基于中国A股数据的资本资产定价模型实证研究

任凯 石玺 王智城

摘要

本文试图运用2000年到2019年中国A股市场的数据，对资本资产定价模型进行实证检验。首先，我们采用Fama-MacBetch回归对标准 CAPM 模型进行验证，尽管得到的结果很好地支持了三个核心假设，但是无法得出“收益和风险之间显著正相关”这一重要结论。其次，我们又分析了Fama-French三因子模型，想要把规模因素和价值因素考虑到资本资产定价中，发现该模型表现良好。最后，我们考虑了中国股市特有的IPO监管造成的壳价值问题，尝试使用剔除最低市值30%股票的方式来尽量削减影响，对在三因子模型中发现的部分异常做解释。（尽管文章并未提出创新性的思路和方法，但是对模型复现的过程让笔者对资本资产定价模型的实质有了深刻的认识）

1. 资本资产定价模型简介
2. 标准的 CAPM 模型

二十世纪六、七十年代，Sharpe（1964）、Lintner（1965）和Black（1972）将Markowitz理论（1959）延伸成为资本资产定价模型（Capital and Asset Pricing Model，CAPM）。CAPM 用方程表示为：

,（1）<这个1能否右对齐？>

其中是资产的期望收益率，指无风险利率，为市场组合的期望收益率，它在理论上是指由所有的风险资产共同组成的投资组合，表示资产的系统风险，它被定义为资产收益与市场组合收益之间的协方差同市场组合收益方差之间的比：



CAPM模型是在严格的假设条件下进行的理论分析模型，假设包括：

1. **均值方差假设：**投资者通过观察一段时间内证券组合的预期收益率和标准差来评价证券组合，若标准差及其他方面等同，投资者将选择具有较高收益率的证券组合；若预期收益率等同，投资者将选择具有较低标准差的一种证券组合。

2. **投资者一致假设：**投资者计划的投资时点和投资期限相同，组成各个投资者组合的证券数 目相同，投资者对证券的预期收益率 、标准差 、协方差看法一致，从而保证市场有效边际只有一个；投资者选择不同投资组合的原因只是风险偏好不同。

3. **完全市场假设：**市场不存在交易成本和税收，所有资产完全可分割、可交易，市场是完全竞争的，信息成本为零，所有市场参与者同时接受信息，且都是理性的。

标准CAPM模型主要结论是：任何资产的期望收益率与其值呈线性关系。如果市场达到均衡，市场上的所有证券的风险收益定价关系都应在证券市场线上。Black、Jensen 和 Scholes（1972）以及Fama和MacBetch（1972）也都发现股票的平均收益和存在着显著的正向线性关系。

1. Fama-French 三因子模型

二十年代后期，一些实证分析结果向CAPM模型发起挑战，其中最引人瞩目的就是Banz（1981））提出的规模效应，他发现市场权益ME（股票的价格乘以流通股数目）增加了值提供的对平均收益率的横截面解释。小股票（低ME）的平均回报率在用估计时偏高，而大股票则偏低。

Fama和French（1992a）也像 Ringanum（1981）以及Lakonishok 和 Shapiro（1986）一样，都发现了美国股票市场1963至1990年期间，风险和平均收益之间的关系消失了。他们注意到：股票平均收益率与规模、杠杆、E/P以及账面市值比之间的单变量关系很强。在多变量测试中，规模和平均收益率之间的负相关关系是稳健的；账面市值比与平均收益率之间的正向关系也持续存在，进而认为如果资产的定价是理性的，那么股票风险应该是多维度的，风险的一个维度是由规模来代表的，另一个维度是由账面市值比来代表的。

他们的实证结果表明：规模（ME）和账面市值比（BE/ME）为1963-1990年期间的平均股票收益截面提供了一个简单而有力的描述，在此基础上提出了“Fama-French三因子模型”（Fama 和 French 1993）：



其中是表示时间点，、、分别表示时投资组合的收益率、无风险利率以及市场组合的收益率。、为在标准CAPM模型（1）的基础上增加的两个因子，分别表示小市值组合收益率-大市值组合收益率、高账面市值比组合收益率-低账面市值比组合收益率。该模型分析结果表明：对于时间序列回归说，规模和账面市值因素可以解释不同股票平均回报的差异。

本文基于中国A股数据分对标准CAPM模型和Fama-French三因子模型进行实证分析，其余部分将按照如下步骤进行。第二节将说明数据来源，第三节将介绍采用Fama-MacBatch回归的方式对标准CAPM模型的分析结果，第四节则是构建了Fama-French三因子模型并进行分析，第五节讨论了上述两种方法的分析结果，考虑到分析结果比较差提出现阶段最新研究给出的猜想。第六节总结了我们的结论。

# 二、 数据来源

本文使用的数据，包括股票的交易数据和指数的交易数据均来自RESSET数据库，无风险利率采用一年期定期存款利率。

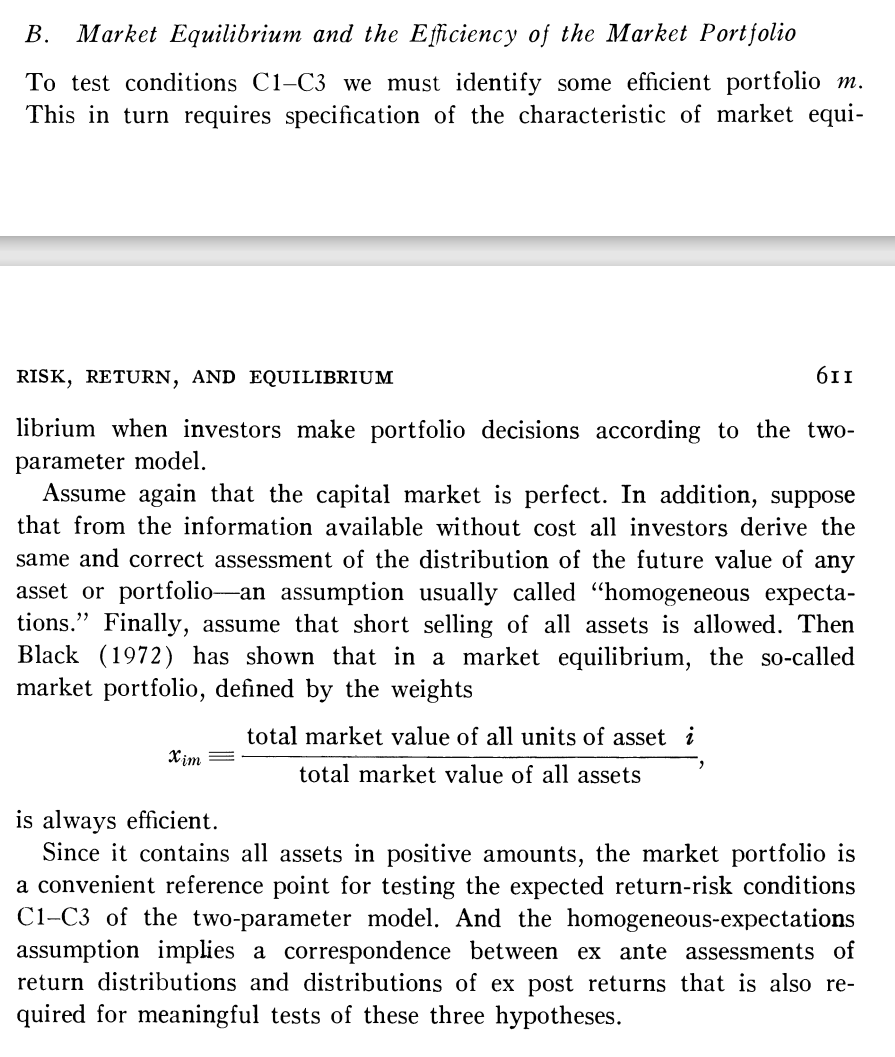
我们所分析的数据频率为月度，时间是从2000年1月至2019年12月。中国A股市场，始于1990年上海和深圳交易所的成立。我们把研究重点放在2000年与2019年之间，主要有三个原因：首先是为了保证会计数据的统一性。在中国，有关财务报告各方面的规则和条例的实施直到1999年才基本成型。虽然1993年开始有了公平交易和财务披露原则，但企业在遵守这些原则方面得到的指导很少，各公司自由发挥，强加自己的标准，限制了各个公司会计数据的可比性。直到1998年和1999年关于交易和财务报告的法律和法规才得到更加彻底的设计和实施，直到1999年会计准则的统一才得以广泛实现，因此2000年以后的投资组合回报样本所依赖的会计数据比前几年更具有可比性。其次是为了确保有足够数量的观察数据（画一个图呈现数据的变化）。最后，考虑到2020年以来疫情对市场产生了巨大的冲击，将这一部分的异常数据进行剔除。

# 三、 采用Fama-MacBatch回归对标准 CAPM 模型进行验证

**3.1 假设**

方程（1）有着3个需要被验证的基础条件，分别是：（C1）在任何有效组合中，证券的预期收益和它的风险之间的关系是线性的；（C2）是有效组合中证券的风险的完整度量，（1）中没有出现证券的其他风险度量；（C3）在一个由风险规避型投资者组成的市场中，较高的风险应该与较高的预期收益相关，也就是说，。

A．



***B． 收益的随机模型***

方程式（1）是以预期收益为基础的，它必须用逐期的证券和投资组合收益的数据来检验，这些数据实际是不可得的。因此，我们希望选择一个逐期收益的模型，这样才能使用观察到的平均收益率来验证C1~C3，所以本文对（1）进行如下的随机概括：

（2）

其中，指的是第期，是证券从期到期的单期百分比收益，方程（2）允许和在不同期内随机变化。条件C3成立对应于方程（1）中的斜率是正的，对应于方程二中（2）中风险溢价的期望是正的，也即。

方程（2）中引入是为了检验线性，条件C1成立所对应的等价假设为。类似的声明也适用于（2）中加入的，它是对证券风险的某种衡量，与没有确定性关系。条件C2所对应的假设是。当然对比（1）和（2）有一个显而易见需要被验证的假设那就是

***C． 模型的核心假设***

总的来说，考虑到（2）是对（1）的随机概括形式，因此本文给出标准形式CAPM模型对预期收益的可检验假设为：

*H1（线性）：；*

*H2（是风险的完整度量）：*

*H3（正预期的收益—风险交易）：*

*H4（Sharpe-Lintner假设）：*

注意，我们在设计可检验假设时的确发现Fama（1973）还提到另外一个假设，即市场有效性假设，但综合考虑到这一假设主要侧重于市场有效性并且实现起来较为复杂，本文后续并未对其进行检验。

**3.2 方法**

***A．方法概述***

要想检验标准CAPM模型，不得不面对“变量误差”问题，预期收益—风险方程（1）需要采用相对风险度量的真实值，但在分析中只能使用估计值。此处，我们采用下式进行估计：



其中**和**是使用月度收益数据对**和**的估计，在这里我们选用“上证综合指数”作为**的近似替代。（对上证综合指数做一个简要说明）

Blume（1970）证明对于任何的组合，如果是用权重来定义，那么组合的风险衡量就可以表示为：



经过严谨数学推导后，不难发现：如果的误差大大低于完全正相关，那么投资组合的就可以比单支证券的更精确地估计真实的，**因此相较于本小组在平时作业中采用单只股票检验CAPM，本文最大的改进之处就在于将采用构建投资组的方式完成检验。**

同时，为了减少因使用投资组合而不是单个证券造成的风险—收益检验中的信息损失，一般都是在对单个证券值进行排序的基础上构建投资组合，进而获得一般投资组合的值。但需要注意的是，如果只是简单地进行操作很可能会导致严重的“回归现象”（Regress Phenomenon）：考虑到在截面上，观察到的较大的往往高于相应的真实，观察到的较小的往往低于真实，进而在对排序的基础上构建投资组合会造成投资组合内正负抽样误差的堆积，其结果是，一个较大的投资组合风险会倾向于高估真实的，而较小的则倾向于低估真实值。

为了尽可能地避免“回归现象”，本文采用如下方法来完成投资组合的构建以及后续模型的计算：

* 步骤1. 把基于一个时间段内数据计算出的证券进行排序，作为构建投资组合的依据，我们把这段时间称为“组合构造期”；
* 步骤2. 用随后时间段中的数据，估计出这些投资组合中每一个证券的初始，把这段初始估计的时间称为“初始估计期”；
* 步骤 3. 将初始估计期按年向前滚动，估计后续时间段内投资组合每一年的，同时获取月度的收益率均值数据，基于此完成模型的回归估计，将这段时间称为“模型检测期”。

这样的方式使用后面时间段的“新”数据，能够保证在一个投资组合中单个证券的误差在很大程度上是随机的，进而使得在一个投资组合的中回归现象能够被最小化。

***B．方法细节***

图1.以2000到2012年数据区间为例，展示了本文进行参数估计的细节。设为证券总数，设为等于或小于的最大整数。首先使用前4年（2000—2003）“组合构造期”的月度收益数据估计出每一支证券的，将其排序并作为构造20个投资组合的参考，1号投资组合到20号投资组合中证券的递减，也即最大的证券在1号组合中，最小的证券在20号组合中。这样的话，如果是偶数，第一个和最后一个组合各有个证券，如果是奇数则最后一个投资组合（最小）有相对多出的证券。

接下来的5年（2004—2008年）“初始估计期”的数据被用来重新计算，这些数据在投资组合内的不同证券之间进行平均，以获得20个初始投资组合，用于风险—收益检验。和递增下标表示在接下来的4年（2008—2011年）的每年中这些被重新计算为单个证券的简单平均值，从而逐年调整组合，以考虑个别证券的退市。注意到，本文和Fama（1973）所用的方法在此处并不完全一致，其对进行的是逐月调整，但本文考虑到数值的稳定性即变动应以年为周期，故采用逐年的方法来进行调整。

另外，我们使用作为对证券的非风险的衡量，其来自下述市场模型（3）最小二乘残差的标准偏差：

（3）

标准差是在以下意义上对非风险的衡量：一种与投资组合理论相反的风险观点认为，证券的风险是由其收益分布的总离散度来衡量的。考虑到一个由风险规避者主导的市场，这个模型将预测一个证券的预期收益与它的总收益分散性有关，而不仅仅是与该证券对有效组合的收益分散性的贡献有关。如果那么在（3）中，进而有：

 （4）

因此，从（4）中，我们可以说是对证券收益分布中与不直接相关的那部分分散性的估计。

在2009—2012年的4年期间，20个投资组合的逐月回报率也被计算出来，每个月的单个证券的权重相同。对于这一时期的每个月，运行以下横截面回归：

（5）

自变量是上述投资组合中证券的的平均值；是这些的平方值的平均值；而同样是投资组合中证券的的平均值。

回归方程（5）是（2）对一个投资组合中的证券进行平均化，估计值，，以及作为解释变量，并对随机系数，，以及进行最小二乘估计。方程（5）的结果，也即对2009—2012年4年间逐月回归系数，，以及，是我们对这一时期标准CAPM模型假设检验的依据。要想获得其他时期的检验结果，只需重复执行上述步骤：用7年的数据来形成投资组合，接下来的5年用来计算（5）中独立变量的初始值；然后在接下来的4年中逐月拟合（5）的风险—收益回归。

表1显示本文基于数据选取的3个不同的投资组合构造期（除第一个都是7年的长度）、5年的初始估计期和测试期（除了最后一个都是4年）。选择4年的测试期是为了平衡计算成本。选择7年的投资组合形成期和5年的初始估计期来估计风险—收益回归中的独立变量和，反映了平衡从静止过程中获得的大样本的统计能力与p的任何非恒定性所引起的潜在问题的愿望。因此我们要对不满足需要的数据进行剔除，本文为了操作方便，所以选取了2000年到2019年20年间持续进行交易的股票数据，共有724支股票，但是其中有部分股票个别月份停止交易，这对整体计算影响并不大故不作考虑。

表1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 时期 |  |  |
|  | 1 | 2 | 3 |
| 组合构造期 | 2000-2003 | 2001-2007 | 2005-2011 |
| 初始估计期 | 2004-2008 | 2008-2012 | 2012-2016 |
| 模型检测期 | 2009-2012 | 2013-2016 | 2017-2019 |

***C．方法实施示例***

本文附录代码FamaMacBatch.do文件对方法的整体过程做了具体实现，为了方便理解，此处我们以表1.中的时期1（2000年至2012年）为例，展示方法的实施具体实施步骤。

步骤1. 2000年至2003年，基于48个月的“组合构造期”中的数据，通过市场模型（3）估计出每一支股票的：



对所有股票的进行排序，由小至大等分为20份，相应地构建起20个投资组合：最小5%对应的股票为投资组合1（38支股票），其次为投资组合2（36支股票）以此类推，最大的5%对应的证券为投资组合20（38支股票）。

步骤2. 2004年至2008年，基于60个月“初始估计期”中的数据，用同样的方法估计出该区间中每一支股票的以及，按照上述步骤1.得到的投资组合划分，依靠等权求平均的方式，获得20个投资组合在当期的以及：



然后计算“模型检测期”第一年2009年12个月20个投资组合的平均收益率



向前滚动1年，基于2004年至2009年72个月的数据，重复上述操作，得到当期以及：



然后计算“模型检测期”第二年2010年12个月20个投资组合的平均收益率



继续向前滚动1年，基于2004年至2010年84个月的数据，重复上述操作，得到当期以及：



然后计算“模型检测期”第三年2011年12个月20个投资组合的平均收益率



继续向前滚动1年，基于2004年至2011年96个月的数据，重复上述操作，得到当期以及：



然后计算“模型检测期”第四年2012年12个月20个投资组合的平均收益率



步骤3. 对2009年至2012年4年的数据做汇总，得到如表2.所示的数据结构，为后续回归估计做准备。

在时期2和时期3上重复上述操作，最终得到20个投资组合在132个月份上的数据。

表2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 组合 | 时间 |  |  |  |  |
| 组合1 | 2009年1月 |  |  |  |  |
|  |  |  |  |  |
| 2010年12月 |  |  |  |  |
|  |  |  |  |  |
| 2012年1月 |  |  |  |  |
|  |  |  |  |  |
| 2012年12月 |  |  |  |  |
|  |  |  |  |  |  |
| 组合20 | 2009年1月 |  |  |  |  |
|  |  |  |  |  |
| 2010年12月 |  |  |  |  |
|  |  |  |  |  |
| 2012年1月 |  |  |  |  |
|  |  |  |  |  |
| 2012年12月 |  |  |  |  |

***D．初步观察***

表3.显示了20个投资组合在三个“初始估计期”中的的值和他们的标准误。同时显示的还有：也即和之间的回归系数；以及，来自（3）市场模型的投资组合残差的标准差，注意其和的不同，后者是单个证券的平均值，也在表中显示出来。

在这样的假设下：对于一个给定的证券，（3）中的干扰是连续独立的，与无关，并且通过时间完全分散，那么的标准误差为：

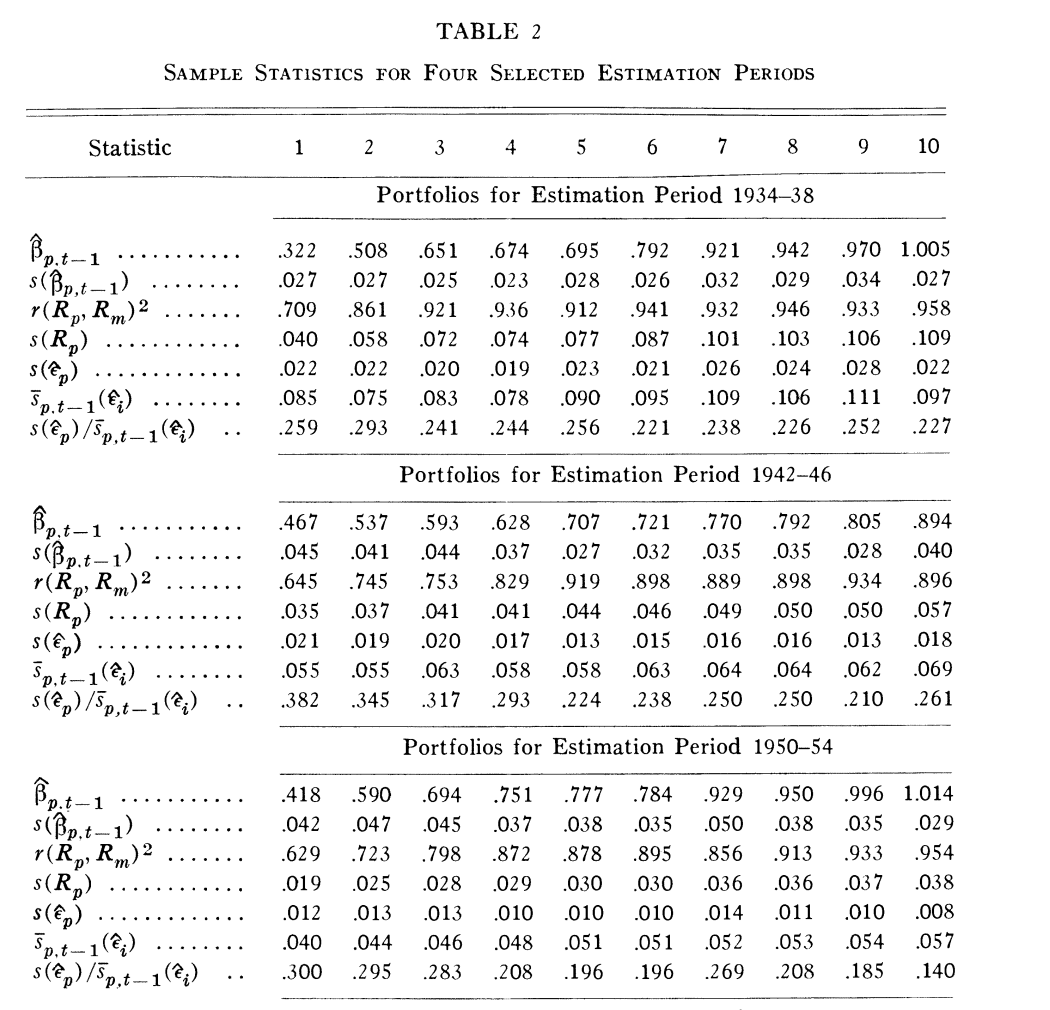


其中是用于计算的月数，类似的：



在表（2）中可以看出要比小，这也证明了：通过组合来评估要比通过单支证券来估计要精准。

表3.



**3.3 结果**

表4.中展示了对标准CAPM模型的检验结果，本文对全样本区间（2009—2019）和三个子样本区间（2009—2012、2013—2016、2017—2019）做了分析。结果显示了四个不同版本的风险—收益回归方程（4）。D组是基于（4）本身，在A—C组中，（4）中的一个或多个变量被去除。表格中显示了，逐月回归系数估计值的平均值；每月估计值的标准差，本表还预先提供了用于检验的假设的统计量：



其中是时期内的月份数，也是用于计算和的数量。

在解释这些统计量时，我们应该记住Fama（1965a）和Blume（1970）的结论，这些结论表明，相对于正态分布而言，普通股票的收益分布是“厚尾”的，而且可能更符合非正态对称稳定分布，而不是正态的。从Fama和Babiak（1968）的观点来看，这一结论意味着，当人们在基础变量为正态的假设下解释大的统计量时，得到的概率或显著性水平可能是高估的。但需要注意的是，除了假设H3（正的预期收益—风险权衡）之外，向上偏的概率水平会导致拒绝双参数模型的假设的偏差。**因此，如果在正态性假设下解释****统计量时不能拒绝这些假设，那么当我们考虑到经验收益分布的厚尾时，这些假设就有了更坚实的基础。**

表4.



|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 参数 | | | | | | | | |
| 时间区间 |  |  |  |  |  |  |  |  |  |
| A： |  |  |  |  |  |  |  |  |  |
| 2009—2019 | .0107 | 1.06 | .0013 | .1765 | ··· | ··· | ··· | ··· | .09 |
| 2009—2012 | .0066 | .32 | .0103 | .70 | ··· | ··· | ··· | ··· | .04 |
| 2013—2016 | .0306 | 2.02 | -.0070 | -.74 | ··· | ··· | ··· | ··· | .10 |
| 2017—2019 | -.0071 | -.67 | .0004 | .02 | ··· | ··· | ··· | ··· | .15 |
| B： |  |  |  |  |  |  |  |  |  |
| 2009—2019 | -.0393 | -.21 | .0993 | .27 | -.0477 | -.26 | ··· | ··· | .17 |
| 2009—2012 | -.1608 | -.31 | .3268 | .32 | -.1488 | -.30 | ··· | ··· | .12 |
| 2013—2016 | .0113 | .12 | .0402 | .22 | -.0270 | -.32 | ··· | ··· | .17 |
| 2017—2019 | .0553 | .81 | -.1250 | -.93 | .0595 | .96 | ··· | ··· | .22 |
| C： |  |  |  |  |  |  |  |  |  |
| 2009—2019 | .0207 | .17 | .0008 | 0.11 | ··· | ··· | -.0738 | -.08 | .17 |
| 2009—2012 | -.2253 | -.13 | .01 | .74 | ··· | ··· | .2254 | .17 | .14 |
| 2013—2016 | .0767 | .31 | -.0091 | -1.02 | ··· | ··· | -.3437 | -.18 | .16 |
| 2017—2019 | .0039 | .02 | .0005 | .03 | ··· | ··· | -.1128 | -.08 | .21 |
| D： |  |  |  |  |  |  |  |  |  |
| 2009—2019 | -.0372 | -.14 | .1085 | .27 | -.0528 | -.27 | -.0236 | -.02 | .23 |
| 2009—2012 | -.2823 | -.50 | .3874 | .37 | -.1807 | -.35 | .7314 | .54 | .22 |
| 2013—2016 | .0233 | .04 | .0375 | .13 | -.0248 | -.18 | -.0912 | -.03 | .22 |
| 2017—2019 | .1974 | .82 | -.1686 | -1.14 | .0804 | 1.16 | -.9403 | -.60 | .28 |

1. ***核心假设检验***

先来看*H1*，表4.中B组和D组的结果**没有拒绝标准CAPM模型的条件C1**，即预期收益率和**之间的关系是线性的。不难发现，B组和D组之中无论是全样本时间段还是子时间间段内的均十分接近于0，尽管子时间段2017—2019内的较大，但也都小于1.8，仍然无法显著拒绝原假设。

再考虑标准CAPM模型的条件C2所对应的假设*H2*，即除了**以外，没有任何其他风险会系统性地影响预期收益。表4.中C组和D组的结果没有拒绝这一假设，可以发现都十分接近于0，且的符号随机分为正和负。

如果只看前面两个假设的话，标准CAPM模型似乎能很好地解释数据。然而，如果关键条件C3对应的假设*H3*被拒绝，以上得到结论的价值都会大打折扣。也就是说，除非风险和收益之间平均存在正的相关性，否则我们对标准CAPM模型的验证就无法达到满意的地步。然而，令人失望的是本文的结果和Fama（1973）的结果大相径庭，所有模型的任意时间区间内的均十分小，也就是说表4.中的结果让我们能够拒绝*H3*。

然而值得欣慰的是，在对*H4*检验上，我们的数据结果似乎并不像Fama（1973）中提到的是“ambiguous”的，四组模型中仅有A组中两个阶段的数据支持我们拒绝*H4*，其余部分的均十分接近于0，表4.中的结果让我们无法拒绝*H4*。

**3.4 对标准CAPM模型检验的结论**

综上，我们的结果部分说明了标准CAPM模型的有效性。本文的结果有效说明了，尽管在不同时期存在“随机非线性”，但是我们不能拒绝这样的假设，即在做出投资组合决策时，投资者应该假设证券的投资组合风险和其预期收益之间的关系是线性的。我们的结果同样也证明：不能拒绝除了投资组合风险外，没有任何风险措施会系统地影响平均收益这一假设。而且也无法拒绝无风险利率在标准CAPM模型中扮演的重要角色。但是不可否认的是，我们基于中国A股数据无法得到“平均而言，收益和风险之间应该有一个正向的关系”这一重要结论，这也促使我们进一步展开了接下来对Fama-French三因子模型的实证研究。

# 四、Fama-French 三因子模型

**4.1 数据补充说明**

由于后续操作将引入新的因子，故在此处对上述数据进行补充说明。首先，我们和Fama（1992）一样排除了金融公司，因为这些公司的指标可能与非金融公司的意义不同。另外，我们还剔除了状态为ST（特别处理）、\*ST（退市风险警示）或PT（特别转让）的股票，以及账面价值为负的股票，避免异常指标对数据的干扰。

本文把公司资产负债表中披露年的账面所有者权益作为年账面价值，把公司披露的年的市场总值作为年的市场价值。同时，不再直接采用上证综合指数作为市场回报率，而是采用流通市值加权平均法计算得到。

注意，规范信息披露制度下上市公司年度财务报表一般在次年3、4月份公布，且在一季度财务报表公布之前披露，即4月底为年报披露截止期。财务报表信息披露的滞后性造成金融数据库中财务数据与市场数据的时间不一致，即年末的财务数据因年报公布的滞后性在年4月份披露财务报表时才得到更新，因此选取年5月至年4月作为组合构建周期。例如某只股票母公司在2002年4月底才披露财务报表，那么2002年1月至4月只能依靠2001年披露的财务报表数据来进行分组，而2002年5月至2003年4月可以采用2002年4月底披露的数据进行分组，以此类推。

**4.2 时间序列回归的输入**

不同于上一部分对采用的截面回归，分析Fama-French三因子模型需要进行的是时间序列回归。正如（2）式所呈现的那样，解释变量包括市场组合的收益率、市值、账面市值比三个因素，被解释变量为根据市值和账面市值比形成的25个股票组合的收益率。下面将详细介绍获取这些输入数据的方法，同时对输入数据的结构做展示，最后做了一些基本的描述性统计。

***A.股票三因子***

***添加因子的动机——***市值和账面市值比似乎是解释平均股票收益的特别变量，我们有理由相信它们能代表收益中的基本风险因素。Fama和French（1992b）记录了市值和账面市值比与收益率之间的关系：高账面市值比（）的公司往往有较高的资产收益，至少在五年前和五年后的对账面市值比公平的测量中收益率较高。市值也与盈利能力有关，控制账面市值比变量，小公司往往比大公司的资产收益率高。但是，收入的市值效应仅仅到了20世纪80年代。直到1981，控制账面市值比因子，小公司的利润只比大公司略高一些。但对于小公司来说，1982年的经济衰退变成了长期的经济萧条。同时还由于某种原因，小公司没有享受到20世纪80年代中期和后期的经济繁荣。

小公司可能会遭受长期的收益率低落，而大公司则可以避免这种情况。这表明市值与一个市场风险因子有关，这可能解释了市值和平均收益率之间的负相关关系。与之类似，账面市值与收益率之间的关系表明相对盈利能力是收益率中一个共同市场风险因子的来源，这可能解释了和平均收益率的正相关关系。三因子模型的提出正是为了测量收益率中与市值和BE/ME有关的共同变化。

***计算因子的组合构建——***为了研究月度的经济基本面，Fama和French（1992b）使用了由在和上的排序的各种股票形成的六个投资组合。本文使用同样的方式来形成投资组合，旨在模仿与市值和账面权益相关的收益中的基本风险因素。这确保了回报率中的基本风险因素研究与我们对经济基本面的补充研究之间的对应关系。

表x.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 低（） | 中（） | 高（） |
| 小（） |  |  |  |
| 大（） |  |  |  |

根据Fama和French（1992a）的证据表明，账面市值比在解释股票收益率上比市值具有更大的作用。因此本文决定将根据将公司分为三类，根据将公司分为两类。但是，这种分类是任意的，依据的FAMA和French（1992b）的敏感性检验。正如表x.所示，一方面以年4月A股所有股票中位数将公司分为小和大（S和B）两组，另一方面根据此时股票底部30%（Low）、中间40%（Medium）以及最高30%（High）三个断点将公司分为低、中和高三组（L、M和H）。从我们对的两个分组以及对的三个分组中，我们构造了六个投资组合（，，，，，）。例如投资组合包含在小组中的股票，它们也在低组中，组合包含大的股票，它们也在高组中。六个投资组合的月加权收益率从年5月至年4月计算，投资组合在年4月份重新分组。正如上面提到的，本文选择从年5月开始计算收益率，以确保年的市值以及账面市值比这两大指标是已知的。

***市值因子——***我们用投资组合计算来模仿与市值相关的收益率的风险因子，其是每个月3个小市值投资组合（，，）以及3个大市值投资组合（，，）在平均收益率之间的差额：



因此，是小型和大型股票投资组合的收益率之间的差额，将账面市值比作为控制变量。这种差异应该在很大程度上不受的影响，而是集中在大型与小型公司的不同平均收益率。

***账面市值比因子——***类似的，我们采用模仿与账面市值比相关收益率的风险因子。是每个月在两个高投资组合（和）收益率的简单平均值和两个低投资组合（和）收益率的平均值之间的差额：



的两个组成部分是高投资组合和低投资组合的收益率，它们的加权平均市值大致相同。因此，两个收益率之间的差异应该在很大程度上与市值因子无关，而是集中在高和低公司不同的收益率行为。

注意，和标准CAPM模型检验中不同的是，此处投资组合收益率是由个股市值加权而来的，而不是直接平均得到的。价值加权一方面是为了最小化方差，另一方面可以更为准确地模拟投资组合，捕捉小股和大股不同的收益率行为。或高和低的股票，以一种符合现实投资机会的方式。

***市场因子——***股票收益率的市场因子反映的是超额市场收益率，即。其中，是由个股流通市值加权平均法计算得到市场组合收益率， 是一年期定期存款利率。

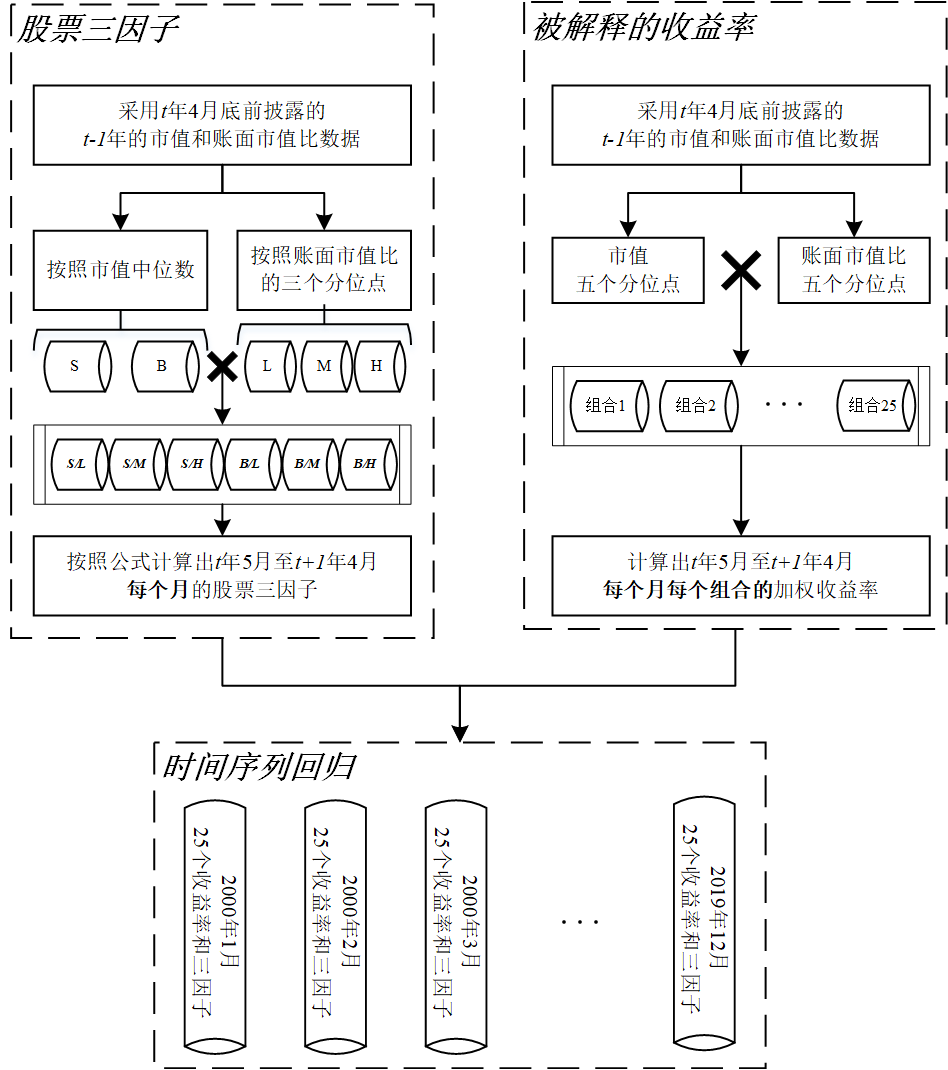
***B.被解释的收益率***

我们依据市值以及账面市值比划分25个投资组合，并将其超额收益率作为时间序列回归的被解释变量。之所以使用按市值和账面市值比构成投资组合，是因为我们要确定投资组合的和是否涵盖了与市值和账面市值比有关的股票收益率中的市场因子。

此处我们构建25个投资组合的方式很像前面说明的6个的投资组合。在每年5月，我们将A股所有股票按市值分和账面市值比分类，和都是由公司年4月底前披露的年相关数据测量的。我们依靠A股所有股票的5个市值的分位点以及5个账面市值比的分位点，进而建立了25个投资组合，并计算加权投资组合从年5月至年4月的月度收益率。这25个投资组合在2000年1月至2019年12月的月度超额收益率是股票在时间序列回归中的因变量。

***C.数据处理结果展示***

图2.展示了构建时间序列回归输入数据的流程，对上面提到的股票市场因子构建以及被解释收益率构建的过程做了总结。表A.则呈现了处理完成进行回归的数据结构。



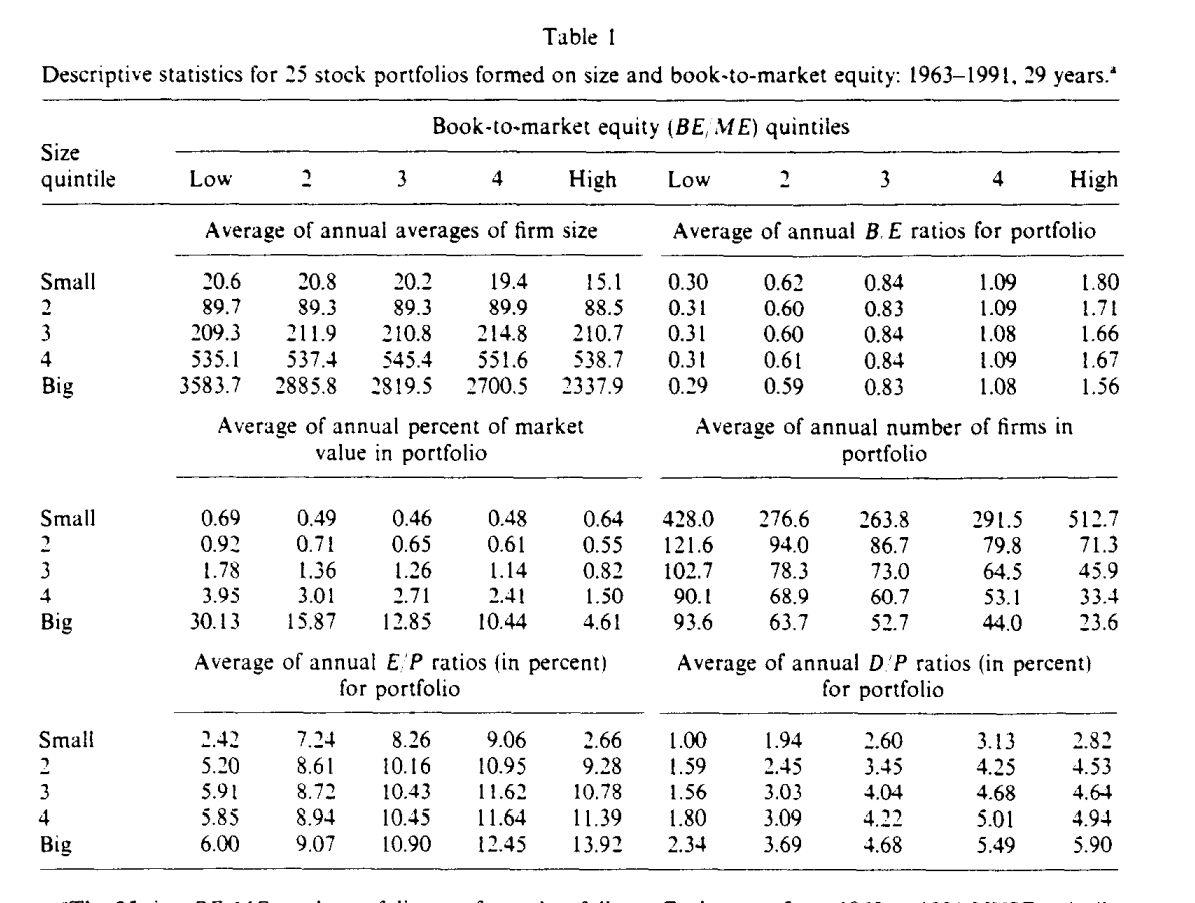
表A.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 组合 | 时间 | 被解释收益率 | 股票三因子 | | | | |
|  | |  | |  |
| 组合1 | 2000年1月 |  |  |  | |  | |
|  |  |  |  | |  | |
| 2000年12月 |  |  |  | |  | |
|  |  |  |  | |  | |
| 2019年1月 |  |  |  | |  | |
|  |  |  |  | |  | |
| 2019年12月 |  |  |  | |  | |
|  |  |  |  |  | |  | |
| 组合25 | 2000年1月 |  |  |  | |  | |
|  |  |  |  | |  | |
| 2000年12月 |  |  |  | |  | |
|  |  |  |  | |  | |
| 2019年1月 |  |  |  | |  | |
|  |  |  |  | |  | |
| 2019年12月 |  |  |  | |  | |

***D.数据的初步描述性统计***

表 Y.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Low | 2 | 3 | 4 | High |
|  | Average of annual averages of firm size | | | | |
| Small | 9.29e+08 | 9.02e+08 | 8.32e+08 | 7.68e+08 | 7.17e+08 |
| 2 | 1.48e+09 | 1.48e+09 | 1.47e+09 | 1.48e+09 | 1.46e+09 |
| 3 | 2.27e+09 | 2.28e+09 | 2.26e+09 | 2.25e+09 | 2.25e+09 |
| 4 | 3.76e+09 | 3.72e+09 | 3.74e+09 | 3.69e+09 | 3.73e+09 |
| Big | 1.38e+10 | 1.15e+10 | 1.57e+10 | 1.90e+10 | 2.04e+10 |
|  | Average of annual B.E ratios for portfolio | | | | |
| Small | 0.3388104 | 0.5817435 | 0.7905607 | 1.077802 | 2.239342 |
| 2 | 0.3462208 | 0.5744277 | 0.7865289 | 1.061606 | 1.94653 |
| 3 | 0.3468638 | 0.5774163 | 0.7847416 | 1.062772 | 1.994762 |
| 4 | 0.3526669 | 0.5748885 | 0.7788596 | 1.064191 | 2.105844 |
| Big | 0.3228064 | 0.5714194 | 0.7791914 | 1.069154 | 2.35344 |
|  | Average of annual percent of market value in portfolio | | | | |
| Small | 0.0058976 | 0.0081746 | 0.0086967 | 0.0099799 | 0.0100035 |
| 2 | 0.0122601 | 0.0153409 | 0.0166725 | 0.01553 | 0.0124677 |
| 3 | 0.0196512 | 0.0227636 | 0.0231727 | 0.0225617 | 0.0184021 |
| 4 | 0.0373463 | 0.0382316 | 0.0349546 | 0.0310797 | 0.0274931 |
| Big | 0.168753 | 0.0921243 | 0.1017382 | 0.0971186 | 0.1495857 |
|  | Average of annual number of firms in portfolio | | | | |
| Small | 30.05417 | 46.275 | 60.25417 | 90.30833 | 111.0792 |
| 2 | 60.9 | 70.3125 | 75.09583 | 72.40833 | 60.94167 |
| 3 | 67.6625 | 74.90417 | 76.0625 | 69.62917 | 52.9625 |
| 4 | 77.69167 | 82.86667 | 72.14583 | 60.97917 | 49.67083 |
| Big | 98.125 | 67.70417 | 59.20417 | 50.85417 | 69.24583 |



表Y.展示了我们使用A股所有股票交易数据断点来形成25个投资组合的基本统计信息。可以看出，我们的结果和Fama（1993）基于纽约证券交易所的断点来形成投资组合的描述性统计有所不同，股票的数量和市值大小以及账面市值比之间没有明显的关系，分布较为均匀。正如我们所想的一样，5个最大的投资组合的市值约为市场总价值的xx，并且最大市值和最低账面市值比的股票占据了25个组合总价值的16%以上。

表Z

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 加权组合收益率 | | | | | |
| 市值 | Low | 2 | 3 | 4 | High | High-Low |
| Small | 1.96147 | 2.06339 | 1.913168 | 1.841052 | 1.502822 | -0.459 |
| 2 | 1.437059 | 1.589977 | 1.573204 | 1.606666 | 1.27866 | -0.158 |
| 3 | 0.8527741 | 1.069181 | 1.382624 | 1.220779 | 1.13915 | 0.286 |
| 4 | 0.757337 | 0.8446231 | 1.004628 | 1.028897 | 1.009717 | 0.252 |
| Big | 0.494437 | 0.7579765 | 0.6871445 | 0.7847376 | 0.7893696 | 0.295 |
| Big-Small | -1.467 | -1.305 | -1.226 | -1.056 | -0.713 |  |

如表Z.所示，根据市值和账面市值比形成的25个股票组合产生了范围比较广的平均超额收益，从每月0.49%到2.06%。基于本表，我们可以得出与Fama(1992a)中部分相同的结论：一方面，五组中，所有的平均回报率都趋于从小到大的组合中减少，采用Newey-West t 统计量分别检验市值最高组和最低组的差异也是显著的，即规模和平均回报之间存在负相关。但是另一方面在每一个市值的五分位数中，采用Newey-West t 统计量分别检验账面市值比最高组和最低组的差异并不显著，甚至Small和1组中为负值，并且平均回报率也没有呈现“随着的增加而增加的趋势”，这一点我们将会在后面进行补充讨论。

**4.3 时间序列回归结果**

在时间序列回归中，斜率和值是不同风险因素是否捕捉到股票收益变化的直接证据。本文和Fama(1993)一样，通过散布来验证验证三因子在收益中的作用，分别是：

（a）只使用市场因子超额市场收益进行回归；

（b）使用市值因子和账面市值比因子作为解释变量的回归；

（c）使用三因子、和进行回归。

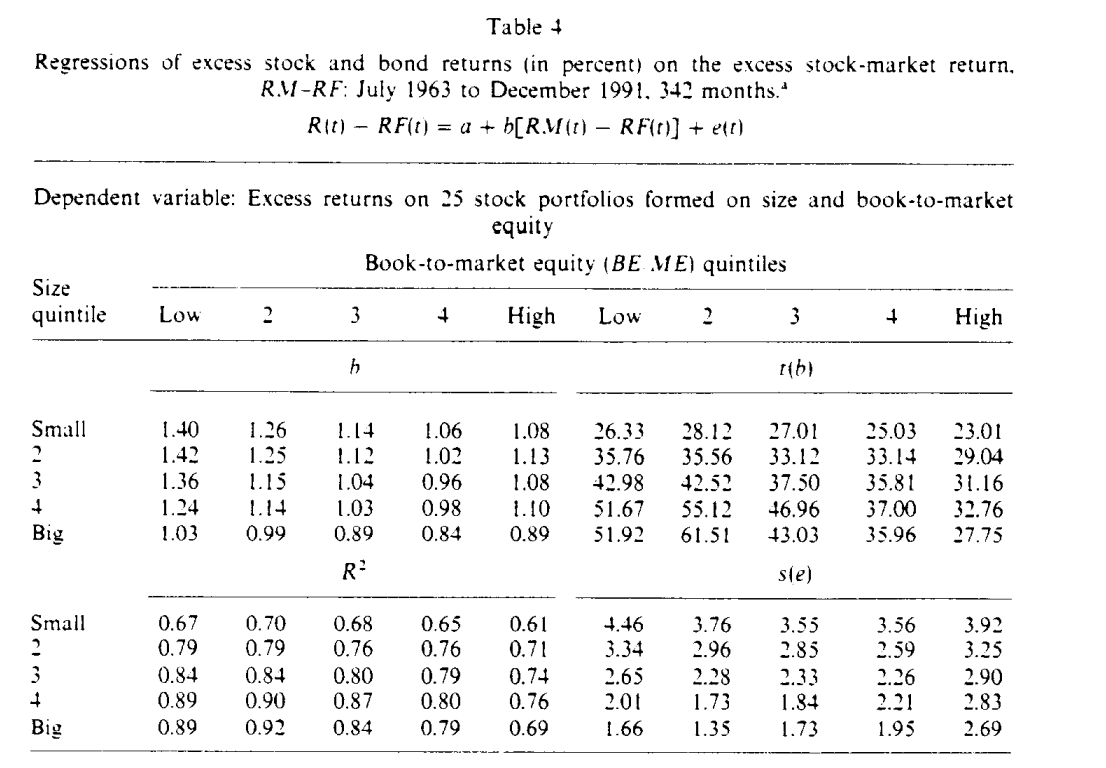
本节将报告上述三步的回归结果，分析三因子对股票市场超额收益的解释能力。

***A．市场因子***

表B.显示，股票市场组合的超额收益的确能捕捉到股票超额收益率的变化，这很正常，然而，更重要的是，市场因子在股票收益中留下了许多可能由其他因素解释的变化。25个组合回归的值中，只有4个较高市值的组合超过了0.9，而对于小市值股票组合，值大都在0.7至0.9之间，小市值和小账面市值比的组合甚至都低于了0.7。而正如我们接下来将要展示的那样，对于这些股票组合，市值和账面市值因素（和）将有更好的解释力。

表B.（依托Excel 制作）



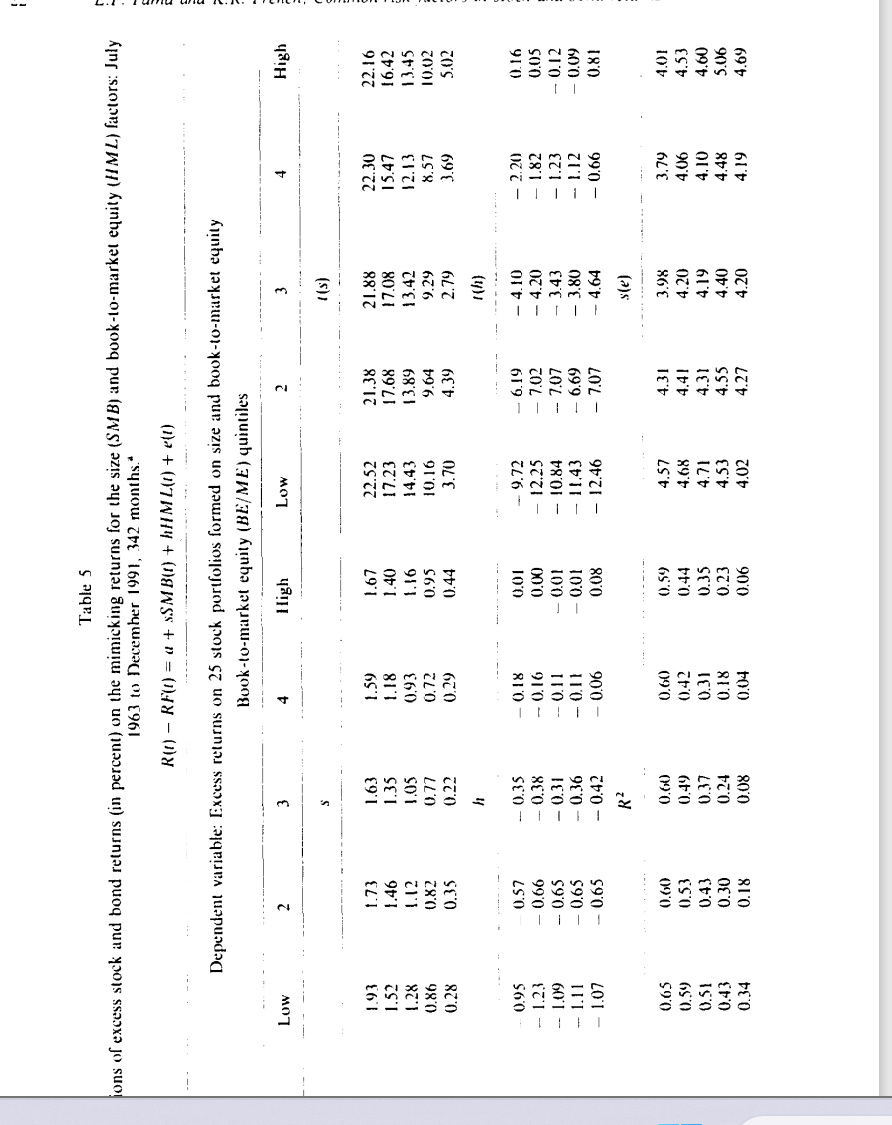


***B．SMB和HML***

表C.显示，在没有市场因子的情况下，和的确也能捕捉到股票收益的少量时间序列变化：25个值中有20个高于0.1，12个高于0.3。然而，特别是对于规模较大的五分之一的投资组合，和对股票收益的解释能力并不强，这刚好和表B形成了“互补”。同样对于表B中较低的几个组合表C中的明显较高。

表C.





***C．三因子***

表C显示，单独使用和对股票超额收益的解释力并不高，但似乎与市场因子形成互补。表D.显示把三因子全部包含在回归中的结果，这个表可以说是本部分最重要的一张表，令人欣慰的是，在我们的数据上，Fama-French三因子模型得到了较好的验证，这三个股票因子的确捕捉到了股票超额收益的变化。

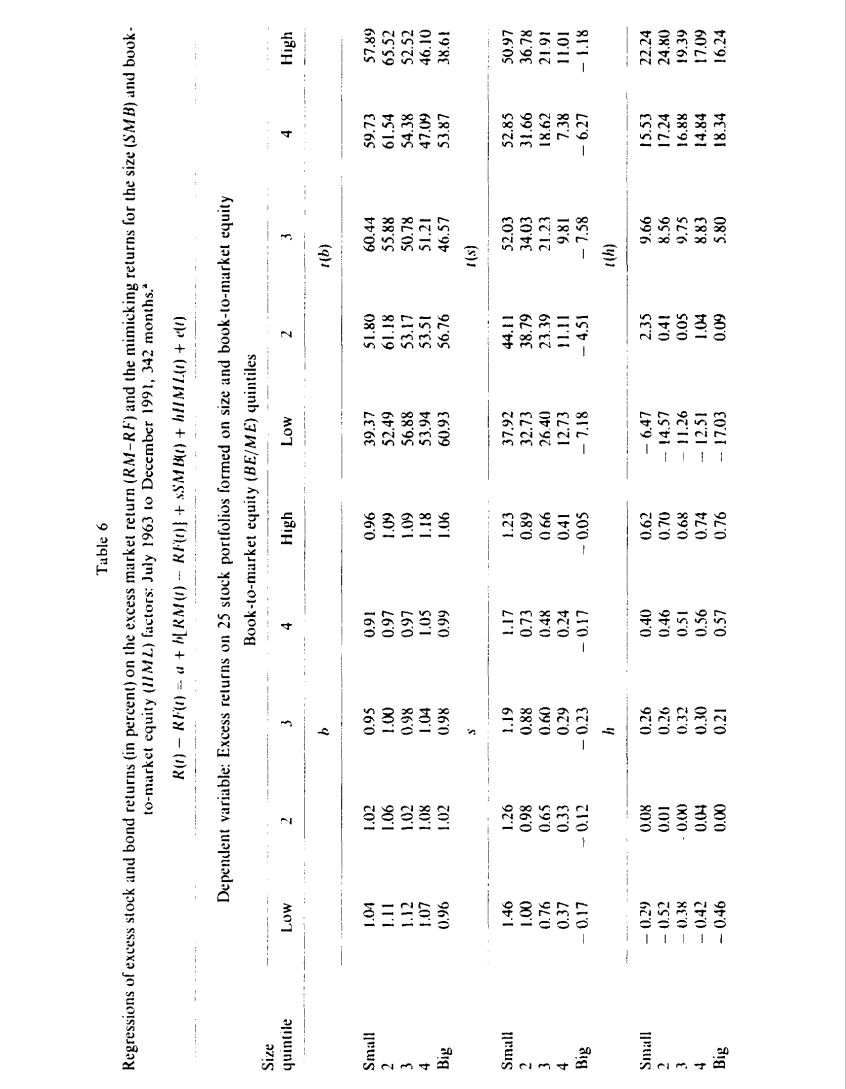
首先，因子斜率的统计量几乎都大于2，大多数都大于10。这是十分显著的，作为规模因子的模拟回报，明显捕捉到了市场因子和所忽略的股票回报的共同变化。此外，正如我们想的那样，的斜率与市值大小有关，在每一个账面市值比的五分位数中，的斜率从较小的五分位数到较大的五分位数单调地下降。

其次，对于的斜率，即模拟账面市值因素的回报，与账面市值比也有系统的关系。在每个市值的五分之一的股票中，的斜率单调地增加，从最低账面市值比五分之一的显著负值到最高五分之一的显著正值，而且几乎所有的值都在1%的显著性水平下是显著的。明显捕捉到了市场因子和因子所忽略的股票收益的共同变化，这与账面权益有关。

最后，在上述因子和因子的显著斜率下，三因子共同回归必然导致的大幅增加。表B中单独使用市场因子回归，只有四个组合的值大于0.9，而表D中值小于0.9则变成了“凤毛麟角”（25个中只有1个小于0.9）。对于规模最小的五分之一的五个投资组合，从表B的0.65到0.73增加成为了表D中的0.87和0.96之间。在表D.中即使是最小的（最小市值和最小账面市值比组合的0.88），也远远大于市场单独产生的0.65。

表 D.





当然我们也注意到了，在回归中加入和对股票的市场有一个有趣的影响。在表B的单因子回归中，处于最小市值和最低账面市值比五分位数的股票组合的是1.13，在另一个极端，最大市值和最高账面市值比五分位数的股票组合的单变量是0.94。而在表D的三因素回归中，这两个组合的分别是1.02和0.97，在回归中加入和会使股票的向1.0塌陷：低向1.0移动，高则向下移动。

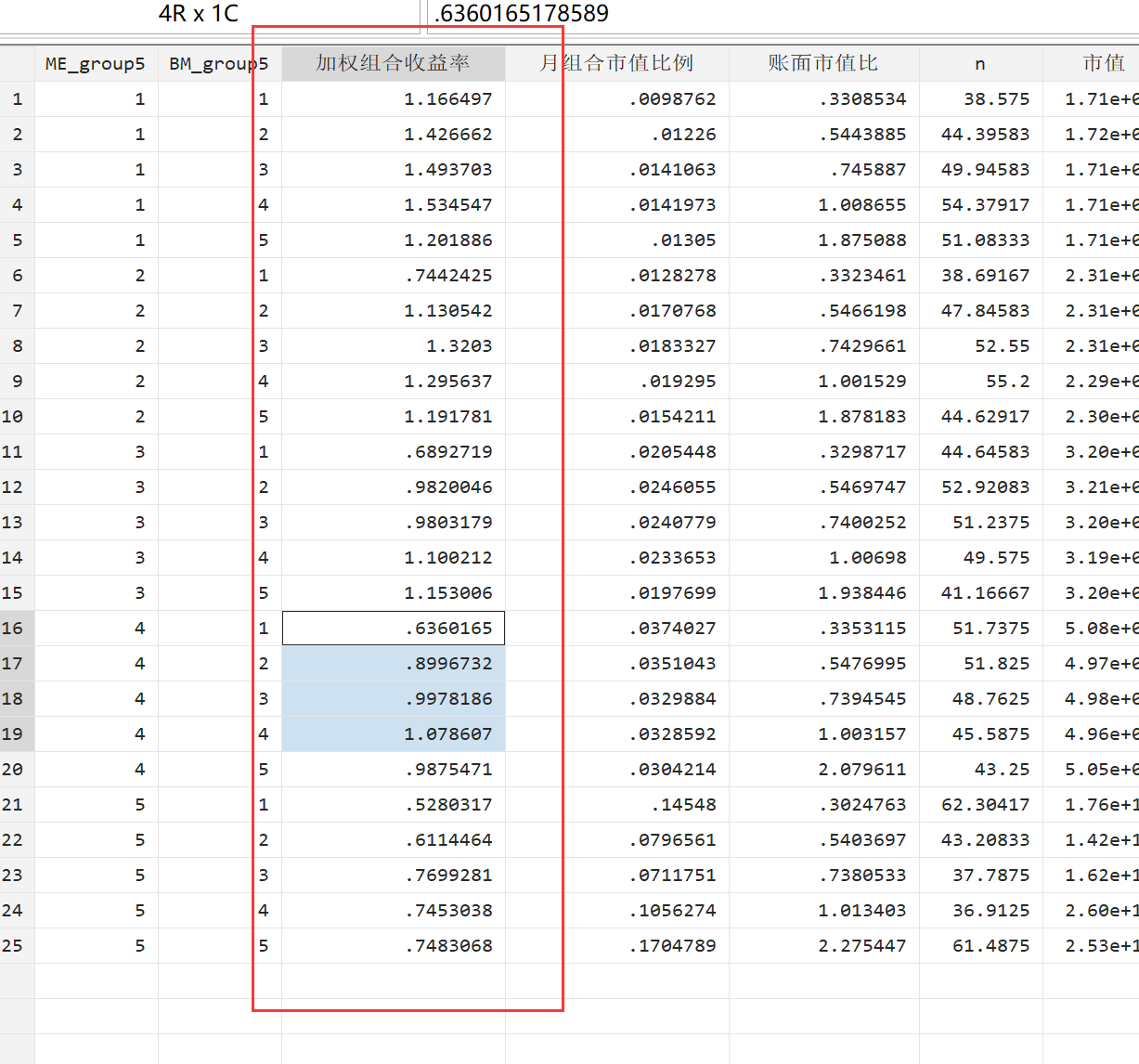
**4.4 对Fama-French三因子模型验证的结论**

在我们数据上对Fama-French三因子模型的验证的结果，几乎和Fama（1993）如出一辙，表B到D充分说明了模型中三因子的确可以充分地反映股票超额收益率的变化。但表Z中的部分异常结果，还是让本文更深一步地探讨了中国股市特有的IPO监管造成的壳价值问题。

# 五、Fama-French 三因子模型

表Z1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 加权组合收益率 | | | | | |
| 市值 | Low | 2 | 3 | 4 | High | High-Low |
| Small | 1.96147 | 2.06339 | 1.913168 | 1.841052 | 1.502822 | -0.459 |
| 2 | 1.437059 | 1.589977 | 1.573204 | 1.606666 | 1.27866 | -0.158 |
| 3 | 0.8527741 | 1.069181 | 1.382624 | 1.220779 | 1.13915 | 0.286 |
| 4 | 0.757337 | 0.8446231 | 1.004628 | 1.028897 | 1.009717 | 0.252 |
| Big | 0.494437 | 0.7579765 | 0.6871445 | 0.7847376 | 0.7893696 | 0.295 |
| Big-Small | -1.467 | -1.305 | -1.226 | -1.056 | -0.713 |  |



# 六、结论