影响；3) 有时产生于计量经济模型所研究问题的本身等。异方差性产生的主要原因较为符合我们所研究的模型，即模型可能存在异方差性，需要做进一步检验。

3.4. 异方差性

在线性回归模型的研究中，对异方差的检验和修正是我们常常遇到的问题，它的存在对回归分析有极强的破坏作用，所以检测模型的异方差性是很必要的：通过绘制残差平方序列的散点图来确定模型的异方差性。

首先进行异方差的图性检验：绘制残差平方序列 E_2 与解释变量 X_2 、 X_5 和 X_6 的散点图。

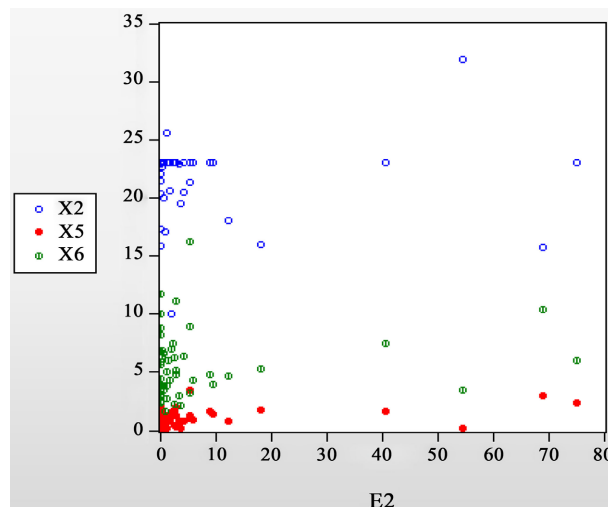


Figure 2. Scatter plot of residual squared series
图 2. 残差平方序列的散点图

残差平方序列 E_2 与解释变量 X_2 、 X_5 和 X_6 的散点图主要分布在左下角, 大致看出有变动的趋势, 因此, 模型在很大程度上可能存在异方差。为了确定由图 2 得出的结论, 我们进行进一步的检验——White 检验, 除了图形检验之外这是最常用的检验异方差性的方法。下面我们对模型进行 White 检验, 由此确定模型存在异方差。

对模型进行 White 检验, 无交叉乘积项的检验结果如表 7 所示:

Table 7. White test results

表 7. White 检验结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	42.7408	68.9797	0.6196	0.5390
X_5^2	28.6836	6.8367	4.1955	0.0001
$X_5 * X_6$	-6.8208	3.0145	-2.2626	0.0292
$X_5 * X_2$	-1.1750	1.1410	-1.0298	0.3093
X_5	8.3427	29.2861	0.2849	0.7772
X_6^2	0.0982	0.4634	0.2118	0.8333
$X_6 * X_2$	0.1565	0.3930	0.3982	0.6926
X_6	2.7742	7.6312	0.3635	0.7181
X_2^2	0.1953	0.0855	2.2842	0.0277
X_2	-6.7063	4.8997	-1.3687	0.1787
R-squared	0.6947	Mean dependent var		7.1936
Adjusted R-squared	0.6261	S.D. dependent var		16.5281
S.E. of regression	10.1069	Akaike info criterion		7.6412
Sum squared resid	4085.9430	Schwarz criterion		8.0236
Log likelihood	-181.0290	Hannan-Quinn criter		7.7868
F-statistic	10.1158	Durbin-Watson stat		1.6708
Prob (F-statistic)	0.0000			

取显著水平 $\alpha = 0.05$, 由于 $X_{0.05}^2(9) = 16.92 < nR^2 = 34.7378$, 所以拒绝原假设, 即模型存在异方差。

调整异方差——利用加权最小二乘法(WLS)调整, 我们选择权重为“ $1/abs(E1)$ ”, 再对模型进行 White 检验, 得到有交叉乘积项的检验结果如表 8 和表 9 所示:

Table 8. WLS test results

表 8. WLS 检验结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-16.8927	1.2877	-13.1186	0.0000
X_2	0.8079	0.0550	14.6778	0.0000
X_5	17.5473	0.4106	42.7354	0.0000
X_6	-0.5201	0.0823	-6.3182	0.0000

Continued

Weighted Statistics			
R-squared	0.9896	Mean dependent var	13.1621
Adjusted R-squared	0.9889	S.D. dependent var	14.4245
S.E. of regression	0.9436	Akaike info criterion	2.7984
Sum squared resid	40.9603	Schwarz criterion	2.9514
Log likelihood	-65.9615	Hannan-Quinn criter	2.8567
F-statistic	1455.9610	Durbin-Watson stat	1.9106
Prob (F-statistic)	0.0000	Weighted mean dep.	17.2784

Table 9. White test results after WLS regression

表 9. WLS 回归后的 white 检验结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.1523	0.1633	7.0581	0.0000
$X_2^2 * WGT^2$	0.0155	0.0235	0.6599	0.5132
$X_2 * X_5 * WGT^2$	-0.2125	0.3430	-0.6196	0.5391
$X_2 * X_6 * WGT^2$	-0.0095	0.0391	-0.2440	0.8085
$X_2 * WGT^2$	-0.3471	0.9035	-0.3842	0.7029
$X_5^2 * WGT^2$	0.5610	1.0057	0.5577	0.5802
$X_5 * X_6 * WGT^2$	-0.1320	0.4304	-0.3067	0.7607
$X_5 * WGT^2$	4.7183	8.4056	0.5613	0.5778
$X_6^2 * WGT^2$	0.0045	0.0498	0.0896	0.9291
$X_6 * WGT^2$	0.2779	0.7837	0.3546	0.7247
WGT^2	-0.8451	9.8784	-0.0855	0.9323
R-squared	0.2829	Mean dependent var		0.8192
Adjusted R-squared	0.0991	S.D. dependent var		0.9291
S.E. of regression	0.8819	Akaike info criterion		2.7780
Sum squared resid	30.3303	Schwarz criterion		3.1986
Log likelihood	-58.4500	Hannan-Quinn criter		2.9382
F-statistic	1.5390	Durbin-Watson stat		1.5087
Prob (F-statistic)	0.1625			

取显著水平 $\alpha = 0.05$, 由于 $X_{0.05}^2(10) = 18.31 < nR^2 = 14.1484$, 模型已经不存在异方差, 可见对模型进行了 WLS 调整(权数为 $1/abs(E1)$)后, 已消除异方差。为使调整的结果更合理, 我们再选择一个权数($1/E2$)对模型进行调整, 调整后的结果以及检验结果如表 10 和表 11。

取显著水平 $\alpha = 0.05$, 由于 $X_{0.05}^2(10) = 18.31 < nR^2 = 93.6512$, 模型依然存在异方差, 可见对模型进行了 WLS 调整(权数为 $1/E2$)后没有消除异方差。所以我们选择表 8 和表 9 的结果作为最后的模型:

$$Y = -16.8927 + 0.8079X_2 + 17.5470X_5 - 0.5201X_6 \quad (2)$$

Table 10. Adjusted WLS test results
表 10. 调整后 WLS 检验结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-16.4156	0.4240	-38.7160	0.0000
X_2	0.7905	0.0184	42.9247	0.0000
X_5	17.7337	0.1426	124.3609	0.0000
X_6	-0.5766	0.0271	-21.2558	0.0000
Weighted Statistics				
R-squared	0.9991	Mean dependent var		8.4086
Adjusted R-squared	0.9990	S.D. dependent var		21.2141
S.E. of regression	0.2259	Akaike info criterion		-0.0607
Sum squared resid	2.3476	Schwarz criterion		0.0922
Log likelihood	5.5182	Hannan-Quinn criter		-0.0024
F-statistic	16443.29	Durbin-Watson stat		1.5882
Prob (F-statistic)	0.0000	Weighted mean dep.		21.4789

Table 11. White test results after adjusted WLS regression
表 11. 调整后 WLS 回归后的 white 检验结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0507	0.0014	35.9036	0.0000
$X_2^2 * WGT^2$	-0.0014	0.0021	-0.6812	0.4998
$X_2 * X_5 * WGT^2$	-0.0358	0.0201	-1.7756	0.0836
$X_2 * X_6 * WGT^2$	0.0042	0.0020	2.1135	0.0410
$X_2 * WGT^2$	0.0782	0.0745	1.0499	0.3002
$X_5^2 * WGT^2$	-0.0323	0.0467	-0.6927	0.4926
$X_5 * X_6 * WGT^2$	0.0092	0.0252	0.3640	0.7178
$X_5 * WGT^2$	0.7115	0.5019	1.4175	0.1643
$X_6^2 * WGT^2$	-0.0070	0.0025	-2.7539	0.0089
$X_6 * WGT^2$	0.0079	0.0456	0.1747	0.8622
WGT^2	-1.2639	0.6730	-1.8781	0.0679
R-squared	0.7930	Mean dependent var		0.0469
Adjusted R-squared	0.7399	S.D. dependent var		0.0169
S.E. of regression	0.0086	Akaike info criterion		-6.4722
Sum squared resid	0.0029	Schwarz criterion		-6.0516
Log likelihood	172.8061	Hannan-Quinn criter		-6.3120
F-statistic	14.9427	Durbin-Watson stat		1.5444
Prob (F-statistic)	0.0000			

可以看出, $R^2 = 0.9896$, 修正的可决系数 $\bar{R}^2 = 0.9889$, 可决系数和修正后的值较大, $t = (-13.1186), (14.6778), (42.7354), (-6.3182)$, 即最终模型(2)拟合效果较好。

3.5. 假设检验

已知原假设为 $H_0: \beta_0 = \beta_1 = \dots, \beta_n = 0$, 备择假设为 $H_1: \beta_i, (i=0,1,\dots,n)$ 不全为 0。取显著水平 $\alpha = 0.05$, 查询 F 分布表可知自由度为 $k-1=3$ 和 $n-k=46$ 时的临界值 $F_\alpha(3,46) = 2.81$ 。由表 11 得到 $F = 14.9427$, 由于 $F > F_\alpha(3,46)$, 应拒绝原假设 H_0 , 说明回归方程显著, 即“发行市盈率”、“每股收益”、“发行前每股净资产”三个变量联合起来确实对“新股上市价格”有显著影响。

查 t 分布表可得, 自由度为 $n-k=46$ 时的临界值为: $t_{0.025}(n-k) = 2.0140$ 。由 $t = (-13.1186), (14.6778), (42.7354), (-6.3182)$ 可知, 其绝对值均大于临界值 2.0141, 即在给定的显著性水平 $\alpha = 0.05$ 下, 分别都应当拒绝原假设 H_0 , 即当在其他解释变量不变的情况下, 解释变量 X_2 、 X_5 和 X_6 分别对被解释变量新股定价 Y 有显著的影响。

4. 模型预测

为了说明最终模型的拟合效果, 我们根据模型对 Y 值进行预测, 观察实际值和预测值之间的误差, 用以证明模型是可靠的并且是否能够用在样本外的预测, 所以接下来我们将从样本内预测和样本外预测两个角度对模型进行预测并分析预测效果。

4.1. 样本内预测

此时仍然选择建模时使用的 50 只股票为对象进行预测(我们在预测时选取的是静态预测)。图 3 中, 左侧方框内的蓝线表示预测值 YF ; 红线表示置信区间(即估计区间), 通常是在单点预测的基础上加减置信水平的分位数乘以估计方差, 此时图中位于蓝线上方的红线表示预测值+2*S.E。位于蓝线下方的红线表示预测值-2*S.E。在图 3 上方的虚线方框中给出的是第 10~15 只股票的预测值与其置信区间的详图, 我们可以清晰地看到每一个位置的预测值都落在了置信区间之内, 预测效果较好。

图 3 中, 右侧方框内给出的是衡量预测准确度的统计值: Root Mean Squared Error: 表示均方误差根, 是估计值与真值之差平方的期望值, 用以衡量数据的变化程度, 此时为 2.6827; Mean Absolute Error: 表示平均绝对误差, 是将平均误差绝对值化, 反映实际预测误差的大小, 此时为 1.7867; Mean Abs. Percent Error: 表示平均相对误差, 对平均相对误差取绝对值, 它的计算公式乘以了 100%, 所以输出结果中其值的度量单位为百分比, 此时为 15.5411%。Theil Inequality Coefficient (Theil 不等系数), 只能取 0 到 1 之间的值, 其值为 0 说明预测十分准确, 其值为 1 说明预测值非常不准确, 此时为 0.0667, 此值接近于 0, 说明预测效果较好。

Theil 值被分为三部分: Bias: 表示预测值是否有系统性的偏差, 此时为 0.000031, 说明存在的系统性偏差很小; Variance: 此时为 0.013083, 说明实际值的波动与预测值的差别不大; covariance: 此时为 0.986886, 表示的是非系统的预测误差。上述三部分的值之和: $0.000031 + 0.013083 + 0.986886 = 1$ 。即 Theil 值的三部分相加等于 1。

综合上述, 模型的预测结果表明预测过程中没有出现错误, 预测值 Y 和实际数据 Y 值误差不大, 预测效果很好。为了进一步分析预测效果, 接下来计算预测值和实际值的误差, 它反映了预测值与实际值的偏离大小。将误差区间分为四个层次, 统计每个误差对应的单位, 计算其比重, 结果如表 12 所示。

由表 12, 绝对误差在 ± 2.00 以内的有 36 只, 占比为 72%, 绝对误差最高为 8.44, 最低为 -8.68。此结果说明我们最终的模型的拟合程度是比较高的, 预测的误差分布大部分在较小的范围之内, 只有几只的预测误差较大, 从统计学的角度上看, 产生这个程度的误差是正常的。

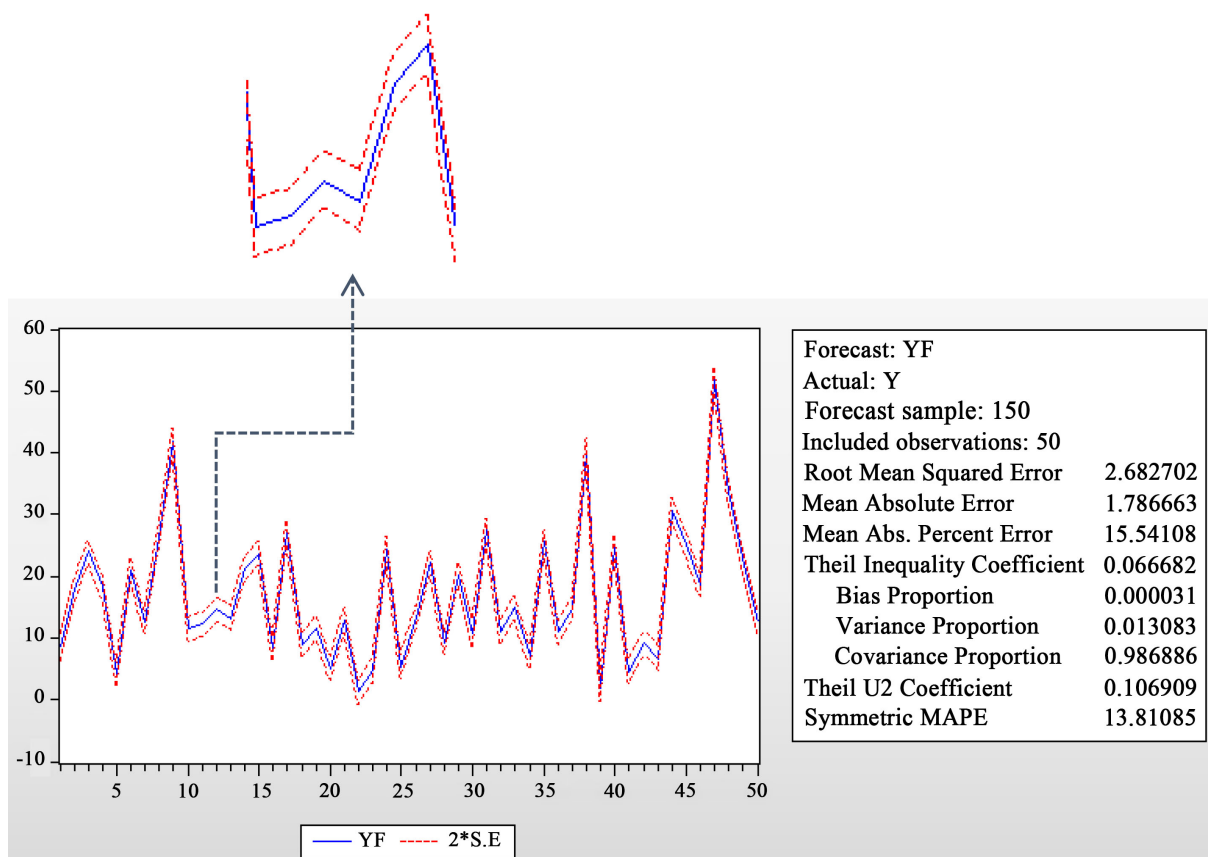


Figure 3. Sample prediction results

图 3. 样本预测结果

Table 12. Error distribution for in-sample forecasts

表 12. 样本内预测的误差分布

绝对误差	新股(只)	比重(%)
-2.00 到 2.00	36	72%
-4.00 到 4.00	9	18%
大于 4.00	4	8%
小于-4.00	1	2%

4.2. 样本外预测

在已知的 50 只股票范围外重新选取其他的新股作为对象进行预测, 目的是为了进一步证明预测结果的准确性。

我们选取 2019 年上市的 10 只新股作为预测对象进行预测, 这 10 个样本的上市时间为 2019 年 1 月 8 日至 2019 年 7 月 2 日, 分别为: 华林证券、威尔药业、威派格、永冠新材、迪普科技、博通集成、日丰股份、鸿合科技、泉峰汽车和天准科技。它们全来自 A 股, 发股类型和所属行业也与样本中股票相符。发股类型和所属行业的占比如图 4 所示。

根据预测结果, 对实际值 Y 与预测值 YF 进行比较, 计算出绝对误差, 如表 13。

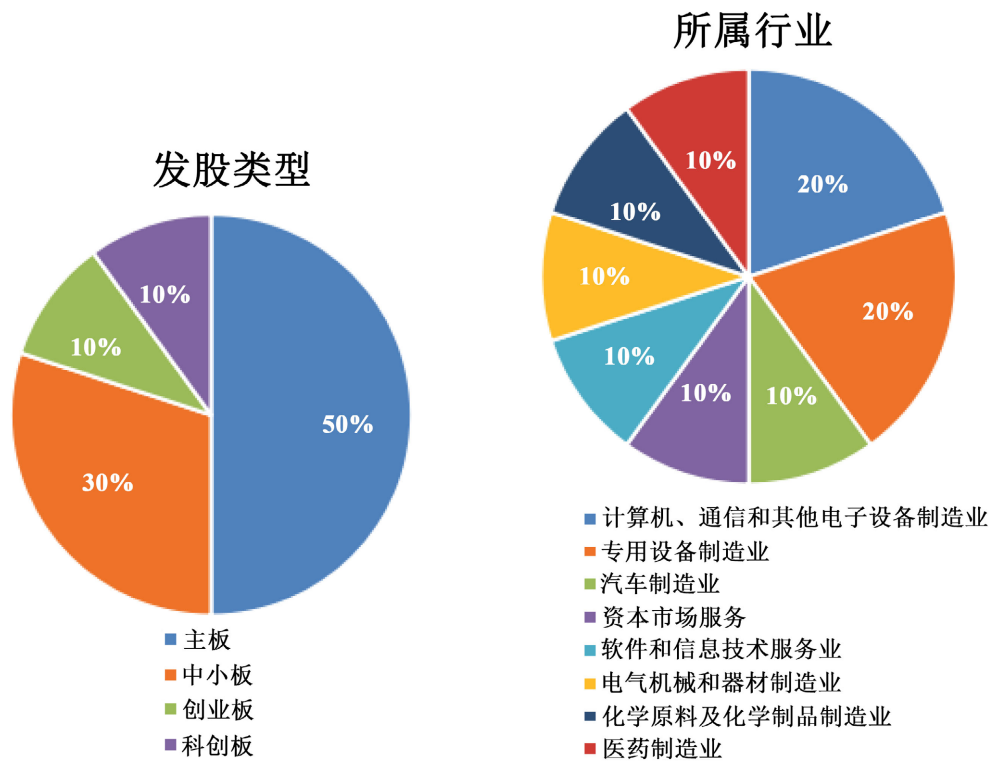


Figure 4. Type of issue and industry
图 4. 发股类型和所属行业

Table 13. Out of sample forecast
表 13. 样本外预测

实际值 Y	预测值 YF	绝对误差
3.62	3.29	-0.33
35.50	35.44	-0.06
5.7	5.90	0.20
10.00	16.47	6.47
11.23	12.47	1.24
18.63	12.32	-6.31
10.52	17.05	6.53
52.41	54.47	2.06
9.79	9.16	-0.63
25.5	39.63	14.13

由表 13, 样本外数据的实际值与预测值的绝对误差(除天准科技(最后一项)外)都在样本数据预测误差范围之内且总体的误差较小, 说明模型的预测效果是可靠的, 预测结果也合理, 这一结果对企业发行新股的定价具有重要意义。由于天准科技这一只新股不符合我们的预测, 所以我们现在来单独对它进行分析。天准科技是在科创板上市的, 科创板是独立于现有主板市场的新设板块, 天准科技的定价与预测价

的绝对误差为 14.13, 超出正常误差范围, 所以由主板、中小板、创业板上市的这些新股的定价模型并不能适用于科创板上市的股票的定价。

5. 结语

在模型的建立过程中, 我们消除了异方差以及多重共线性, 对模型的 F 值、 t 值和可决系数做了分析, 最后根据模型作出的预测, 将预测值与实际值进行了比对, 还做了外延数据的预测, 证明了建模效果较好。从模型的结果可知, 在本文选取的七个影响因素中, 新股上市价格在 98.9578% 以上的概率可以用“发行市盈率”、“每股收益”、“发行前每股净资产”来解释。公司想要从发行股票这一行为中募集到更多的资金, 就应该重点关注每股收益的变化; 市盈率的升高, 一定程度上会使股票的上市价格提高; 发行前每股净资产对其定价的影响虽然是呈现出负相关的, 但是其影响程度十分小。此模型对于投资者来说, 有参考作用, 虽然不能完全有把握预测, 但是对个人投资者来说是足够的, 可以根据自己的需求, 参考模型中的分析结果来评估某只新股的定价是否合理、是否值得投资、能否达到预期效果等。

参考文献

- [1] 张效梅. 中国股市现状分析[J]. 中国证券期货, 2012(11): 5.
- [2] 陶冶. 影响新股定价的因素分析[J]. 湖南大学学报, 2003, 17(3): 34-36.
- [3] Ibboston, R.G. (1975) Price Performance to Common Stock New Issues. *Journal of Financial Economics*, 2, 235-272. [https://doi.org/10.1016/0304-405X\(75\)90015-X](https://doi.org/10.1016/0304-405X(75)90015-X)
- [4] 张炜. 我国新股发行定价问题探讨[J]. 经济师, 2005(9): 123-125.
- [5] 成惠芝. 新股上市首日价格影响因素的研究[J]. 东南大学学报, 2011(13): 76-77.
- [6] 胡志强. 我国新股发行的价格行为、市场实际与 IPO 浪潮研究[M]. 武汉: 武汉大学出版社, 2011.
- [7] 童艳, 刘煜辉. 中国 IPO 定价效率与发行定价机制研究[M]. 北京: 中国金融出版社, 2010.
- [8] 孔玉生, 王忠杰. 新股上市首日定价预测模型研究[J]. 中国管理科学, 2003, 11(1): 6-9.
- [9] 杨文涛, 陈杰. 新股发行定价的影响因素分析[J]. 金融财税, 2011(8): 91-93.
- [10] 黄灿林. 我国农林牧渔业上市公司市盈率影响因素的实证研究[D]: [硕士学位论文]. 合肥: 安徽农业大学, 2011.