

# 投资思想篇





# 第 1 章 量化投资在中国股票市场的应用

量化投资是一种基于经济、金融理论，借助严谨的实证研究方法，进行投资标的和投资时机选择的投资方法。量化投资在美国已经有 40 年的发展历史，并涌现出桥水（Bridge-water）、AQR 等以量化投资为主、管理资产超千亿美金的对冲基金。

量化投资是否适用于中国的股票市场？在谈到在中国股票市场做投资的时候，很多人就会摇头。大家常说，中国的股票市场不规范，靠内幕消息，股票价格由庄家操控。我们在高校的大学生可能会用专业一些的词，比如，我们的市场不是有效的。也常有人认为，我国的上市公司根本不具有投资价值，它们上市就是为了圈钱，根本不考虑给投资者回报。我们身边，也有亲戚、朋友，在股票市场上一而再、再而三地亏钱，甚至有些股票买了之后就一直跌，一两年的时间里能下跌百分之六七十之多。

此外，中国股票市场目前仍然有较强的做空限制，且股指期货目前交易受限。而且，对量化投资最核心的批评恐怕是：量化投资所基于的投资理论在中国是不适用的，我们的股票市场和美国的不一样。

但实际上，中国股票市场具有发挥量化投资功效的独到优势；中国股票市场的有效性为量化投资的适用奠定了良好的基础；量化投资在中国股票市场上大有可为。

## 1.1 在中国股票市场进行投资的独到优势

虽然中国股票市场有各种待完善的地方，但中国市场在应用量化投资方面有两大独到的优势。

第一，数据优势。中国上市公司的财务数据披露完整，数据库更新迅速。通过常用的金融数据库，我们能够很容易地获取上市公司的财务数据。并且，数据库提供的数据也包含上市公司高管、股权结构、董事会、股东交易等三大财务报表以外的数据，这些都为量化投资提供了丰富的数据基础。此外，中国金融数据库更新数据非常及时。比如，季报的截止时间是 10 月 31 日，我们在 11 月 1 日就可以从数据库中提取季报的数据，保证了数据分析的时效性。

丰富的数据是量化投资发挥功效的前提。缺乏数据的量化投资是缺乏生命力的，也正所谓“巧妇难为无米之炊”。相对债券、期货市场，股票市场的数据库丰富很多。比如，上市公司的财务报表就有三大张：资产负债表、利润表和现金流量表。每张表都有数十个项目，而且表与表之间也存在钩稽关系。丰富的数据，为量化投资在股票市场发挥巨大功效奠定了良好的基础。

第二，市场优势。量化投资的核心，在于利用模型从市场中提取并处理与未来收益率相关的信息，找到并未被其他投资者注意到的投资机会。做投资，其实是一个和市场上其他投资者博弈的过程。更快捷、有效的信息提取和处理是在股票市场上长期获利的基础。这就好

比打牌——你能否赢，很大程度上取决于你的对手有多弱。在美国、中国香港等发达的资本市场，主要的投资者是机构。比如，根据美国中央银行的数据，美国股票市场超过 80% 的交易量来自机构投资者。所以在美国股票市场，放眼望去，都是专业的机构投资者，具有极强的信息收集和处理能力，但凡有一丁点超额收益的机会，机构投资者就会立刻发觉并利用。然而在中国股票市场上，根据上海证券交易所（以下简称上交所）的数据，87% 的交易来自散户。所以我们只要比一般散户处理信息更有效率，获利的可能性就更大。因此，相比美国、中国香港等以机构投资者为主的市场，量化投资在中国市场上处理信息的优势更为明显，获利的概率也更大。

实际上，以散户为主的特点，使得当期的中国股票市场成了人类历史上，迄今为止最适合做投资的市场。美国在 20 世纪 70 年代完成了由以散户为主的市场向以机构为主的市场的转变。在此之前，虽然美国市场是以散户为主，但那个时候计算机技术的应用还处在萌芽之中，搜集信息难度之大，非今日信息时代的人们所能想象。中国目前有超过 3600 家 A 股上市公司，市值 50 万亿元。在现在这样一个信息时代，中国经济、政治稳定，股票市场规模体量如此之大，居然还是一个以散户为主的市场，可见它包含了非常多的可获利的机会，等待我们去发觉和利用。可以这么说，中国的股票市场就像一个宝库，即使是简单的量化投资技术，都能帮助我们在这个宝库中获得长期、丰厚的回报。

## 1.2 中国股票市场的有效性为量化投资的适用奠定了基础

量化投资的核心，在于提取与未来股票收益率相关的信息。这就需要市场满足两点：第一，市场是有效的，即与收益率、定价相关的信息终究会反映在价格当中。如果市场完全无效，价格不受信息的驱动，那么即使将一只股票方方面面的信息都掌握透彻，也无法对未来的收益率进行预测。第二，与股票定价相关的信息并不能够立刻、完全反映在价格当中，而是需要一定的时间。如果价格立刻反映信息，不给量化投资留任何分析、建仓的时间，那么量化投资也是没有用武之地的。也就是说，量化投资的适用，需要市场是有效的，但又不是那么有效，这样利用量化投资方法提取信息，才能对未来做出预测。

那么，中国股票市场的有效性如何？是否像许多人所说的那样，市场是无效的，价格并不反映价值呢？我们通过以下几组案例来分析。

### 1.2.1 上市公司股票价格是否体现价值

我们先来看第一个案例：我们希望找到满足这些条件的公司：（1）上市超过 10 年；（2）从 2003—2014 年，12 年时间里，它们的平均年度股票回报率，包括股利、回购等，超过 20%；（3）在这 12 年里，有 9 年的年度收益率超过 15%。市场上是否有满足这三个条件的公司呢？大家可以看到，要满足这三个条件是不容易的。我国股票市场，2003—2005 年是熊市，2005—2007 年是牛市，2008 年金融危机，2009 年因为 4 万亿人民币的刺激而上涨，但从 2010—2014 年，又跌跌不休。要在这个区间里，找到平均年化收益率超过 20%、平均年化收益率超过 15% 的年份超过 9 的公司，难度可想而知。表 1-1 给出了满足这三个条件的公司，共 10 只股票，它们的平均年化收益率如表中所示。

表 1-1 10 只股票组合列表

股票代码	股票名称	平均年化收益率（2003—2014 年）
000050.SZ	深天马 A	38%
000538.SZ	云南白药	31%
000550.SZ	江铃汽车	35%
002008.SZ	大族激光	55%
002038.SZ	双鹭药业	71%
600211.SH	西藏药业	22%
600420.SH	现代制药	35%
600501.SH	航天晨光	24%
600636.SH	三爱富	22%
600647.SH	同达创业	24%

接下来我们看看这 10 只股票的基本面情况，看看它们是否值得投资。表 1-2 列出了这个股票组合从 2005 年到 2013 年每年的平均现金流与资产的比例、股东权益回报率、盈利增长率，以及市盈率。

表 1-2 10 只股票组合的基本面情况

年份	现金流/资产	股东权益回报率	盈利增长率	市盈率
2005	8.33%	12.85%	16.48%	28
2006	7.42%	14.42%	22.32%	69
2007	7.56%	17.06%	14.78%	32
2008	3.73%	10.46%	-13.95%	37
2009	6.80%	12.90%	30.81%	46
2010	10.67%	18.37%	46.53%	44
2011	10.86%	23.86%	7.99%	20
2012	8.80%	17.86%	7.62%	33
2013	7.74%	16.17%	2.11%	30
平均	7.99%	15.99%	14.97%	38

从表 1-2 可以看出，这 10 只股票平均而言，每年都产生了正的，而且是大量的现金流，其平均现金流对资产的比例接近了 8%。这些公司的平均权益回报率（Return on Equity, ROE）达到了 16%，平均主营业务盈利增长率除了在 2008 年金融危机期间为负以外，其余年份全部为正。它们的市盈率平均为 38。这些数据表明，这 10 家上市公司在过去 9 年里极少发生亏损。它们通过经营产生了大量的现金流入，给股东带来了相当可观的经济收益。而与此同时，它们由市盈率反映的市场估值并不高。如果投资者从 2005 年 1 月开始持有这个组

合，一直到2014年7月，10年时间，累计收益率超过10倍，如图1-1所示。

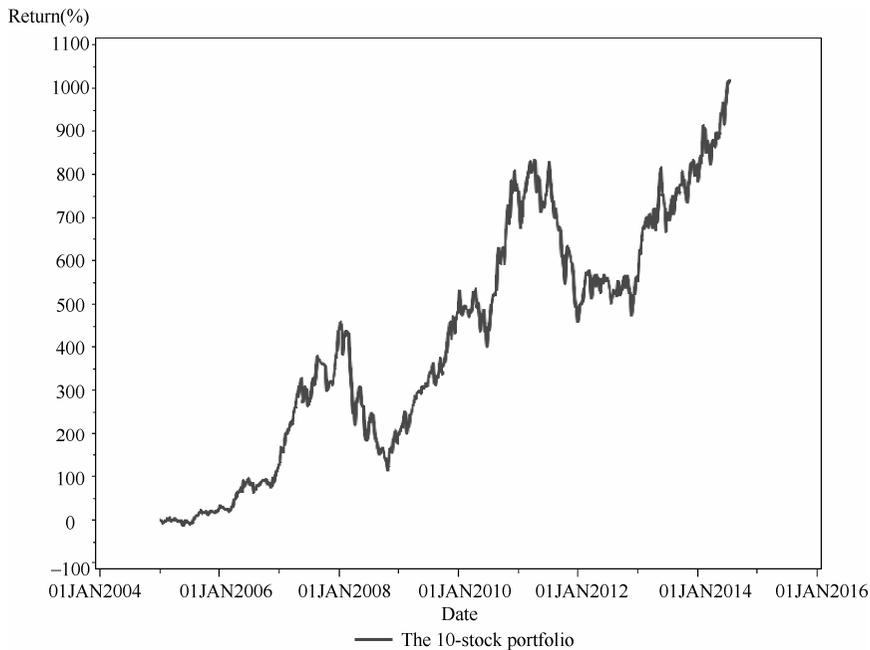


图 1-1 10 只股票组合的累计收益率（2005.1—2014.7）

从这10只股票的例子，我们可以看出，事实上，在中国的股票市场上是有值得投资的股票的，而且它们的价值反映在股票市场上，能够给投资者带来非常可观的回报。

第二个案例：我们来看满足另一组条件的公司。（1）2005—2013年，公司每年的主营业务利润都为正，（2）从2005年到现在，累计收益率超过20倍。这9只股票的组合列表如表1-3所示。

表 1-3 9 只股票组合列表

Code	Name	Cum. Return (2005.1~2014.7)
000651.SZ	格力电器	21.1
000712.SZ	锦龙股份	20.2
000826.SZ	桑德环境	22.1
002038.SZ	双鹭药业	22.6
600111.SH	包钢稀土	36.0
600315.SH	上海家化	25.5
600517.SH	置信电气	20.4
600587.SH	新华医疗	22.7
600804.SH	鹏博士	29.9

这9只股票，收益率最高的是包钢稀土，达到36倍，最低的是锦龙股份，也有20.2倍。那么，这9只股票的业绩如何？我们一起来看表1-4。这张表反映的是2005年至2013年公

司的营业收入、主营业务利润、现金流和资产的增长幅度。

表 1-4 9 只股票组合的公司基本面增长幅度（2005—2013 年）

股票代码	股票名称	营业收入增长	主营业务利润增长	现金流增长	资产增长
000651.SZ	格力电器	5.6	18.4	20.5	9.5
000712.SZ	锦龙股份	-0.3	6.6	24.2	5.0
000826.SZ	桑德环境	19.7	21.2	21.1	12.3
002038.SZ	双鹭药业	8.8	16.0	16.0	5.7
600111.SH	包钢稀土	8.6	69.5	111.3	8.1
600315.SH	上海家化	1.3	15.8	15.3	1.5
600517.SH	置信电气	9.4	12.5	18.8	9.0
600587.SH	新华医疗	9.1	15.0	17.7	6.0
600804.SH	鹏博士	8.4	297.2	233.8	19.6
Average		7.8	52.5	53.2	8.5

从表 1-4 可以看出，这 9 家上市公司，从 2005—2013 年，平均而言，营业收入增长了 7.8 倍，主营业务利润增长了 52.5 倍，现金流增长了 53.2 倍，资产增长了 8.5 倍。这些都是非常优秀的业绩。在过去长达 9 年的时间里，这些优秀的上市公司在收入、利润、现金流和资产上都取得了非常高的增长。它们的业绩成长也反映在了股票价格当中。如果从 2005 年就持有这 9 只股票所构成的组合，累计收益率如图 1-2 所示，截止到 2014 年 7 月，累计收益会超过 30 倍。



图 1-2 9 只股票组合的累计收益率（2005.1—2014.7）

上述两组例子，给我们的是同一个信息，那就是中国市场上有好公司，有值得投资的公司，并且，这些优秀公司的价值在中国的股票市场上的确是得以体现了。

## 1.2.2 上市公司的股票价格是否反映信息

我们再通过一项研究结果，了解中国股票市场的有效性，探寻上市公司的股票价格是否反映信息。这项研究，是关于上市公司在大股东买卖这家公司的股票之后的价格走势情况的。中国从 2007 年开始，也就是股权分置改革之后，有比较多的大股东交易。大股东通常是指持有一家公司股票比例超过 5% 的股东。这些大股东主要是公司的控股公司、高管，也包含机构投资者和个人投资者。这些大股东因为持股比例较高，所以对这家公司内部运营的了解会比小股东更多，他们掌握的信息更完全。如果一家公司，它的大股东不仅继续持有，而且大量买入了这家公司的股票，这通常说明大股东对这家公司非常有信心，认为这家公司的股票未来会上涨，这就传递出提升股价的信号。如果市场是有效的，那么我们应该看到这家公司在大股东买入股票之后的一段时间里，其股票价格会出现显著的提升。反之，如果一家公司的大股东选择抛售这家公司的股票，通常这对公司价值而言是负面的消息，其股票价格在未来一段时间里会出现下跌。根据这个逻辑，我们可以统计每周发生大股东交易的公司其未来股票的价格波动情况。研究发现，正如所预料的，那些大股东买入的公司股票在未来至少 9 个月的时间里都持续上涨，而那些被大股东抛售的公司股票则持续下跌。并且，依据这个规律，我们可以比较两个不同的交易策略，每周分别购买上周公布有大股东净买入和净卖出的股票，并且持有一周。这两个交易策略所得到的累计投资收益如图 1-3 所示。从图 1-3 我们可以看出，如果我们持续购买大股东买入的股票，累计收益率从 2007 年开始就一直非常稳健地向上，到 2013 年年底达到 150% 左右。而如果我们一直买入大股东卖出的股票，收益率是一直向下的，到 2012 年，用 6 年的时间就会亏完所有本金。

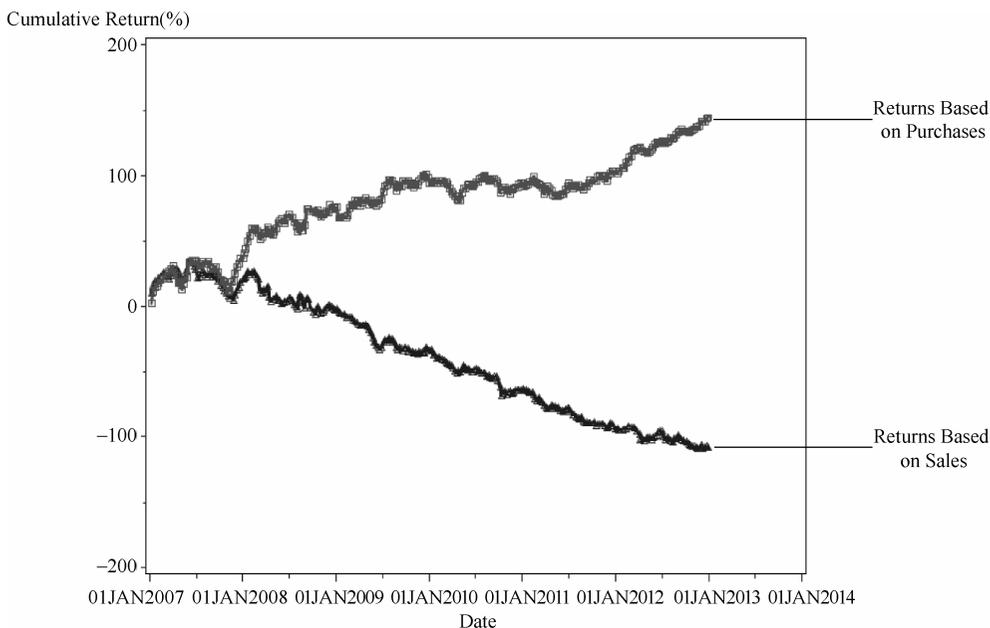


图 1-3 基于大股东交易的投资策略累计收益率

所以大家通过这个例子可以看到，我们的市场价格是反映价值，反映信息的。但从图 1-3 中又可以看出，这些信息并不是立刻就反映在股票价格上，而是在持续的一段时间里都影响着价格的走势。可见，中国股票市场虽然有效，但又不是那么有效，因此，中国股票市场的有效性为量化投资的适用奠定了良好的基础。

## 1.3 量化投资在中国股票市场上大有可为

### 1.3.1 基于基本面信息的投资策略

下面就利用一些例子说明量化投资在中国股票市场的应用。首先看基于公司基本面数据的投资策略。在股票定价理论中，有一个经典模型，称为股利贴现模型，在一系列假设下，股票价格可以通过下列公式计算：

$$\text{股票价格} = \frac{\text{股利}}{\text{贴现率} - \text{增长率}}$$

这个模型是一个极其简化的股票估值模型。但它为我们进行公司价值分析提供了一个指南。什么样的公司能够支付股利？自然是盈利高，有充足的现金流的公司。而且公司价值与增长成正比，所以利润增长快的公司其价值也会比较高。此外，我们希望公司的市场估值并不是很高，这样我们投资的安全边际就会比较大。因此，我们可以选取四项指标：现金流/资产、权益回报率、主营业务利润增长率和市盈率。我们研究这样一个问题：如果投资者只根据单个指标进行选股，能否获得可观的绝对收益？对于单个指标，选取的条件是：现金流/资产 > 10%，权益回报率 > 15%，主营业务利润增长率 > 20%，市盈率 < 20。表 1-5 给出了每年满足单个条件和所有条件的股票数量。事实上，每年我们的市场上都有很多公司满足这些条件。

表 1-5 每年满足基本面选股条件的股票数量

年份	现金流/资产 > 10%	权益回报率 > 15%	主营业务利润增长率 > 20%	市盈率 < 20	所有条件都满足
2005	313	238	410	236	38
2006	364	353	562	45	11
2007	366	562	761	341	56
2008	385	385	386	140	30
2009	533	503	666	227	47
2010	430	618	846	275	49
2011	342	602	834	575	51
2012	480	509	624	607	51
2013	425	558	894	521	60
2014	481	566	931	125	13

假设我们从 2005 年开始，每年 6 月底买入满足条件的股票，并持有 1 年，计算到 2015 年

的累计收益率，并和大盘表现进行比较。图 1-4 给出了过去 10 年时间里的累计收益率曲线。

图 1-4 表明，从 2005 年 1 月到 2015 年 1 月，如果投资者每年都坚持用这四个指标进行选股、交易，累计收益率会达到 8 倍至 10 倍。而同期，我们的大盘只涨了不到两倍。

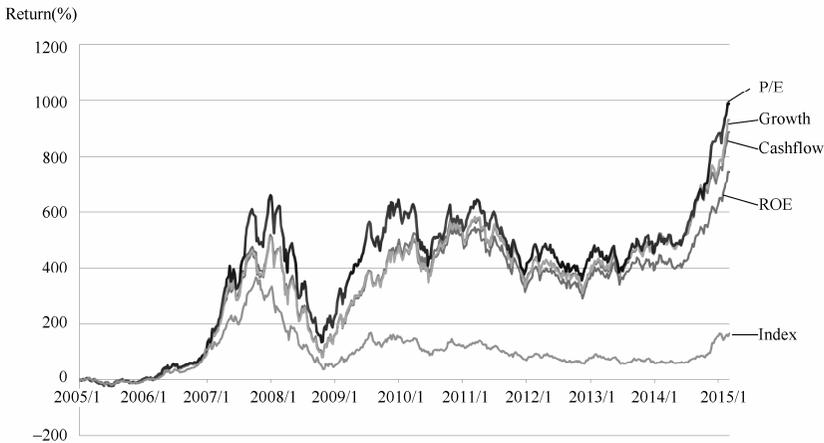


图 1-4 利用基本面指标选股交易的累计收益率曲线（2005—2015 年）

### 1.3.2 基于技术指标的量化投资策略

我们接下来考察一个基于技术指标的量化投资策略。相对强弱指标（Relative Strength Index, RSI），反映的是价格反转的可能性。

$$RSI = 100 - \frac{100}{1 + RS}$$

相对强度（RS 是股票价格在一段时期内，平均上涨幅度与平均下跌幅度之比）以 5 日 RSI 为例。如果 5 天里价格全部上涨，则 RSI=100。如果 5 天里价格全部下跌，则 RSI=0。如果价格的上涨等于价格的下跌，即这周没涨没跌，那么 RSI=50。价格上涨越多，RSI 值越高；价格下跌越多，RSI 值越低。当股票上涨过多时，会有回调的可能；当股票下跌过多时，可能会反弹。RSI 就是用来衡量股票价格反转的指标。RSI 过高时，比如高于 80，股票下跌的可能性较大。RSI 过低时，比如低于 20，股票上涨的概率较大，所以 RSI 与未来股票收益率成反比。

根据 RSI 的金融学含义，我们可以按 RSI 建立投资策略。图 1-5 将每周市场上所有股票按 RSI 分成 10 组，分别计算每组的未来周度平均收益率，如图 1-5 所示。

从图 1-5 中可以看出，除了刚开始 3 组有上下起伏的状况，从第 4 组开始到第 10 组，收益率单调下降。且 RSI 成功地区分出了未来收益率下跌的组合和上涨的组合。

图 1-6 是按照平均收益率最高的那一组建立的投资组合，计算其累计收益率。从图中可以看出，从 2009 年到 2014 年，该策略获利 180%，而同期上海证券综合指数（以下简称“上证综指”）涨幅不到 20%。

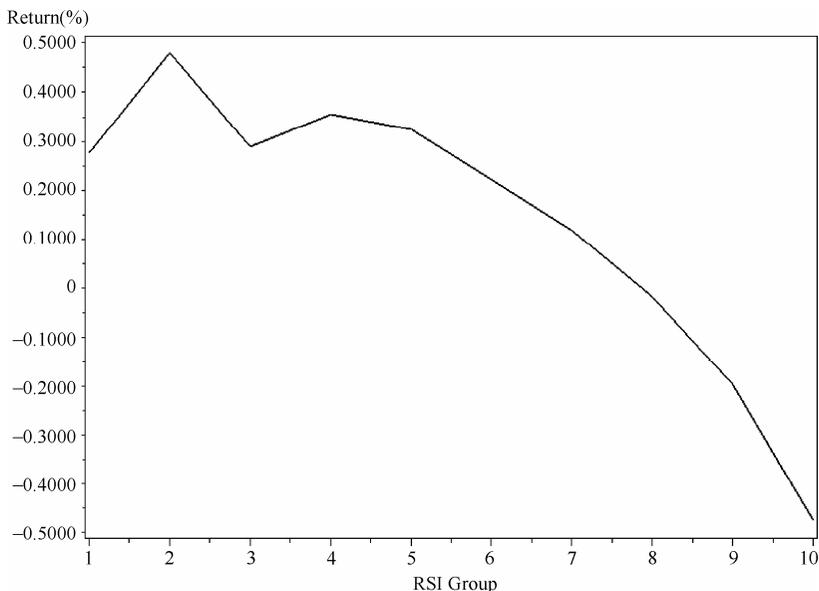


图 1-5 所有股票按 RSI 分 10 组获得的未来周度平均收益率

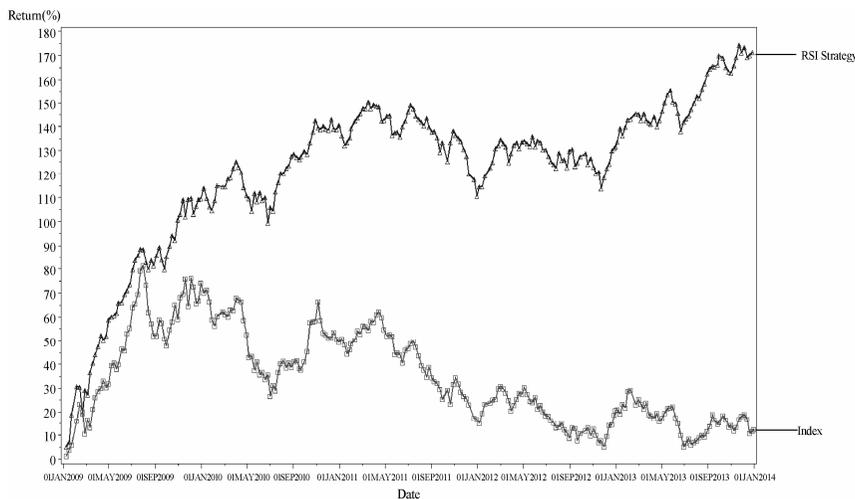


图 1-6 按照 RSI 建立的量化投资组合的累计收益率

### 1.3.3 基本面指标与技术指标的结合

上述两个策略，分别利用的是基本面指标和技术面指标。那么，我们可否综合利用这两类指标，提升策略的表现？答案是肯定的。

在以 RSI 建立的策略中，我们可以另外加上基本面指标。比如，我们要求公司的 RSI 不仅要低，而且该公司需要有正的利润增长、正的现金流和较低的估值。我们得到的新的策略，如图 1-7 所示。

在图 1-7 中，红线代表的是仅用 RSI 构建的投资策略，累计收益率为 300%。蓝线是将基本面指标与 RSI 结合构建的投资策略，累计收益率为 500%。

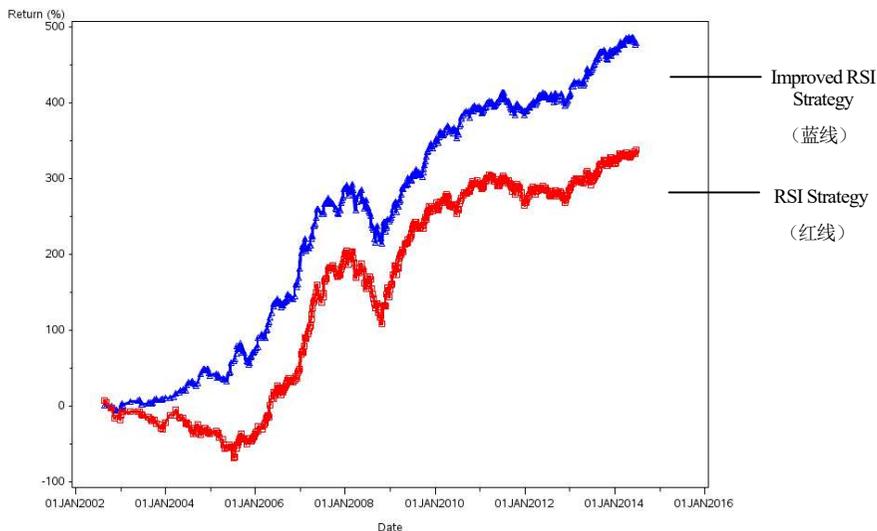


图 1-7 将基本面指标与 RSI 指标结合的量化投资策略获得的累计收益率

## 1.4 从量化投资看坚持理性投资的重要性

之前所介绍的是基于严谨的量化投资分析做基于基本面分析的理性投资，可见即便是在中国市场上，理性投资也能获得绝对收益。那么，为什么我们需要强调对理性投资的坚持？原因在于，市场是波动的，而这样的波动会动摇我们对投资的信心，使我们错过市场长期的回馈。

我们再回过头去看图 1-4。根据基本面选股，累计投资收益率是 10 倍。但作为投资者，你能做到从 2005 年一以贯之地坚持下去吗？在 2008 年，如果你坚持使用这个交易策略，你的亏损会超过 70%。站在 2008 年的 11 月，你会怎么想？你会不会质疑中国的股票市场？你会不会质疑自己？你会不会质疑你的策略？此外，从 2010 年到 2014 年下半年，将近 5 年时间，你都是亏损的。5 年的时间，你没有通过这个策略获得收益，你看到自己的账户一直是亏损的状态。坚持 1 年，容易，2 年，也行，3 年，勉强，但这可是 5 年！站在 2014 年的 11 月，当你回顾自己这 5 年的时候，你是否还会相信股票市场？你是否还会相信价值投资？

但越是艰难，越能体现坚持的宝贵。在 2008 年 11 月，当中国政府宣布 4 万亿刺激计划的时候，股票市场立刻止跌反弹，并开始了一波波澜壮阔的上涨。用不到 1 年的时间，你就能把 2008 年亏损的都赚回来。时间来到了 2014 年 11 月。在长达 5 年的亏损之后，你只要再多坚持一两周，就等到了中国股票市场上最宏大的一次牛市，从 2014 年 11 月到 2015 年 6 月，你会盈利 4 倍。如果你从 2005 年到 2015 年，坚持用这个策略投资，虽然经历了 2008 年的大跌和 2010 年到 2014 年的熊市，但在 2015 年，你的累计收益率依然有 10 倍。然而，如果投资者在 2008 年或 2014 年就离开了股票市场，这些收益就都成了空中楼阁。所以，理性投资，贵在坚持。

反过来，如果投资者不坚持理性投资，而是盲从，听消息，追逐市场热点，那么收益会是怎样的情况？通常，市场热点都是那些在一段时间内收益率高、吸引众多投资者去买的股

票。所以，我们可以模拟这样一个投资策略：投资者每周买入上周收益率和换手率都在前 5% 的股票，并持有一周，如此反复，从 2010 年开始至 2014 年的累计收益率，如图 1-8 所示。

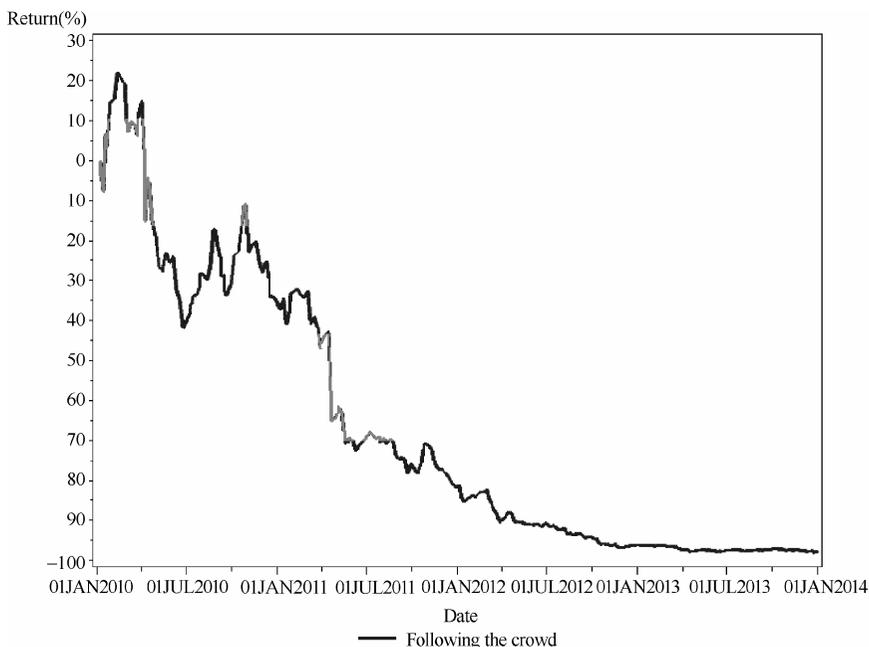


图 1-8 追逐市场热点获得的累计收益率

图 1-8 刻画了追逐市场热点获得的累计收益率。如果投资者从 2010 年开始，只用不到 2 年的时间，就会亏完所有的本金。而且，并不只是从 2010 年开始是这样的情形，无论从 2005 年、2008 年或 2009 年开始计算累计收益率，通通不超过 3 年，投资者一定会亏得一干二净。而这也解释了为什么有那么多人市场上亏钱，而且亏了很多。所以，请务必坚持理性的投资，坚持严谨的分析，这样才有可能在中国股票市场上获取正的绝对收益。

## 总结

尽管中国股票市场尚待完善，且做空机制受限，但由于中国股票市场交易以散户为主，且数据丰富、更新及时，市场价格反映价值与信息，市场有效性支持量化投资在中国股票市场发挥显著的功效。文中的几组案例也表明，即使是非常简单的量化投资策略，在我国股票市场上也能获得较高的收益，从而印证了量化投资在中国股票市场的适用性，也强调了理性投资的重要性。

## 第 2 章 如何分析宏观经济

做投资，就必须懂得分析宏观经济运行情况。谈到分析宏观经济运行，大家可能会觉得这是知名经济学家、专业的宏观经济分析师才能去做的事情。其实不然，用一个简单有效的框架，通过对核心指标进行分析，你也能了解宏观经济的运行情况，并指导投资决策。

### 2.1 为什么要做宏观经济分析

股票市场，其实与宏观经济的相关性是很大的。而个股的收益受市场的影响也非常大。

我们来看图 2-1。这张图是将每个季度的 GDP 增长率和当季的上证综指收益率放在一起进行比较的。

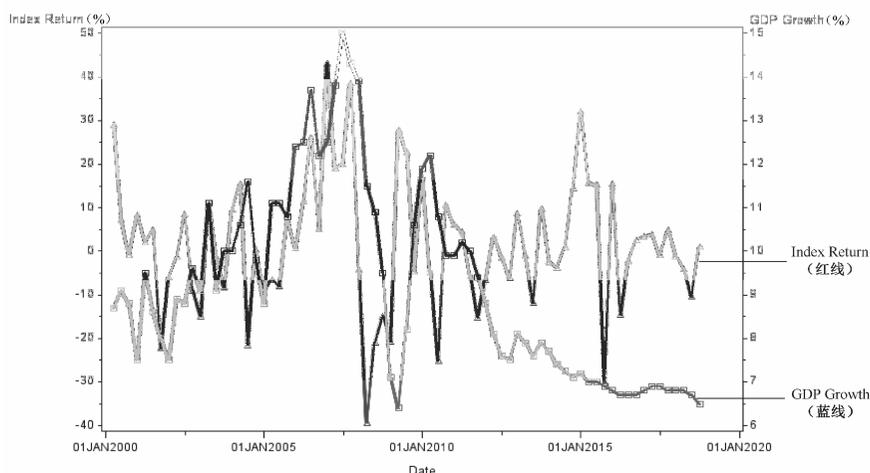


图 2-1 2000—2018 年上证综指收益率与 GDP 增长率对比

从图 2-1 中我们可以看出，GDP 增长率与上证综指收益率具有一定的关联度，而且有些时候基本上是同涨同跌的。

与此同时，我们来看图 2-2。图 2-2 反映的是平均来看个股收益率的波动有多大比例来自于市场的波动。我们看到这个比例最高的接近 60%，中位数的水平大致是 40%。因此，宏观经济影响市场，市场影响个股，这也是为什么我们在做投资的时候，需要非常注重宏观经济分析。

从本质上看，我们做宏观经济分析是为了在投资中进行择时。我们希望通过宏观经济分析，明确当前是否适合进行股票投资，通过择时，来控制我们的投资风险。

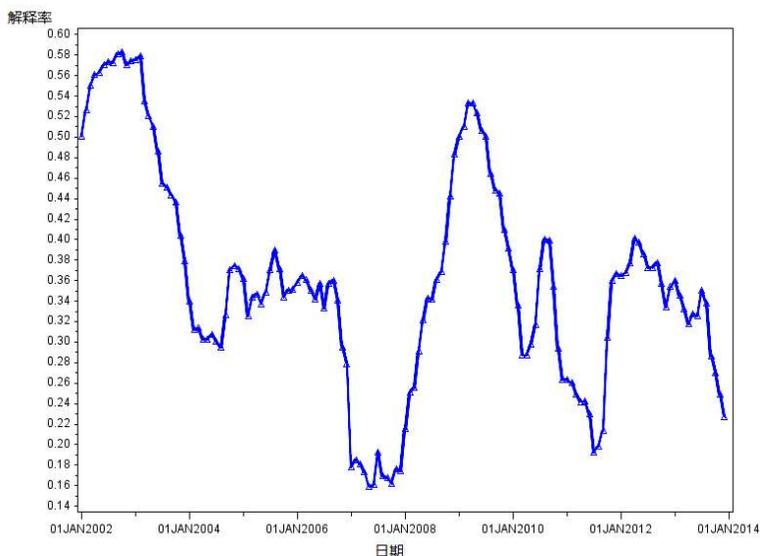


图 2-2 市场波动对个股收益波动的平均解释率

## 2.2 宏观经济分析的框架

宏观经济分析的框架，从国民经济的核算公式开始：

$$\text{GDP} = \text{消费} + \text{投资} + \text{净出口} + \text{政府开支}$$

这个公式表明了不同指标的作用。国内生产总值（GDP）作为经济总量指标，反映经济运行的总体情况。消费、投资、净出口是经济的三驾马车，是经济总量的三大构成部分，是经济基本面的分析要素。政府开支是政府对宏观经济施加影响的一个方面。考虑到政府对宏观经济的影响还有政策、政治稳定等，我们需要专门设一个部分，分析政府的影响。此外，由于我们是为了股票投资而进行宏观分析的，所以还需要包含对资本市场总体的分析。概括来说，为投资而进行的宏观经济分析，分为三个部分：第一，经济基本面分析，包括经济总量、消费、投资以及净出口；第二，政府政策与政治影响，包括经济政策和政治与社会稳定；第三，金融市场的分析，包括市场改革、流动性和投资者情绪。接下来我们逐一介绍各部分的内容。

## 2.3 经济基本面分析

经济基本面分为四部分，即经济总量、消费、投资以及净出口。

### 2.3.1 经济总量

经济总量，我们关注两大指标：GDP 增长率和采购经理指数（PMI）。

## 1. GDP

GDP 是衡量一国经济总体发展水平最直接，也是最重要的指标。

图 2-3 给出的是自 2000 年第一季度至 2018 年第三季度的 GDP 增长率。从图中可以看出，我国的 GDP 增长率在 2008 年金融危机前后跌宕起伏。首先是受金融危机影响，GDP 增长率从 2007 年的 15% 断崖式地下降至 2008 年的 6.5%，之后在 4 万亿元人民币刺激计划的拉动下，上升到 2010 年的 12%。但在 2010 年之后，我国的 GDP 增长率呈现显著下降的趋势。

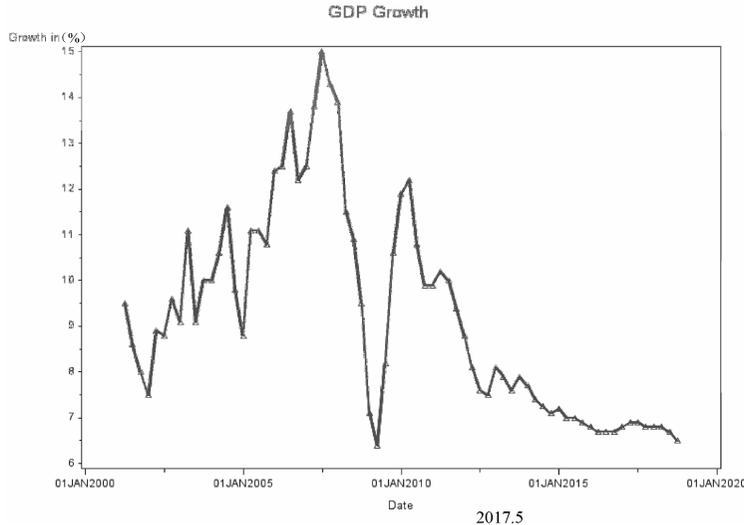


图 2-3 2000—2018 年中国 GDP 增长率

细心的读者会发现，增长率的下降趋势是放缓的。2010—2012 年，GDP 增长率从 12% 下降到了 7.5%，平均每年下降 1.5%。这是中国经济过去 10 年中最困难的时期。但从 2013 年到 2015 年年底，3 年时间，GDP 增长率只下降了 1%。而且，更让人欣喜的是，从 2016 年开始，GDP 增长率出现了止跌企稳，甚至小幅上升的趋势。但从 2018 年开始，GDP 增长率就呈现出加速下降的趋势，由一季度的 6.8% 降到二季度的 6.7%，在第三季度加速下降低至 6.5%，经济形势非常严峻。

我们再来看图 2-1 中展示的上证综指收益率与 GDP 增长率的对比。以 2008 年为分水岭，我们可以看到在 2008 年之前，上证综指收益率与 GDP 增速基本具有相同的波动态势。在 2008 年，受金融危机的影响，股指大幅回落。2009 年，GDP 增长和股票市场收益都大幅提升。2012 年以前，GDP 增长率和股票市场的表现基本是一致的。但从 2012 年开始，GDP 增长一路下行，而股票市场却不断走高，两者在 2015 年第二季度背离度达到最大。

从统计意义上来说，我们也可以关注 GDP 增长率与上证综指收益率之间的相关系数，由此检测两者之间是否存在显著的正相关性。计算可得，两个指标同期数据的相关系数为 0.17，具有一定的正相关性。但由于核算和数据公布的滞后性，我们在当季，只能使用上一季度的 GDP 数据。通过计算滞后一期（即前一期）的 GDP 增长率与当期上证综指收益率之间的相关系数，可得这一数值为 0.09，只有同期相关系数的一半。因此，为了在投资中运用更为准确的数据，金融机构聘请了大量的宏观经济分析师、经济学家去预测 GDP 增长率。

## 2. 采购经理指数 (PMI)

尽管 GDP 是衡量宏观经济运行最直接的指标, 但 GDP 数据每季度只更新一次。作为补充, PMI 也是一个重要的衡量宏观经济整体运行情况的指标。相对于 GDP, PMI 具有频率更高、快速简便、及时综合、预测性强等优点, 是一个较为可靠的先行指标。

PMI 的中文含义为采购经理指数。它通过对样本企业采购经理发放月度调查问卷采集数据, 分别询问他们关于订单、生产、雇员、配送和存货五个方面的商业活动的现实情况。被调查者只需要对每一个问题做出定性的回答, 即通过与上月进行比较, 选择较上月提升、不变或者恶化三者之一的答案即可。然后, 指数的编制者根据问卷结果计算单项 PMI 指标, 具体计算公式为:

$$PMI=1 \times P_1+0.5 \times P_2-0 \times P_3$$

式中,  $P_1$  代表认为有所提升的答案所占的百分比;  $P_2$  代表认为保持不变的答案所占的百分比;  $P_3$  代表认为出现下降的答案所占的百分比。

单项 PMI 指标能够直观地显示企业某一项生产活动的现状, 指标越高, 表明企业对经济运行状况越乐观。而综合 PMI 指标是对单项 PMI 指标的加权平均, 用于衡量企业供应和采购各个方面的总体发展趋势。综合 PMI 指标也有制造业和服务业之分, 而制造业综合 PMI 指标编制更早, 是刻画经济景气指数的重要指标。

从上述公式我们也可以看到为什么说 50 为 PMI 的荣枯线。如果所有企业的采购经理都认为经济有所好转, 则  $PMI=100$ ; 如果所有人都认为经济下滑, 则  $PMI=0$ ; 如果各占一半, 则  $PMI=50$ 。因此, 如果 PMI 大于 50, 则预示着更多人认为当下经济正在向好发展; 反之, 如果 PMI 小于 50, 则预示着经济发展前景不容乐观。

目前, 我国每月的制造业 PMI 指数都有两个版本, 分别是由国家统计局发布的官方 PMI 和由财新发布的财新 PMI (原汇丰 PMI)。两者的区别主要在于统计范围, 官方 PMI 覆盖了大中小型企业, 而在财新 PMI 的统计范围中, 中小企业占比更高。从历史数据来看, 两者大体同步, 偶有背离。

与 GDP 的分析方法类似, 我们可以通过折线图来比较官方制造业 PMI 指数与上证综指收益率之间的联动关系, 如图 2-4 所示, 我们仍能够从图中发现明显的同期正相关性。统计上, 两者相关系数为 0.150。除了同期的正相关性, 为了衡量 PMI 指数的预测能力, 通过计算滞后一期 PMI 指数与上证综指收益率之间的相关系数 (0.071), 我们可以看到单独的 PMI 指数的预测力仍旧不佳。

此外, 我们用同样的方法计算财新 PMI 与上证综指收益率的相关性。结果显示, 同期的相关性只有 0.044, 而滞后一期的相关性甚至变成了 -0.064。这说明财新 PMI 与上证综指收益率的相关性, 不如官方的 PMI 与上证综指收益率的相关性强。

除此之外, 我们还可以根据企业规模的大小, 把 PMI 指数再细分成大中小型企业。图 2-5 即展现了不同规模企业的 PMI 指数。从图中可以明显地看到在整个样本期间, 不同规模企业的 PMI 指数差异较明显, 从大型企业到小型企业, PMI 指数水平逐渐下降, 波动率则逐渐上升。小型企业与大型企业在市场地位、管理效率、议价能力和规模效应等方面都存在显著差异, 经济波动对其影响更大, 在样本期内基本处于荣枯线之下。这也印证了小企业在实际经营中, 融资难、生存不易等现实问题。但小企业的经营困难也与其企业发展阶段有关。企业初创之时存在对内团队磨合、产品开发, 对外开拓市场的艰难过程, 绝大多数小企业处于存

亡的边缘，其发展阶段很大程度上影响了它们对经营状况的评估。

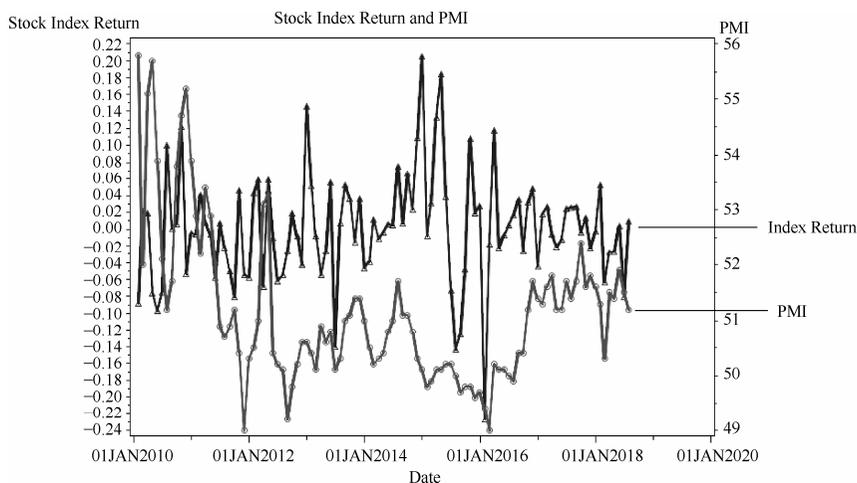


图 2-4 上证综指收益率与制造业 PMI 指数之间的联动关系

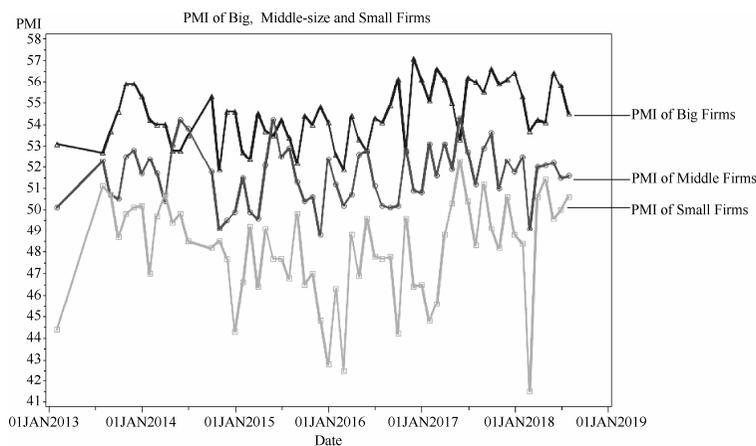


图 2-5 大中小型企业 PMI 指数

## 2.3.2 消费

接下来我们将逐一分析拉动经济增长的三驾马车。针对消费，可以从两方面考察，一方面是物价水平，另一方面则是社会消费总量的变化。

### 1. 物价水平

物价水平可以用商品价格指数表示，具体可分为消费者价格指数（Consumer Price Index, CPI）和生产者价格指数（Producer Price Index, PPI）。两者均由国家统计局于每月月初发布，是衡量经济通货膨胀（以下简称“通胀”）水平的有力指标，全面反映了全国消费品和生产产品的总体价格变化，对于政府货币政策和进行资产定价都具有指导意义。

CPI 是反映居民购买的消费性商品及服务的价格水平随时间变动的相对数值。根据 2016 年 1 月国家统计局对 CPI 统计口径进行的新一轮调整，目前 CPI 的构成成分共有八大类，分

别是食品烟酒、衣着、生活用品、医疗保健、交通通信、娱乐教育、居住和其他用品。

通过图 2-6 我们看到自 2000 年以来我国 CPI 同比增长率的变动情况，除了在 2002 年和 2009 年出现了通货紧缩，其他时期的物价基本维持上涨态势，且通胀率全部处于 10% 以下，这表明我国在这段时间内并未出现恶性通胀，但是我们仍需注意到 2007—2008 年我国曾经出现物价持续上涨的情况，CPI 同比增长率于 2008 年 2 月达到了样本区间内的最高点 8.7%。但在 2008 年受金融危机影响之后，CPI 同比增长率随着宏观经济水平的下行，大幅走低，甚至进入了通货紧缩的状态。在 4 万亿元刺激之后，通胀从 2009 年下半年大幅拉升，在 2011 年下半年再次达到 6.5% 这样一个较高水平。自 2012 年下半年，我国的通胀水平总体保持平稳，在 1%~3% 的区间波动。

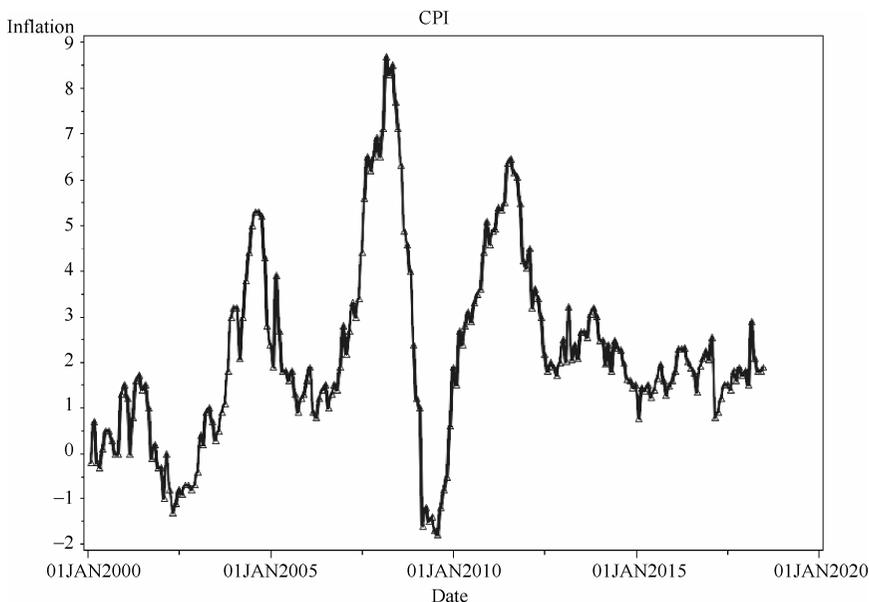


图 2-6 2000—2018 年 CPI 同比增长率变化

我国国家统计局会同时公布食品类 CPI 与非食品类 CPI。由于食品类消费品价格对供需反应迅速，易受气候、自然灾害等因素影响，从图 2-7 我们也可以看到它的波动远比非食品 CPI 的波动剧烈，后者非常稳定，基本在 0 的水平上下徘徊，所以综合 CPI 数据的波动基本来自食品类消费品价格。

除了 CPI，PPI 也是重要的价格指数，所不同的是，PPI 统计范围是生产者购买商品和劳务的价格变化，表明企业面临的购买成本的升降。

我们通过图 2-8 可以明显地观察到 CPI 同比增长率与 PPI 同比增长率之间的正相关性，二者很少出现背离的情况，同时由于 PPI 中能源占比较大，波动更剧烈。PPI 同比增长率自 2012 年开始连续出现负值，从 2012 年一直跌到 2016 年。与此相对应的是，在这期间，我国的煤炭企业、钢铁企业大面积亏损。进入 2016 年，PPI 同比增长率开始止跌回升，并进一步由负转正。这一定程度上说明我国政府在推行去产能政策之后，工业产品的价格出现了显著回升。

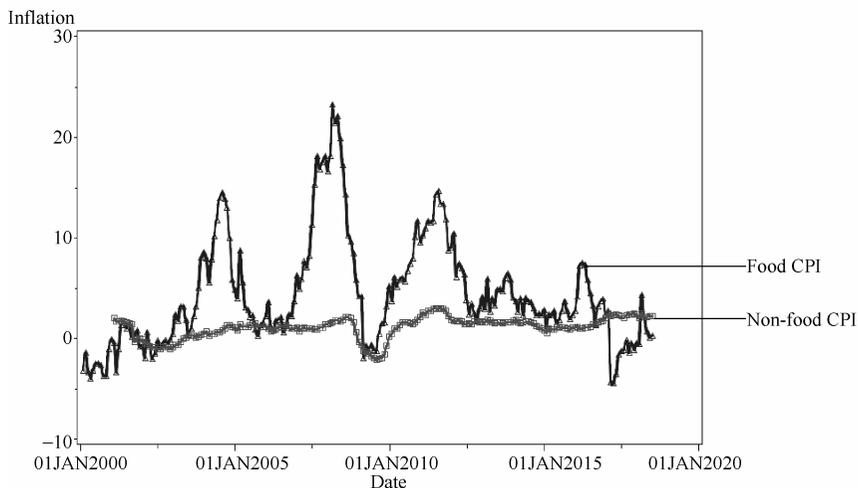


图 2-7 食品类 CPI 与非食品类 CPI 同比增长率

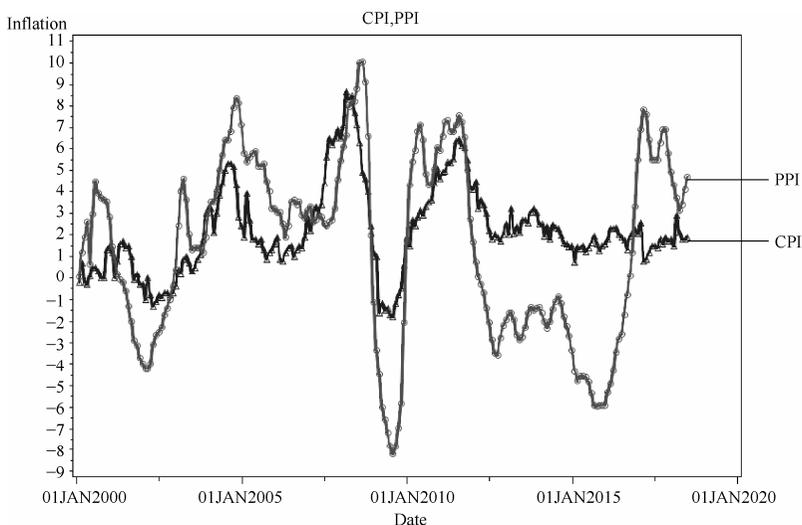


图 2-8 2000—2018 年 CPI 与 PPI 的走势

接下来我们将 PMI 和 CPI 放到一起来评估宏观经济运行。我们从图 2-9 中可以看出，2005—2007 年是中国宏观经济非常繁荣的一个时期。在这段时间，PMI 的均值达到 55，同时通货膨胀率只有 2%左右。这说明宏观经济处于一段非常高质量的增长时期。

但从图 2-10 中我们可以看出，中国宏观经济自 2010 年至 2016 年，PMI 和 CPI 同比增长率都处于整体下降的趋势，这段时期也是我国宏观经济非常困难的一段时间。但从 2016 年开始，PMI 和 CPI 同比增长率都明显反弹，说明经济体内部活动日渐活跃，经济不断向好。但可以看到从 2018 年上半年开始，PMI 已经不断走低，进入了下降通道。

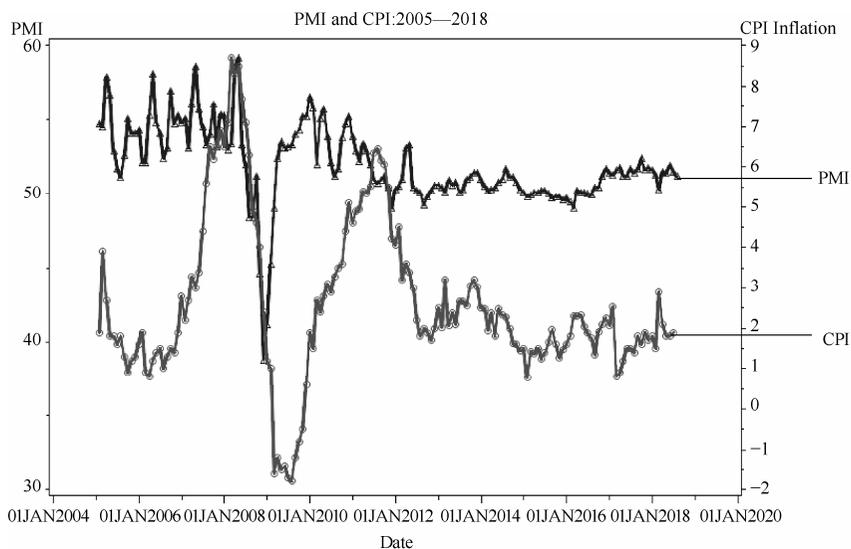


图 2-9 2005—2018 年 CPI 与 PMI 的走势

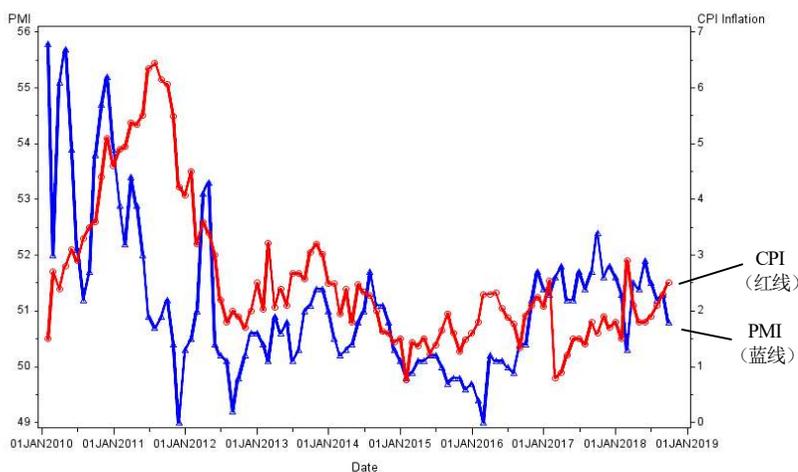


图 2-10 2010—2018 年 CPI 同比增长率与 PMI 的走势

## 2. 社会消费总量的变化

除了价格水平维度，消费变化也可从国内消费总量水平上体现。图 2-11 显示的是社会消费品零售总额的同比增速。

从图 2-11 可以看出，2008 年是消费增长的转折点，从之前的逐步上升，转为下降。尽管在过去 10 年，我国电商企业崛起，为消费者提供了便捷的服务，但我们的消费增长率不仅没有上升，反而从 20% 降到了目前的 10% 左右。这样的消费增长率的下降，有可能是由于房地产价格的大幅上升，对消费造成了挤出，导致消费的增长持续下滑。

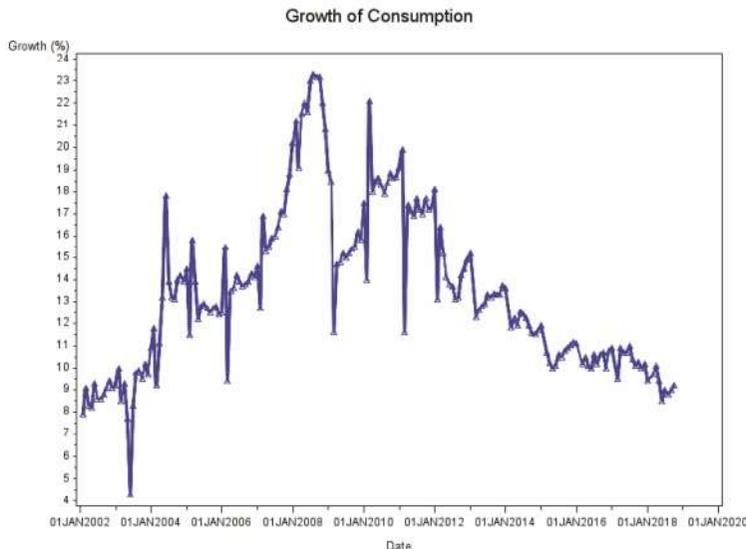


图 2-11 2002—2018 年社会消费品零售总额同比增速 (%)

### 2.3.3 投资

投资是另一驾拉动经济发展的马车，是政府刺激经济关注的重点，也是我们分析宏观经济的重点。分析投资时，我们关注三点：①固定投资完成额增长率；②工业附加值增长率；③货币供应的增长率。

#### 1. 固定投资完成额增长率

如图 2-12 所示，从总体来看，我国的固定资产投资完成额的增长率自 2002 年以来均在 10% 以上，并出现过两波明显的投资热潮。2004 年曾出现过一波投资过热的浪潮，固定资产投资完成额增速高达 50%。之后随着中央调控政策的出台和实施，才逐步回落至正常水平。除此之外，2008 年 4 万亿元刺激计划的出台也曾带来新一波投资热潮。我们可以看到除了投资过度膨胀的时期，2012 年之前固定资产投资的增长率基本都稳定在 20% 至 30% 的区间。值得注意的是 2012 年以来，固定资产投资完成额增速从 20% 的高位逐渐滑落，刺激计划余温消退，经济逐步走入下行通道，目前降至 6% 左右，与这段时期的 GDP 等指标相吻合。

#### 2. 工业附加值增长率

工业附加值，作为企业在生产过程中，对投入带来的价值增加，可以理解为经营成果或利润。工业附加值的增长，在过去几年逐年下降，目前已经降至较低的水平。当企业利润增速较低的时候，企业投资的意愿会进一步受抑制，这预示着未来我国投资的增速有可能会进一步放缓。图 2-13 显示了 2000—2018 年工业附加值同比增速。

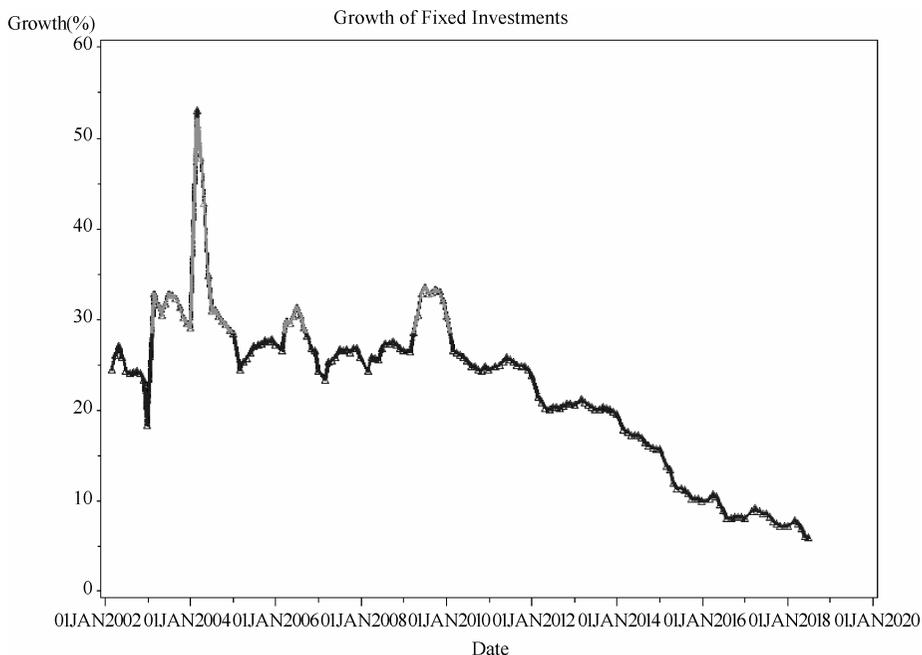


图 2-12 2002—2018 年固定资产投资完成额累计同比增速 (%)

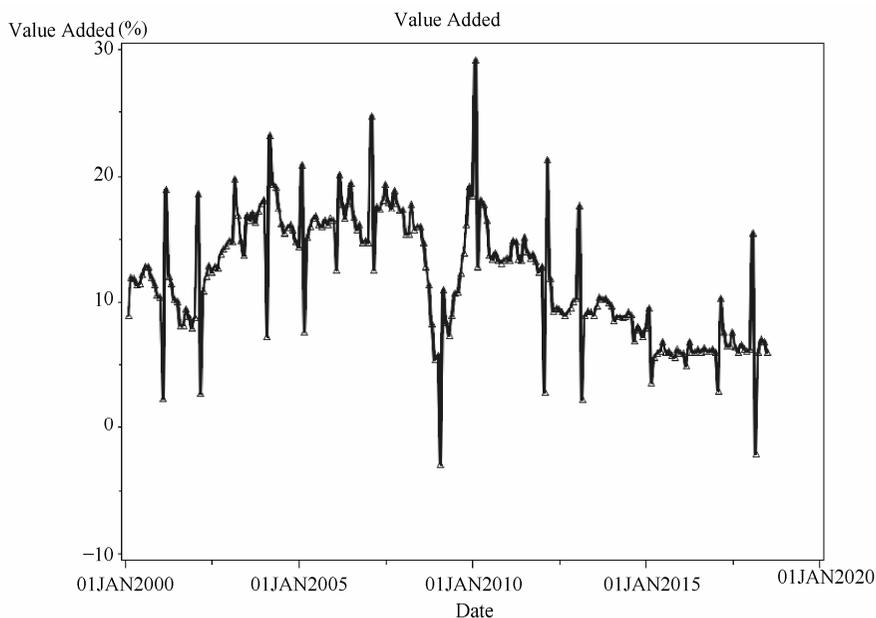


图 2-13 2000—2018 年工业增加值同比增速 (%)

### 3. 货币供应的增长率

从投资资金来源的角度分析，多数投资都依靠借款融资，因此相关的资金供给指标，如金融机构新增人民币贷款和货币供应量，能够有效反映社会投资需求的变化。图 2-14 显示的就是我国广义货币供应量（M2）的同比增速，在 2009 年有一个明显的波峰，这源自 4 万亿元刺激计划带来的货币供给显著上升。之后，我国的货币政策趋于稳健，货币供给的增速

不断下降。当货币供给增速下降，社会融资成本上升时，企业的投资成本就会进一步增加，企业的投资意愿就会继续下降，导致投资增速的进一步放缓。

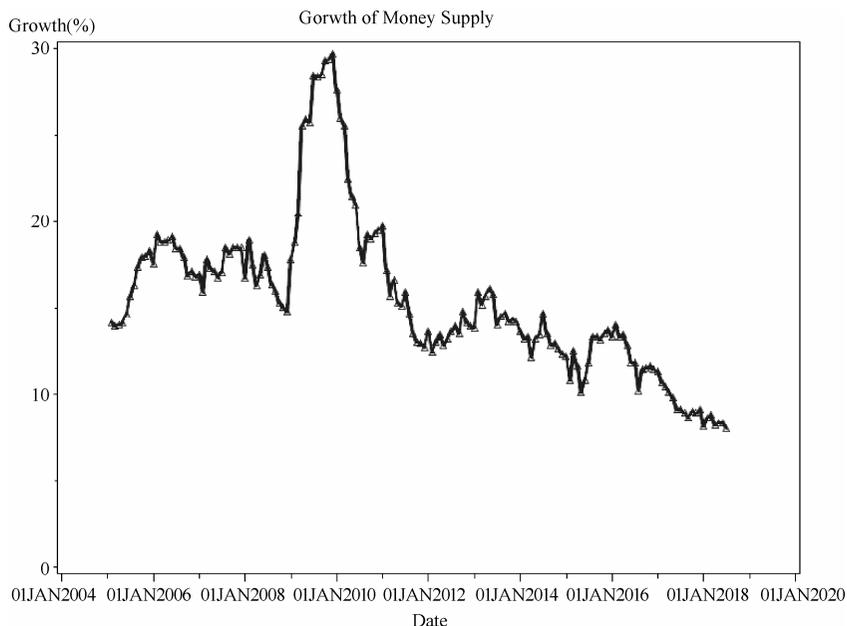


图 2-14 2004—2018 年 M2 每月同比增速 (%)

### 2.3.4 净出口

净出口是GDP的重要组成部分，成为拉动经济的第三驾马车。从图 2-15 我们可以看到，自 2017 年以来，我国的出口金额增速基本处在下滑的通道，特别是在近期中美贸易战的背景下，出口不容乐观。

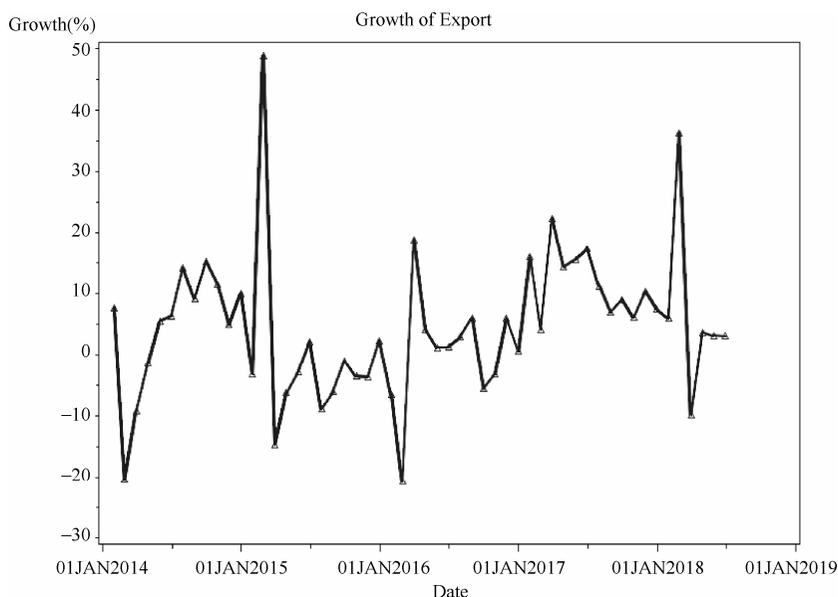


图 2-15 2014—2018 年出口金额同比增速 (%)

另外，从图2-16我们发现，与出口相对应的进口金额的增速从2017年开始也逐步下降，目前维持在10%左右。

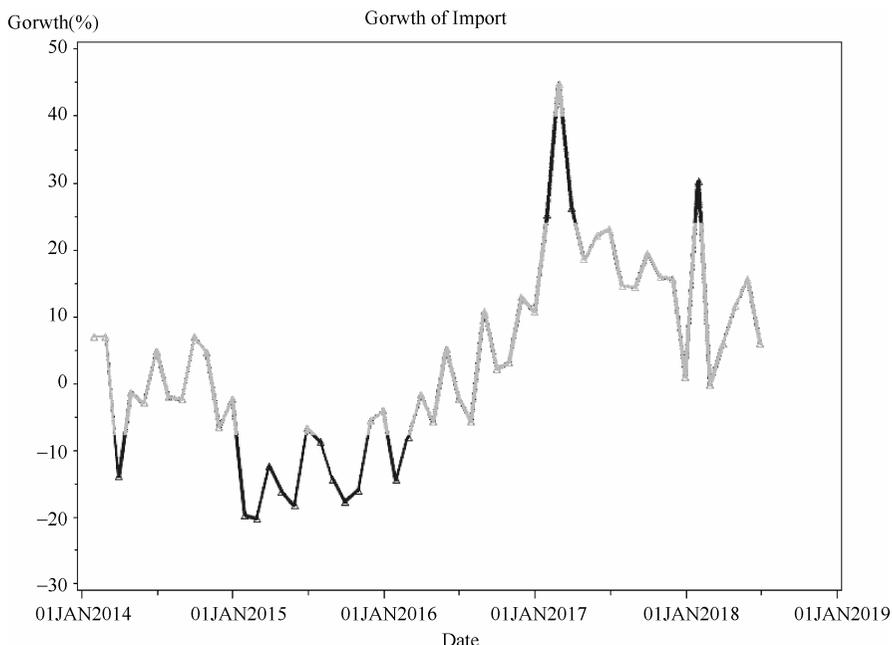


图 2-16 2014—2018 年进口金额同比增速 (%)

## 2.4 政府政策与政治影响

在介绍完经济基本面的分析之后，接下来我们看政府对宏观经济的影响。作为国民经济中的重要部门，政府在经济运行中起到了调控和监管的作用。政府可以通过经济政策，如财政或货币政策达成稳定物价、充分就业、经济增长、国际收支平衡和金融稳定等目标。由于不同资产与利率、资金流动性、经济增长等的相关性各不相同，因此政府重大政策对不同资产价格的作用也各不相同。同时，一国的政治环境是否和平直接影响了投资者的情绪和预期，从而影响资产价格的涨跌，动荡的社会虽然不利于发展金融市场，但是也能为投资者带来投资机会。所以，我们通过分析政府的政策和政治环境的变化，同样能够把握资产价格的走势。

### 2.4.1 经济政策

政府的经济政策是指政府通过直接或间接的手段对经济进行调控。政府的经济政策可分为货币政策与财政政策。其中，货币政策是指中央银行通过货币政策工具调节货币供应量和利率，从而影响资金流动性和宏观经济运行；财政政策是指政府通过财政支出和税收政策等手段来调节总需求。

接下来，我们将以我国政府2008年年底出台的4万亿元刺激计划为例，看看政府政策是如何影响股价变动的。

2008年，全球金融危机的爆发对我国实体经济造成了严重冲击，直接导致我国2008年GDP

增速陡降，并于第四季度跌至 6.5%。同时，上证综指从 2008 年 1 月 14 日的 5522 点，在短短 10 个多月的时间内跌了 70%，于 2008 年 10 月 28 日跌至 1664 点。为了应对金融危机和救市，2008 年 11 月 9 日国务院出台了更加有力的扩大国内需求的措施，推出 4 万亿元投资，主要用于基础设施建设。

4 万亿元刺激计划发出的宽松信号和对实体经济的直接刺激，扭转了人们对经济的悲观预期，在当时迅速有效地将中国经济从危机的泥潭中拉出。刺激计划的公布直接带动了 2009 年股市的新一轮持续上涨，将上证综指从 2009 年年初的低位拉至 2009 年 8 月 4 日的 3478 点，如图 2-17 所示。



图 2-17 2008 年年底 4 万亿元刺激计划前后上证综指的变化

## 2.4.2 政治与社会稳定

只有稳定的政治和社会环境才能够保证企业得到公平和充足的发展机会，提升投资者的信心，它是企业效益蒸蒸日上、金融市场蓬勃发展的摇篮。相反，社会环境和政治的动荡会加剧投资的不确定性，抑制投资，导致经济发展停滞不前。

然而，由于投资者的恐慌心理和悲观预期，社会动荡所带来的资产价格大幅下跌却能够创造不少投资机会。坚持价值分析、目光长远的投资者往往能够从资产的不合理低价中发现机会，等到资产股价重回合理水平时收获丰厚的回报。

## 2.5 金融市场的分析

金融市场本身作为宏观经济的重要组成部分，其与资产价格的关系最为紧密。横向对比各国金融市场的发展现状，我们可以发现在市场机制运行合理、信息传递高效、投资者权益受到保障的金融市场，其市场更加有效，资产价格反应更加迅速到位。我国目前还处于不断探索进行金融市场改革、完善市场运行机制的阶段，改革带来机遇与挑战。而完善的金融市场本身可以通过市场流动性、投资者情绪等影响资产价格。放眼国际，各国利率水平也能够影响全球资产配置结构。

## 2.5.1 市场改革

金融市场自身的发展之路，必定是一条逐渐开放、资源配置效率不断提高、投资者利益不断得到保障的改革之路。纵观中国股市，最近 10 年间发生的最重大的改革莫过于 2005 年启动的股权分置改革，这也是自中国证券市场成立以来影响最为深远的一次改革。

股权分置是指我国在经济转型和资本市场发展中，由于所有制等历史因素，造成了上市公司同时存在流通股（主要为社会公众股）和非流通股（主要为国有股和法人股）的局面。流通股与非流通股之间交易和转让渠道相异，定价机制扭曲，且存在“同股不同权、同股不同利”等弊端，股东之间利益不平衡，严重影响了上市公司的治理效率，制约了中国股市的健康发展。

而我国实施股权分置改革的目的在于逐步放开非流通股的流通比例，并通过“对价”方式补偿因非流通股入市可能会给流通股股东带来的利益损失，平衡各方利益，实现两种股权的对接。股权分置改革采取逐步试点、分散决策的稳定性方案，按部就班，稳步推进。

为了尽量减少对市场的冲击，股权分置改革严格限制了非流通股股份解禁的份额和速度。自每家上市公司改革方案实施之日起，持有上市公司股份总数的 5% 以上的原非流通股股东，在 12 个月之内不得上市交易和转让。上述规定期满之后，可以通过证券交易所交易出售原非流通股股份，出售股份所占总股数比例在 12 个月之内不得超过 5%，在 24 个月之内不得超过 10%。

从改革效果来看，股权分置改革不仅没有对股市造成严重的负面冲击，还消除了改革前的种种顾虑，释放了积极信号，带动了新一轮牛市，绝大多数股票和股指本身在这次改革中出现了明显的上涨。从图 2-18 我们可以看到股权分置改革前后上证综指的变化。

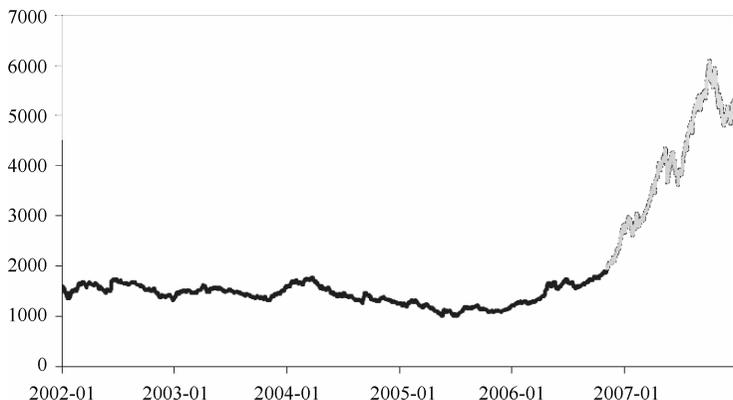


图 2-18 股权分置改革前后上证综指的走势

## 2.5.2 流动性

金融市场资金的流动性能够直接影响资产的价格：当市场流动性充裕时，资金成本低，更多资金流入股市，推动股价进一步抬升；当市场流动性紧缺时，资金撤离，带动股价下跌。

而中国人民银行（以下简称“央行”）通常会根据金融市场的流动性和股价，通过货币政策工具时刻对流动性进行调整，以防止流动性危机和系统性风险的爆发，促进市场稳定。

当市场流动性明显不足、股市大幅下挫时，央行一般会通过降准降息、公开市场操作等工具注入流动性，缓解资金面的紧张。央行的流动性调节可以从 M2 增速、每周公开市场操作货币净投放量等指标衡量。

### 2.5.3 投资者情绪

投资者情绪是指投资者对于未来资产价格走势的乐观或者悲观的预期，是投资者根据市场信息做出的主观判断，从而影响其投资行为。当个体投资者的情绪汇总成为市场整体现象时，能够显著影响资产的价格。当投资者情绪普遍高涨时，会放大市场的利好消息，弱化利空消息，短期内有利于推升股票价格，但是会造成价格泡沫。

衡量投资者情绪的指标有市场成交金额、换手率和新增开户数。市场交易量越大，成交越活跃，投资者情绪越高涨。

## 总结

宏观经济分析是投资分析当中必不可少的重要一步。在对宏观经济运行进行分析时，首先，我们看经济总量指标，关注 GDP 增长率和 PMI。其次，我们会分别分析经济的三驾马车：消费、投资和净出口。在对宏观经济基本面有一个完整认识之后，我们需要对政府政策进行分析，因为政府的政策、政治的稳定，对资产的价格都会有重要影响。此外，金融市场的分析也非常重要，它重点关注的是市场改革、流动性和投资者情绪三个方面。

## 第 10 章 SAS 在量化投资中的应用<sup>①</sup>

本章利用几个具体的案例介绍 SAS 在量化投资中的应用。

### 10.1 7 个应用案例

**案例 10-1:** 在一个包含股票周度收益率的数据集 `weekly_ret` 中, 变量 `date` 是一个数值型变量, 如 2012 年 12 月 31 日是以数字 20121231 存储在观测中。怎样把变量 `date` 转化成 SAS 中的日期格式, 是我们在导入 SAS 数据时经常会遇到的问题。虽然是日期, 但 SAS 会将 20121231 当作一个 2012 万+的数字。我们需要告诉 SAS, 从这个数值型变量中分别提取年、月、日, 再整合到一起得到日期。

```
data weekly_ret;
set weekly_ret;
date_use=put(date,8.);
year=substr(date,1,4);
month=substr(date,5,2);
day=substr(date,7,2);
year_use=year*1;
month_use=month*1;
day_use=day*1;
date_use=MDY(month_use,day_use,year_use);
format date_use date9.
drop year month day year_use month_use day_use;
run;
```

以上就是相应的代码。为了获得年、月、日, 我们首先用 `put` 函数将 `date` 从数值型变量转变为字符型变量。然后, 我们用 `substr` 函数提取字符。比如, 为了提取年, 我们从第一个字符开始提取, 提取四位, 就把年提取出来了; 为了提取月, 我们从第五个字符开始提取, 提取两位就把月提取出来了; 为了提取日, 我们从第七个字符开始提取, 提取两位, 就把日提取出来了。提取出来的年、月、日作为字符, 先放在 `year`、`month`、`day` 里面, 然后再乘

---

① 本章使用的 SAS 代码和数据集, 请扫二维码获得。



以 1 将其转化为数字，再利用 `mdy` 函数将其整合为日期，最后利用 `format` 将其格式定为 31Dec2012 的形式。

**案例 10-2：**每个月（1 月、2 月……12 月）到底哪个行业的股票表现得更好？为了解决这个问题，你需要得到每个月每个行业的股票收益率情况。假设现在你有一个包含月度股票收益率的数据集 `monthly_ret`，包含变量：行业代码 `indus_code`、月份 `month`、年份 `year`、公司代码 `firm_code` 和收益率 `ret`。请利用 SAS 编程得到每个月每个行业股票收益率的均值、标准差、中位数以及 5%、25%、75% 和 95% 分位数的值，同时把数据输出到数据集 `return_month_industry`。

这个问题与投资中的行业轮动有关。我们想了解，在不同的月份，哪些行业的股票收益表现会比较好。比如，夏季家用电器的消费是比较旺盛的，所以在 7~8 月份，家电行业的股票收益可能比较高。为了研究这个问题，我们要做的，其实是一个简单的按月度、行业进行的收益率的统计。

```
proc sort data=monthly_ret;
  by month indus_code;
proc means data=monthly_ret noprint;
  var ret;
  output out=return_month_industry
  mean=m median=md std=sd N=n Min=minimum Max=maximum
  p5=p_5 p25=p_25 p50=p_50 p75=p_75 p95=p_95;
  by month indus_code;
run;
```

以上是相应的代码。首先，用 `sort` 语句将 `monthly_ret` 按照月度和行业去排序，然后利用 `means` 语句。在这里，我们利用 `output out=return_month_industry` 告诉 SAS 我们要把统计的结果输出到 `return_month_industry` 这个数据集。`p5=p_5`，是告诉 SAS 将 5 分位数输出，并命名为 `p_5`，其他类似。

**案例 10-3：**市盈率（PE）是一个重要的预测回报的指标。在股票市场上，我们常常能够发现，一些市盈率较高的股票通常会比市盈率低的股票表现出更高的回报。假设你现在从数据库下载了两个不同的数据集：一个包含周度股价和收益率的数据集 `weekly_ret` 和一个包含年度财务利润的数据集 `earning`。编程合并两个数据集以计算市盈率，注意只需编写合并数据集的代码，而不需要编写后续计算市盈率的代码。`weekly_ret` 数据集中包含变量：年份 `year`、日期 `date`、价格 `price`、总股数 `total_shares` 和公司代码 `firm_code`。`earning` 数据集包含变量：年份 `year`、总利润 `total_earning` 和公司代码 `firm_code`。

这个案例是我们在量化投资中必会遇见的。我们的数据有两类：财务数据和金融市场数据。将这两类数据合并是量化投资中必不可少的步骤。

```

proc sort data=weekly_ret;
by firm_code year;
proc sort data=earning nodupkey;
by firm_code year;
data weekly_ret;
merge weely_ret (in=a) earning;
by firm_code year ;
if a;
run;

```

在这两类数据中，共同的变量为公司代码和年份。上述代码就是将这两类数据按 `firm_code` 和 `year` 来进行合并。这里特别要强调的是在对 `earning` 进行 `sort` 时加了 `nodupkey`，这是为了确保财务数据能够用公司代码和年份进行一一对应。

**案例 10-4：**运用案例 10-3 中的 `weekly_ret` 数据集，编程得到每只股票的下周收益率变量 `ret_next_week`。

这个案例也是在量化投资中经常遇到的。选出股票之后，就要计算股票未来一周的收益率，从而查看选股的效果。

```

proc sort data=weekly_ret;
by firm_code descending date;
run;
data weekly_ret;
set weekly_ret;
by firm_code descending date;
ret_next_week=lag(ret);
if first.firm_code then ret_next_week=.;
run;

```

为了获得未来收益率，我们要利用 `lag` 函数。在用 `lag` 函数之前，要用 `sort` 语句将 `weekly_ret` 按日期倒序，这样对于每只股票，日期从上到下，由距离现在较近依次排到距离现在较远。SAS 在取 `lag` 函数时，默认是取该观察值上面一个观察值，因此就取到了未来的收益。

**案例 10-5：**运用案例 10-3 中的 `weekly_ret` 数据集，根据 2012 年 12 月最后一周的总市值数据，挑选出市值最大的 20 只股票。

这个案例，也是在量化投资中常遇到的，即取一定时间下某指标最大或最小的规定数量

的股票。

```

data stock_val;
set weekly_ret;
if INTCK('DAY ',date,'28Dec2012'd)=0;
mkt_val=price*total_shares;
keep firm_code mkt_val;
proc sort data=stock_val;
by descending mkt_val;
run;
data stock_selected;
set stock_val;
by descending mkt_val;
n+1;
if n<=20;
keep firm_code;
run;

```

上述代码的思路是：首先，提取出 2012 年 12 月 28 日这一天的股票数据；其次按市值倒序，这样市值大的股票就排在前面。然后再利用  $n$  这个变量。 $n+1$  这一行代码的效果，是产生一个变量  $n$ ，从第一行的 1 逐步加 1。因此，当只取  $n \leq 20$  的股票时，就选取了市值最大的 20 只股票。

**案例 10-6：**将一个日度股票交易数据 stock\_daily，只保留每周最后一个交易日的数据，从而将其转化为周度数据。

这个案例是经常会遇到的。我们下载的原始数据通常是日度数据。为了计算周度收益，我们需要将每周最后一个交易日的数据提取出来。具体代码如下：

```

proc sort data=stock_daily;
by firm_code year week descending date;
data stock_week;
set stock_daily;
proc sort data=stock_week nodupkey;
by firm_code year week;
run;

```

在上述代码中，我们用了两步 sort。在第一步 sort 中，我们按公司代码、年、周和日期排序。其中，我们将日期倒序。这样就将每周内数值最大的日期，也就是当周最后一个交易日排在最前面。在第二步 sort 中，我们按公司代码、年、周来排序，利用 nodupkey，每个公司、每年、每周只保留当周第一个观察值，也就是当周最后一个交易日的观察值。

**案例 10-7:** 假设现在有这样一种投资策略: 如果一只股票的财务利润由负转正, 则从盈利宣布之日起持有它三天, 即从  $t$  期持有至  $t+2$  期。为了检验它的投资表现, 首先你需要得到每只股票自盈利宣布之日起三天的收益率。假设现在有两个数据集: 一个数据集 `earning_announce`, 包含变量盈利宣布日期 `date_announce` 和公司代码 `firm_code`; 另一个数据集 `daily_ret`, 包含日期 `date`、日收益率 `ret` 和公司代码 `firm_code`。其中, 日收益率 `ret` 已经过除权除息调整。编程合并两个数据集并得到盈利宣布之日起三天的日收益率。

这个案例, 要求从 `daily_ret` 这个数据集中提取每个 `date_announce` 起三天的股票收益。我们利用这个案例, 介绍一个新的命令: `sql`。

```
proc sql;
  create table ret_data
  from earning as a, daily_ret as b
  as select *
  where a.firm_code = b.firm_code and
  intck('day', a.date_announce, b.date) <= 2 and
  intck('day', a.date_announce, b.date) >= 0;
quit;
```

`sql` 是 SAS 里面非常强大的函数, 可以用来计算变量统计、合并数据集等。我们这里介绍的是如何利用 `sql` 进行条件合并。第二行, `create table ret_data`, 告诉 SAS 要建一个新的数据集, 叫 `ret_data`。第三行, `from earning as a, daily_ret as b`, 告诉 SAS, `ret_data` 是由 `earning` 和 `daily_ret` 这两个数据集而来, 并把 `earning` 简称 `a`, 把 `daily_ret` 简称 `b`。第四行, `as select *` 是指将 `earning` 和 `daily_ret` 的所有变量都放进 `ret_data`。第五行, 用 `where` 语句进行条件合并。`a.firm_code = b.firm_code` 是指两个数据集中的公司代码要对上。第六行, `intck('day', a.date_announce, b.date) <= 2` 是要求 `daily_ret` 里的日期比 `earnings` 里的 `date_announce` 要晚不多于两天。第七行, `intck('day', a.date_announce, b.date) >= 0` 是要求 `daily_ret` 里的日期一定要在 `earnings` 里的 `date_announce` 当天或之后。这样就把 `date_announce` 起三天的收益数据提取出来了。

## 10.2 行业动量效应

行业动量效应是指行业平均收益率有延续原来运动方向的趋势, 过去一段时间收益率较高的行业, 未来表现通常也会比较好。行业动量效应可以指导我们进行行业轮换, 买入过去表现较好的行业股票, 持有一段时间, 然后不断重复更替行业选择。关于行业动量效应, 最重要的文献是马克维茨 (Moskowitz) 和格林布拉特 (Grinblatt) 两位学者于 1999 年发表的。他们利用 1963 年 6 月至 1995 年 6 月的美国股市数据, 发现行业动量是动量效应的主要来源。

具体来说, 使用 SAS 对行业动量效应进行检验可分为以下六步, 分别为:

- ① 获取周度股票收益率数据；
- ② 获取行业分类数据；
- ③ 获取每周行业平均收益率；
- ④ 对行业每周平均收益率进行排序；
- ⑤ 选取收益率最高的行业；
- ⑥ 对行业未来收益率进行统计分析。

接下来，我们分别详细阐述每一步的步骤。

第一步：获取周度股票收益率数据。

- ① 从 Excel 文件中导入每只股票周度收益率数据。

```
proc import
  datafile="C:\data\trading_all_weekly_back_2013.xlsx"
  dbms=xlsx
  out=work.stock
  replace;
run;
```

- ② 剔除停牌（即股票交易量为 0）的观测。

```
data stock;
  set stock;
  if trd_vol_shr>0;
run;
```

- ③ 将数字格式的日期转化成日期格式。

```
data stock;
  set stock;
  year=int(date/10000);
  month=int((date-year*10000)/100);
  day=date-year*10000-month*100;
  date_new=MDY(month, day, year);
  drop date;
run;
```

这里，我们用的是另一种将数字转化为日期的方式。比如 20121231，为了获得年，我们将该数字除以 10000，取整，就得到 2012。将 20121231-2012×10000 得到 1231，我们将其除以 100，再取整，就得到 12。再进一步得到日的数值。

```
data stock;
  set stock;
  date=date_new;
  format date date9.;
run;
```

**第二步：获取行业分类数据。**

- ① 从 Excel 文件中导入行业分类数据。

```
proc import
  datafile="C:\data\stock_infor.xlsx"
  dbms=xlsx
  out=work.infor
  replace;
run;
```

- ② 合并行业分类数据集和股票收益率数据集，并进行变量整理。

```
proc sort data=stock;
  by code;
proc sort data=infor;
  by code;
data stock;
  merge stock(in=a) infor(in=b);
  by code;
  if a & b;
run;

data stock_ret;
  set stock;
  year=year(date);
  week=week(date);
  keep code year week ret indus_sw_3;
run;
```

这里我们采用的是申万三级行业分类即 indws\_sw\_3。

**第三步：获取每周行业平均收益率。**

- ① 计算每周每个行业的平均收益率。

```
proc sort data=stock_ret;
  by indus_sw_3 year week;
proc means data=stock_ret noprint;
  var ret;
  output out=indus_ret
  mean=ret_indus;
  by indus_sw_3 year week;
run;
```

② 获取每个行业下周的平均收益率。

```
proc sort data=indus_ret;
  by indus_sw_3 descending year descending week ;
run;
data indus_ret;
  set indus_ret;
  by indus_sw_3 descending year descending week;
  ret_indus_fut=lag(ret_indus);
  if first.indus_sw_3 then ret_indus_fut=.;
run;
```

**第四步：对行业每周平均收益率进行排序。**

① 统计行业的个数以决定最佳的分组数。

```
data num_indus;
  set indus_ret;
  proc sort data=num_indus nodupkey;
  by indus_sw_3;
run;
```

② 对每周的行业平均收益率进行排序分组。

```
proc sort data=indus_ret;
  by year week;
  proc rank data=indus_ret out=indus_ret groups=100;
  var ret_indus;
  ranks rank_ret_indus;
  by year week;
run;
```

**第五步：选取收益率最高的行业。**

```
data indus_selected;
  set indus_ret;
  if rank_ret_indus>=95;
```

在这里，我们利用 `rank_ret_indus>=95`，选择过去一周收益率排名在前 5% 的行业。

**第六步：对行业未来收益率进行统计分析。**

对其未来收益率进行假设检验（T 检验）。

```
proc ttest data=indus_selected;
  var ret_indus_fut;
run;
```

得到的 T 值，只有-0.43，说明收益并不显著，即行业动量在中国的 A 股市场效果并不好。图 10-1 反映了行业动量策略的累计收益情况。

从图 10-1 中可以看出，行业动量在样本内随着市场的下跌也出现了下跌，几乎没有超额收益。

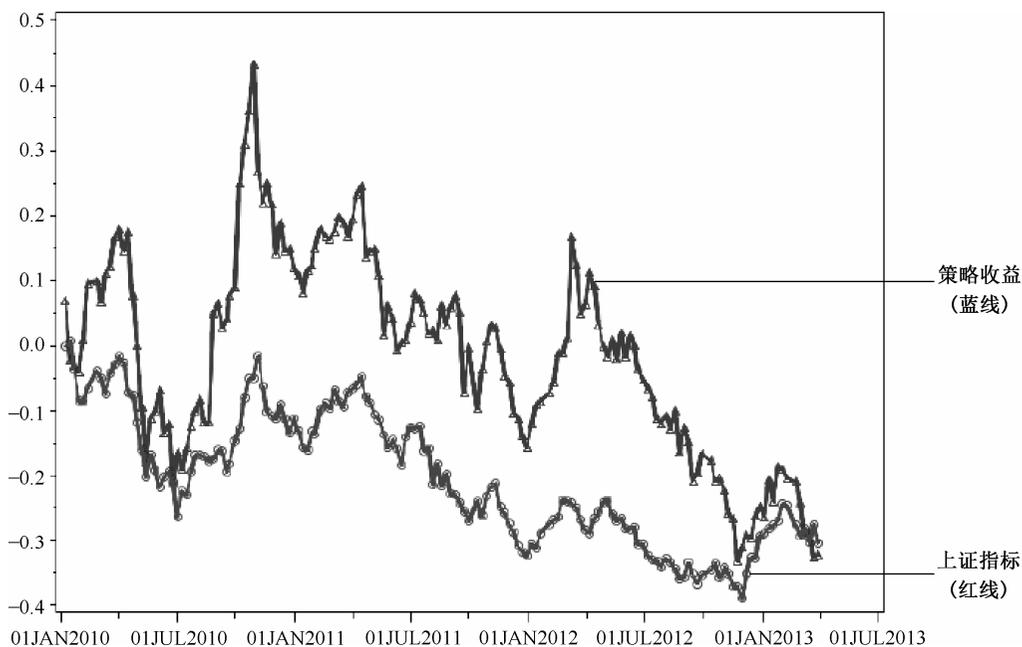


图 10-1 行业动量策略的累计收益

那么，我们能否想到一些改进的办法呢？一个改进的思路是，如果一个行业过去一周收益不错，但其中有些股票并没有涨，那么这些股票在接下来的一周就有可能补涨。按照这个思路，我们将行业的收益率排名合并到个股的数据集中，并计算个股在行业中的排名，同时利用这两个指标来选股，具体步骤如下。

#### 第一步，将行业的收益率排名合并到个股数据集。

```
proc sort data=indus_ret nodupkey;
  by indus_sw_3 date;
proc sort data=stock_ret;
  indus_sw_3 date;
data stock_ret;
  merge stock_ret indus_ret;
  by indus_sw_3 date;
run;
```

第二步，在每个行业内部，将个股按收益率进行排名。

```
proc sort data=stock_ret;
by indus_sw_3 year week;
proc rank data=stock_ret out=stock_ret groups=100;
var ret;
ranks rank_ret_stk;
by indus_sw_3 year week;
run;
```

第三步，计算个股的未来收益率。

```
proc sort data=stock_ret;
by code descending year descending week;
run;
data stock_ret;
set stock_ret;
by code descending year descending week;
ret_stk_fut=lag(ret);
if first.code then ret_stk_fut=.;
run;
```

第四步，选股。

```
data indus_stk_selected;
set stock_ret;
if rank_ret_indus>=50 & rank_ret_stk<=50;
```

在这里，我们选择的是过去一周里收益率最高的 50%的行业里收益最差的 50%的股票。

第五步，T 统计。

```
proc ttest data=indus_stk_selected;
var ret_stk_fut;
run;
```

我们进一步对选股的未来一周收益进行 T 检验，得到的周度收益率平均为 0.22%，年化收益率为 11%，且 T 统计量高达 11.88，说明选择过去高收益行业中低收益个股能够取得显著的收益。该统计样本为 2010 年 1 月至 2013 年 5 月，同期上证综指下跌 30%，本策略累计收益率为 40%，超额收益率达到了 70%。收益如图 10-2 所示。

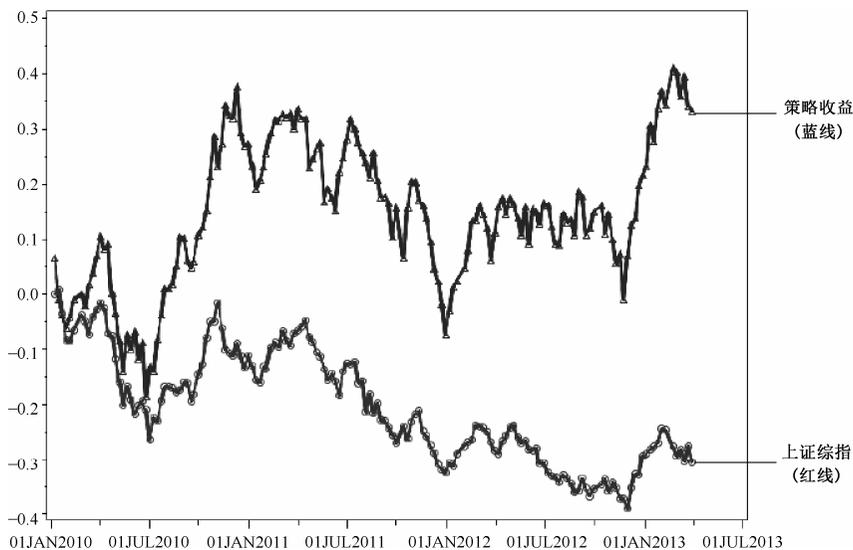


图 10-2 改进后的行业动量策略的累计收益

## 总结

SAS 因其强大的数据处理能力和高效的数据读取方式，在医药、商业等领域的大数据分析中具有广泛且深入的应用。SAS 编程语句简单、严谨，是北美高校金融学博士生从事实证研究的必备技能。