



FIN3080 Project 1 Report

120090651 马可轩 120090633 申恒瑜 119020025 李励嘉

内容概要

在本报告中，小组成员计算了2000年至今中国股票市场所有主板、中小板和创业板上市股票的静态PE/PB月中位数，以及按等权平均法与按流通市值加权平均法的累计月回报率，绘制了相应的图表，探究各指标在不同时期的变化，比较各指标在两板块间的异同，并对其原因进行讨论。同时，我们汇总了主板、中小板和创业板的月度市净率，对可能的影响因子使用回归分析，并做出解释。

本研究所有相关数据均来自与深圳希施玛数据科技有限公司（CSMAR），数据最终解释权归深圳希施玛数据科技有限公司所有。

名词解释

1. PE

市盈率（Price Earnings Ratio），指股票价格除以每股收益的比率，或以公司市值除以年度股东应占溢利。

$$PE Ratio = \frac{\text{Share Price}}{\text{EPS}}$$

本研究选取的数据均为静态市盈率，通常用来作为比较不同价格的股票是否被高估或者低估的指标。

2. PB

市净率（Price/Book value），指每股股价与每股净资产的比率。

$$PB Ratio = \frac{\text{Share Price}}{\text{Book Value Per Share}}$$

一般来说市净率较低的股票，投资价值较高，相反，则投资价值较低；但在判断投资价值时还要考虑当时的市场环境以及公司经营情况、盈利能力等因素。

3. ROE

净资产收益率（Return on Equity），是净利润与平均股东权益的百分比，是公司税

后利润除以净资产得到的百分比率。

$$ROE = \frac{EPS}{\text{Book Value Per Share}}$$

该指标反映股东权益的收益水平，用以衡量公司运用自有资本的效率。指标值越高，说明投资带来的收益越高，体现了自有资本获得净收益的能力。

4. DPR

股息支付率（Dividend Payout Ratio）是向股东分派的股息占公司盈利的百分比。股息支付率指标反映普通股股东从每股的全部净收益中分得多少。

$$DPR = 1 - \frac{EPS - DPS}{EPS}$$

5. Return

回报率为通过投资而应返回的价值，即从一项投资活动中得到的经济回报。对于有 n 支股票的市场，在时间 t 的市场回报率的计算公式为：

$$R_{n,t} = \frac{\sum_n w_{n,t} r_{n,t}}{\sum_n w_{n,t}}$$

使用等权平均法计算时， $w_{n,t} = 1$ ；使用流通市值加权平均法计算时， $w_{n,t}$ 表示股票 n 在 $t-1$ 时间周期的流通市值。若 $r_{n,t}$ 表示的是不考虑现金红利再投资的个股回报率，则 $R_{n,t}$ 为不考虑现金红利再投资的平均市场回报率；反之亦然。

由上，我们可以推导出累计回报率的计算公式为：

$$R_{cml} = \prod_t (1 + R_{n,t})$$

问题一

一、研究方法

（一）数据处理

1. 从CSMAR官方库批量下载2000-2022年主板、中小板与创业板的所有原始数据，用Python pandas库进行后续计算。
2. 利用PE/PB计算公式计算个股单日的静态PE/PB值，并用DataFrame进行保存，便于后续的PE/PB月中位数统计。
3. 板块月中位数获得方法：

得到个股 PE/PB 单日数据后，对所有个股单日 PE/PB 以月为单位求中位数，以代表了个股在每月份中最具代表性的 PE/PB 值。然后对所有股票的月度 PE/PB 再取中位数，得到主板&中小板与创业板各月份的最具代表性的 PE/PB 值，并将数据重新汇总。

（二）图像绘制

1. 绘图数据

图像绘制的数据来自于上述处理的最终结果，利用 `datetime` 函数获取时间序列作为 x 轴数据，将板块月度 PE/PB 纳为 y 轴数据。

2. 图像解释

总图像包含四条曲线，分别为主板&中小板和创业板的月度 PE 和 PB 中位数。图像可以通过代码注释自由选择，便于后续进行对比分析。

二、问题探究

（一）影响PE与PB的相关因素研究

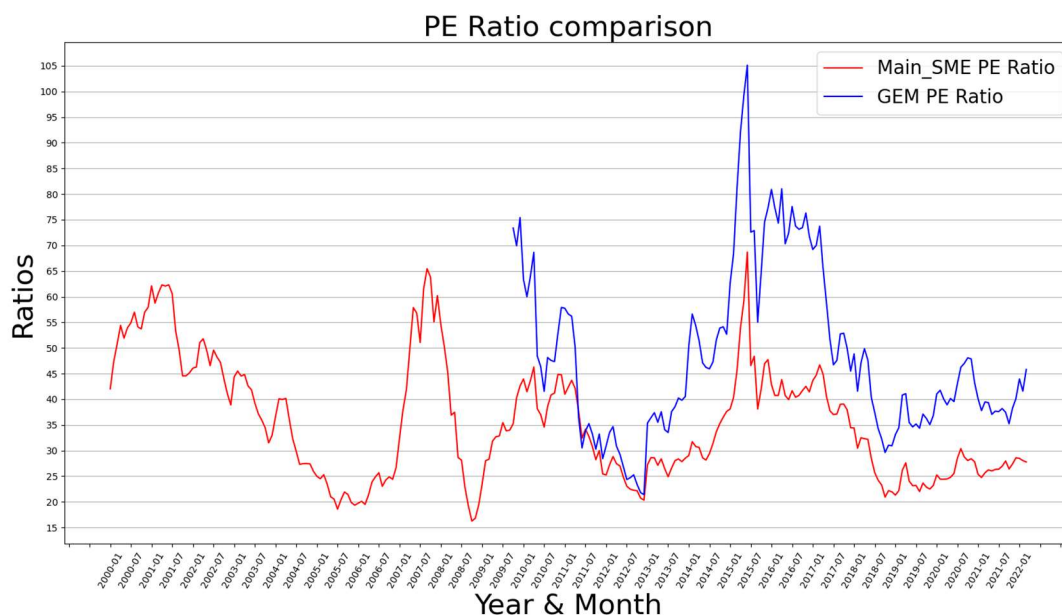


图1：主板&中小板与创业板PE 月中位数对比图

1. PE 的变化趋势概述

（1）主板&中小板 PE：介于 15 至 70 之间，总体波动较大，有三次高峰与四次低谷。

（2）创业板 PE：介于 20 至 105 之间，总体波动较大。其中最高峰值出现于 2015 年。而两次突出的低谷则分别出现于 2012 年和 2018 年。

（3）两板块 PE 变化趋势对比：二者数据在变化趋势上具有同步性，2010 年后的峰值及低谷出现的时间基本吻合。但在 2009 年末创业板成立初期，创业板的 PE 明显高于主板&中小板的同期 PE 值。

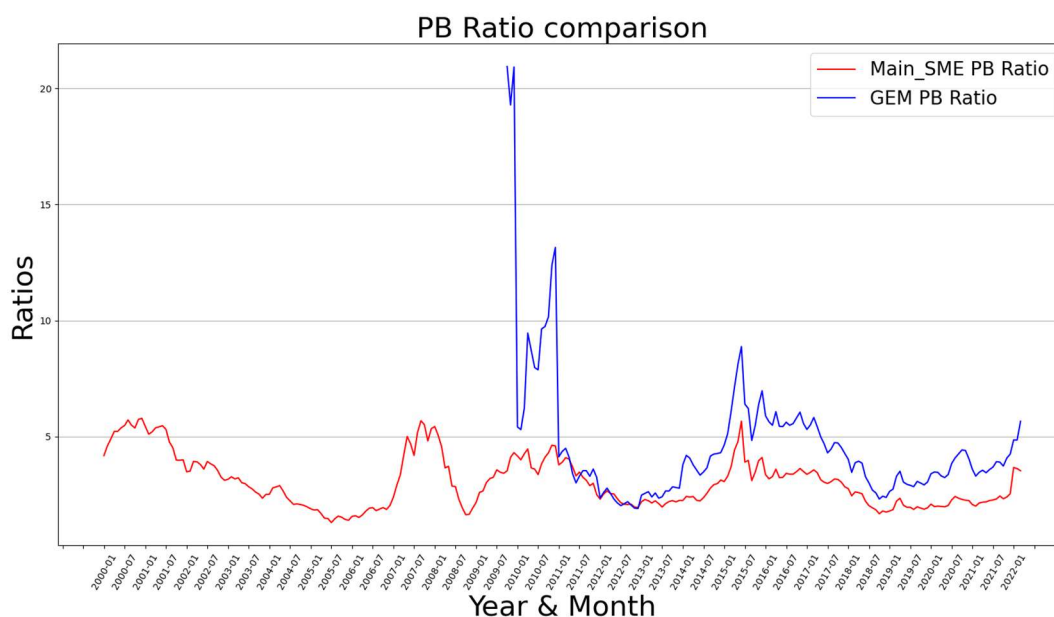


图2：主板&中小板与创业板PB数据对比图

2. PB 的变化趋势概述

(1) 主板&中小板 PB：大部分不超过 5，总体较平稳。其中三次峰值年份分别为 2001 年、2007 年及 2015 年。而较明显的低谷分出现于 2005 年及 2008 年。

(2) 创业板 PB：多介于 0 至 10 之间，除创业板块建立初期，PB 值较高，2011 年开始创业板 PB 值趋于平稳。其中峰值出现于 2009 年末。而两次低谷分别出现与 2012 年及 2018 年末。

(3) 两板块 PB 变化趋势对比：二者数据在变化趋势上具有同步性，2011 年后的峰值及低谷出现的时间基本吻合。但在 2009 年末至 2010 年末，创业板的 PB 明显高于主板&中小板的同期 PB 值。

从图二得出，相较于 PE，PB 的波动幅度较小（除创业板建立初期）。同时，在对比主板&中小板与创业板 PB 数据时发现板块的差别对 PB 具有一定影响。

因此，依据 PE 与 PB 的计算公式及上述图像分析，研究其影响因素要着手于股票的微观数据及中国的宏观经济两个方面。

3. 影响 PE 与 PB 的因素研究

(1) 影响 PE 的微观因素

PE 受 EPS 影响，通过 EPS 的计算公式总结得出影响每股收益的因素主要有：上市公司获利能力的高低及公司的分红政策。因此在股价相对平稳的情况下，一个企业利润越高，普通股的分红越多，则 PE 值越低，反之亦然。

(2) 影响 PB 的微观因素

由于每股账面价值取决于公司净资产，其中净资产的多少是由股份公司经营状况决定的，因此股份公司的经营业绩越好，其资产增值越快，股票净值就越高，因此股东所拥有的权益也越多。而越高的 Book Value Per Share 会导致 PB 越低。

(3) 分析创业板 PE/PB 值初期较高的原因

创业板的 PE 与 PB 值普遍高于主板&中小板。两个板块最为突出的差异出现在 2009 年末至 2010 年末。创业板中的公司在上市初期，企业净资产相对较低，因此在股价平稳的情况下，创业板的 PB 值相对较高。因创业公司的发展特点，净资产会在初期快速增长，后期增速放缓。通过研究创业板成立初期的股票特点分析得出，以下三点因素可以影响 PE/PB：

- a. 预期上市公司获利能力的高低分析
- b. 预测公司未来的成长性
- c. 投资者所获报酬率的稳定性

3. 结合时势分析 PE 与 PB 的变化

2007-2009 年环球金融危机导致全球股价大跌，直接导致 PE/PB 值大幅降低。2015 年中国牛市，中国股票市场股价大涨，使得 PE/PB 值大幅升高。

（二）主板&中小板与创业板 PE/PB 的区别与联系

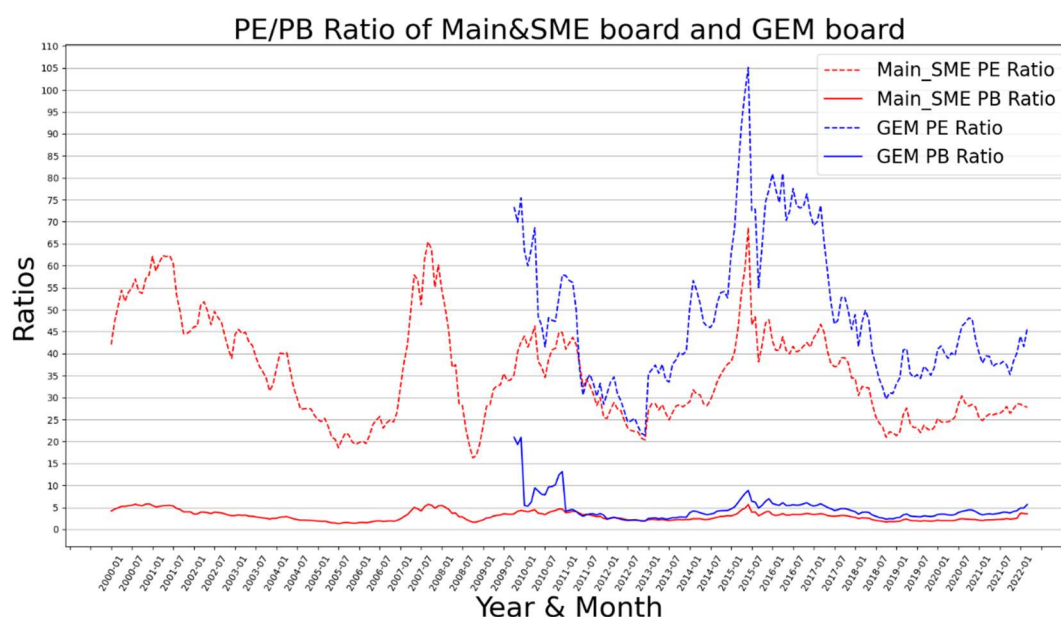


图3：主板&中小板与创业板 PE/PB 数据对比图

1. 两个板块 PE/PB 的不同点

（1）同受股市股价影响

PE 与 PB 的计算公式中，分子均为股票价格，因此二者在股市行情较好的时期都有较高的比值，在图像中体现于峰值的出现在时间上趋同。

（2）创业板 PE/PB 普遍高于主板&中小板

因创业板股票自身的特点，外加 PE/PB 计算公式的相似性，均使得，创业板 PE/PB 普遍高于主板&中小板。

2. 两个板块 PE/PB 的相同点

（1）PB 的数值低于 PE

因 PB 计算公式中 Book Value Per Share 的数值严格大于 PE 计算公式中的 EPS，因此 PB 的数值曲线一定位于 PE 数值曲线下方。

（2）PB 的曲线更为平缓

EPS 受到公司企业的分红政策影响，一般波动幅度较大，而 Book Value Per Share 取决于公司盈利能力，波动幅度远小于 EPS，因此 PB 的曲线更为平缓。

问题二

一、研究方法

（一）数据处理

1. 数据来源

本问题所使用数据选取了 2000-2022 年的主板、中小板和创业板上市股票的证券代码、交易日期、股票价格、每股收益、每股现金分红与每股净资产。

2. 数据处理

（1）本题使用 python 中的 pandas, statsmodel 等库对数据进行处理，因数据量较大，作者将多个表读取后连接，最终输出为 csv 文件 Reg_data.csv, 后使用该 csv 文件进行数据分析。具体处理步骤与注释详见 PB_sheet_calculate.py 文件。

（2）本题使用 VIF 多重共线性检验与指标间的相关系数检验，最后使用线性回归模型以及 t 分布进行数据有效性确认，代码部分详见 PB_Regression.py 或者 PB_Regression_jupyter.ipynb 文件。

（3）数据处理中遇到的 NAN 值，作者均对整行进行数据清除。

3. 数据归纳类型

（1）PB_sheet_calculate.py 文件中，作者将个股每月指标（除回报率）取中位数代表该个股当月表现，回报率指标来自 csmar 数据库，取每月不同市场回报率均值代表市场当月回报率表现。

（2）PB_Regression.py 文件中，作者将当月所有个股指标取中位数得到当月指标，对当月指标进行回归分析。

(二) 统计过程

1. 对 ROE, DPR, ret 三个数据首先进行 VIF 多重共线性检验, 结果如下:

```
In [7]: df2 = df1[['ROE', 'DPR', 'ret']]
vif_data = [variance_inflation_factor(df2.values, df2.columns.get_loc(i)) for i in df2.columns]
vif_data
```

Out[7]: [11.573347447664709, 11.629400862194418, 1.0196409964335038]

当 $VIF \geq 10$ 时, 我们认为变量具有多重共线性, 因此在进行回归分析时, DPR 和 ROE 需进行两次单变量回归分析。

2. 对 PB, ROE, DPR, ret 进行相关系数计算, 结果如下:

```
In [8]: df1.corr()
```

Out[8]:

	PB	ROE	DPR	ret
PB	1.000000	0.340137	-0.194651	0.016167
ROE	0.340137	1.000000	-0.520066	-0.130353
DPR	-0.194651	-0.520066	1.000000	0.013270
ret	0.016167	-0.130353	0.013270	1.000000

如图可分析得到 PB 与 ROE 成正相关, 与 DPR 成负相关, 与 ret 相关系数较小, 我们对 ret 的统计显著性持怀疑态度。

3. 线性回归

- (1) 对 PB 与 DPR 做线性回归, 得到如下结果:

Out[9]: OLS Regression Results

Dep. Variable:	PB	R-squared:	0.038
Model:	OLS	Adj. R-squared:	0.034
Method:	Least Squares	F-statistic:	10.40
Date:	Sat, 12 Mar 2022	Prob (F-statistic):	0.00142
Time:	18:45:45	Log-Likelihood:	-381.71
No. Observations:	266	AIC:	767.4
Df Residuals:	264	BIC:	774.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.9030	0.317	12.324	0.000	3.279	4.527
DPR	-3.2187	0.998	-3.224	0.001	-5.184	-1.253

Omnibus:	24.028	Durbin-Watson:	0.074
Prob(Omnibus):	0.000	Jarque-Bera (JB):	28.831
Skew:	0.804	Prob(JB):	5.49e-07
Kurtosis:	2.871	Cond. No.	17.5

$$PB = -3.2187 \times DPR + 3.9030$$

$$(0.998) \quad (0.317)$$

进行 t 检验后得到 p-value = 0.001，得到 99.9%的可能 DPR 的系数不为 0。

(2) 对 PB 与 ROE 做线性回归，得到如下结果：

Out[10]: OLS Regression Results

Dep. Variable:	PB	R-squared:	0.116
Model:	OLS	Adj. R-squared:	0.112
Method:	Least Squares	F-statistic:	34.54
Date:	Sat, 12 Mar 2022	Prob (F-statistic):	1.26e-08
Time:	18:45:58	Log-Likelihood:	-370.49
No. Observations:	266	AIC:	745.0
Df Residuals:	264	BIC:	752.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.4508	0.421	1.070	0.286	-0.379	1.280
ROE	29.8427	5.078	5.877	0.000	19.844	39.841

Omnibus:	11.115	Durbin-Watson:	0.088
Prob(Omnibus):	0.004	Jarque-Bera (JB):	11.529
Skew:	0.484	Prob(JB):	0.00314
Kurtosis:	2.681	Cond. No.	85.3

$$PB = 29.8427 \times ROE + 0.4508$$

$$(5.078) \quad (0.421)$$

进行 t 检验后得到 p-value = 0.000，得到超过 99.9%的可能 ROE 的系数不为 0。

(3) 对 PB 与 ret 做线性回归，得到如下结果：

Out[11]: OLS Regression Results

Dep. Variable:	PB	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.004
Method:	Least Squares	F-statistic:	0.06902
Date:	Sat, 12 Mar 2022	Prob (F-statistic):	0.793
Time:	18:46:09	Log-Likelihood:	-386.81
No. Observations:	266	AIC:	777.6
Df Residuals:	264	BIC:	784.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.8997	0.064	45.095	0.000	2.773	3.026
ret	0.2039	0.776	0.263	0.793	-1.324	1.732

Omnibus:	23.524	Durbin-Watson:	0.073
Prob(Omnibus):	0.000	Jarque-Bera (JB):	28.060
Skew:	0.795	Prob(JB):	8.07e-07
Kurtosis:	2.932	Cond. No.	12.2

$$PB = 0.2039 \times ret + 2.8997$$

(0.776) (0.064)

进行 t 检验后得到 p-value = 0.793，得到有 79.3%的可能 ret 的系数为 0。故该变量统计性不显著。

二、统计结果分析

由前文结果可得，PB 主要有两个影响因素：ROE 和 DPR。在 PB 基本公式推导中，公司的回报率也是一个因素，但由于本题要求数据量过大，从 CSMAR 数据库中取得的回报率是平均回报率，不能很好的反映各个公司本身的 PB 值，故在统计性检验中没有通过。下面分析 ROE 和 DPR 对 PB 的影响：

1. 对于 PB & ROE:

由两者定义可以推导得到

$$PB = ROE \times \frac{\text{Price Per Share}}{EPS}$$

而前文中 ROE 系数为正，故我们可以得到上市公司股票的价格与该公司净利润状况呈正相关。引申得到当公司营业状况较好时该公司股票价格会呈上涨趋势。

2. 对于 PB & DPR:

由两者定义可以推导得到 DPR 与 EPS 呈反比, PB 与 EPS 成正比。而前文中 DPR 系数为负, 故证明了 PB 与 DPR 成负相关。引申得到若 DPR 越高, 则股市对于该公司 PB 的态度消极, 得到的 PB 呈下降态势。

问题三

一、研究方法

（一）数据处理

1. 数据来源

选取2000年1月至今的等权平均法和流通市值加权平均法计算的考虑现金红利再投资的月市场回报率作为本题研究的原始数据。

2. 数据处理

使用R的tidyverse包和lubridate包对原始数据进行处理。对数据中的缺失值，均予以整行清除，防止后续计算结果出现NA。将日期列转化为R的date对象以便后续操作。由于主板中小板包含了市场类型为1和4的数据，即每个月有两个月收益率的值，因此对其按月份groupby，然后取两个月收益率的平均值作为代表。得到两个板块的月收益率后，使用累乘计算累计月回报率（具体使用的公式已在名词解释部分说明）放入已经设置包含了相应日期的数据框，至此数据处理完成。详细代码与注释请见CumulativeMonthlyReturn.R。

（二）图像绘制

对上述处理完的数据使用ggplot进行绘图，以时间为x轴，累计月收益率为y轴，分主板、中小板与创业板分别绘图，并添加 $y=1$ （累计收益率=100%）的辅助指示线，使回报率盈亏更清晰。

二、问题探究

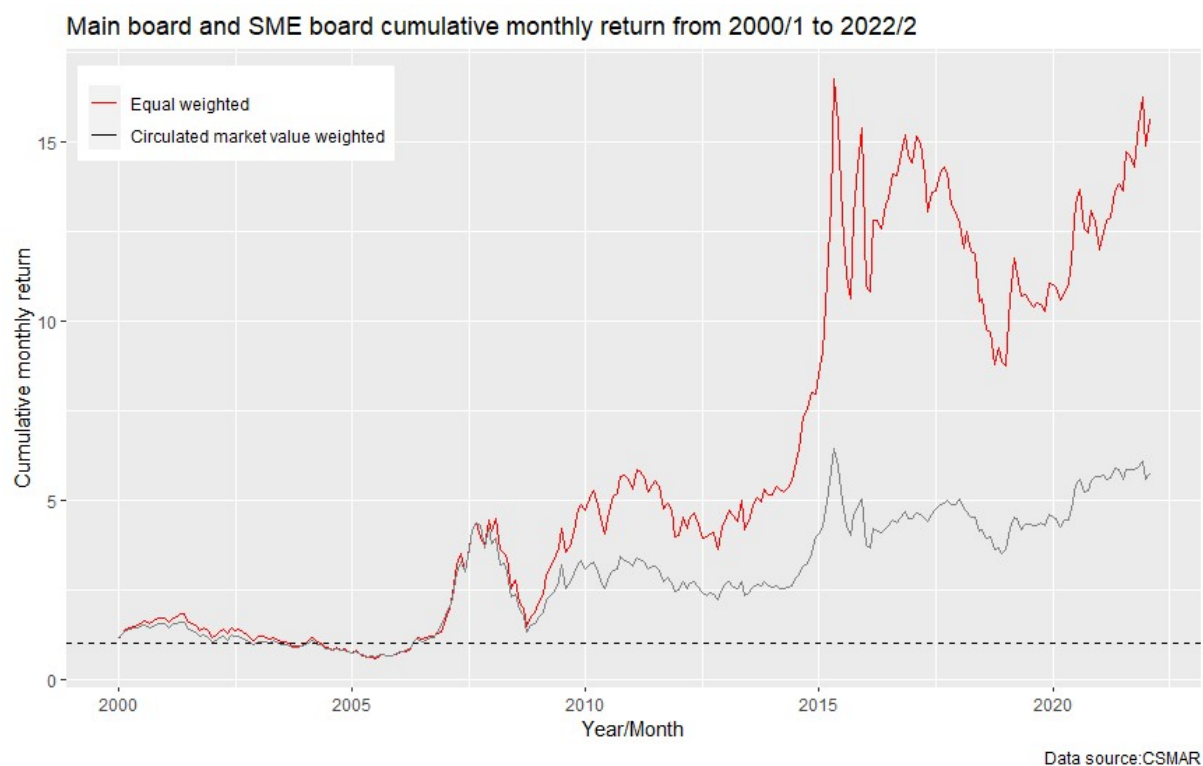


图3：主板、中小板累计月回报率（考虑现金红利再投资）

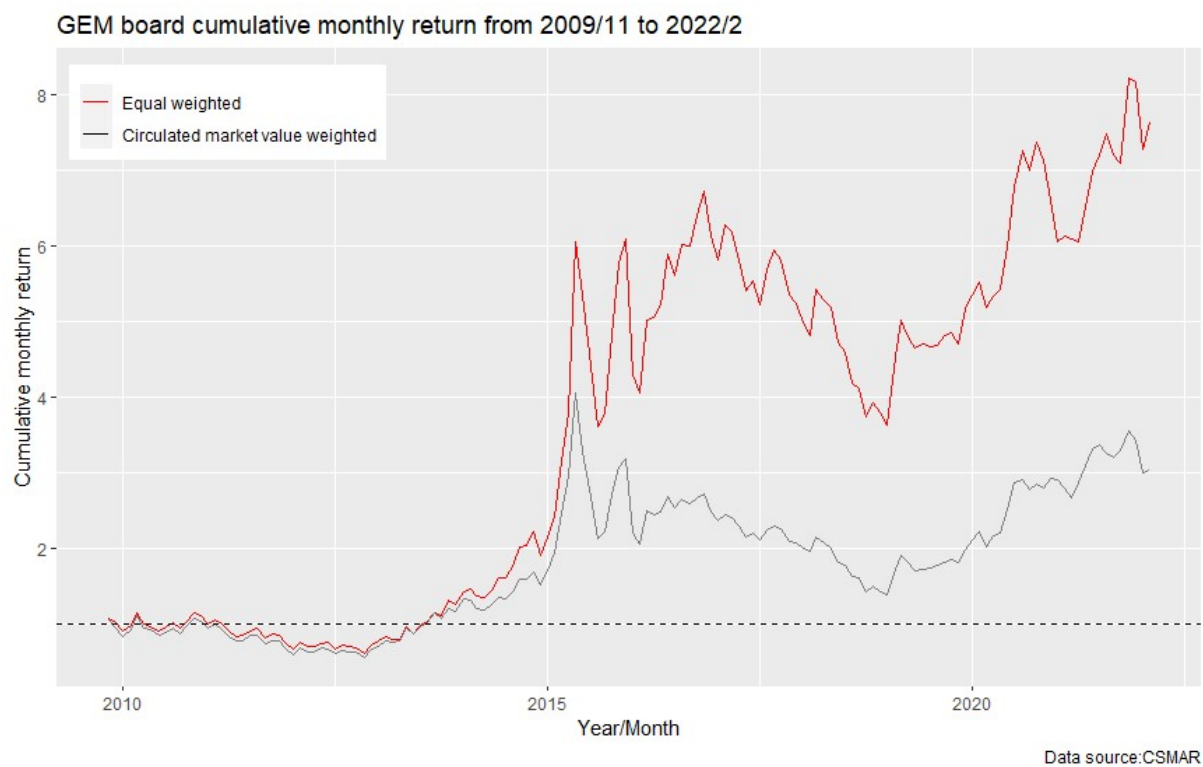


图4：创业板累计月回报率（考虑现金红利再投资）

（一）图像解读

两种方法计算的主板、中小板的股票累计月回报率走势类似。在股票市场建立初期保持平稳增长，在2005年左右一度下滑至亏损的累计月回报率；而在2008年附近先快速增长，再快速消退；随后保持了一段时期的小幅波动后，在2015年大幅增长，又于2019-2020年下滑，2020年后继续保持增长。创业板的两条线初期波动幅度较大，特别是2011-2013年时期都是小于1的状态，2015年后的变化则类似主板、中小板的变化趋势。对两个板块而言，都是等权平均法计算的比流通市值加权平均法计算的累计月回报率高非常多，且波动幅度也相对较大，时间累计越久则越明显。

（二）考虑现金红利再投资与否对累计月回报率的影响

由于考虑现金红利再投资，本应作为dividend散发出去的回报重新作为股本投资，因此会使个股回报率上升。且类似复利的影响，考虑现金红利再投资的累计回报率雪球会越滚越大。

（三）等权平均法与流通市值加权平均法计算的月回报率的区别

等权平均法计算的回报率在不同股票股价百分比变动相似的情况下，等权平均法计算回报率会较好地反映出这一变化，而不受股价绝对水平的影响。其缺点是，在这一算法中，不同规模公司的股价变化幅度对结果带来的影响相同，很多市值较小的公司对市场回报率有着比较大的影响，因此不能很好地反映出对市场的调性。流通市值加权平均法计算的回报率则避免了上述缺点：对股票按流通市值进行加权，可以使市场回报率更好地反映整个股票市场的情况。

因此，在本研究中，等权平均法计算的比流通市值加权平均法计算的累计月回报率高且波动幅度大的原因就是等权平均计算时，把对市场影响较小的公司的影响力扩大了。

（四）宏观经济对累计月回报率的影响

在图像解读中分析的数据变化趋势，均与当时的大背景相一致。如2008年的金融危机，2015年中国牛市，此部分在问题一的探究中已经阐述，在此不再赘述；2019-2020年新冠疫情爆发使许多经济体一度停摆，因此股票月回报率大幅度下跌，导致累计回报下降；2021开始的后疫情时代，全球经济开始复苏，股票收益再次提升。