

专题报告

# 指数成分股调整的事件效应及增强策略

2020 年 11 月 25 日

## ——基于机器学习的视角

- 通常指数编制公司会定期根据指数的编制方案，对指数成分股进行相应的调整。近年来，指数化投资规模持续扩大，指数成分股调整也受到越来越多的关注。本文大致从两个方面进行了探讨，一是预测了 2020 年 12 月指数调整样本股，二是在机器学习视角下试图估计相关样本股受到的事件冲击影响。
- 中证指数公司会根据定期调整规则对市场上的股票进行筛选，选取出符合条件的样本股替换出根据规则应当被剔除指数的股票。本文我们梳理了沪深 300 和中证 500 指数的样本股选取流程，并给出了 2020 年 12 月指数样本股调整结果的预测。
- 根据对历史上定期调整中纳入和剔除股票的收益统计发现，在纳入效应的作用下，新入选指数成份股的样本股在公告日后的 10 个收益日内呈现出明显正向超额收益。在剔除效应的作用下，被剔除出指数成份股的老样本在公告日后的 15 个交易日则呈现出明显的负向超额收益。
- 受市值规模、占指数权重、流动性、中短期走势等因素的影响，样本股受到指数成分股定期调整事件的冲击存在一定的差异。并且这些影响因素呈现出多维化及非线性的特征。因此，我们决定摒弃传统的通过简单指标来判断样本股受冲击的影响程度的做法，而采用了能够有效处理复合问题的随机森林算法来预测调整样本的短期冲击。
- 经过实证研究，采用随机森林算法的能够有效学习并预测样本股受到的纳入效应或剔除效应，并能在一定程度上估计样本股在经过事件冲击后的短期超额收益，从而实现对原有事件效应的进一步增强。

任瞳

86-755-83081468

rentong@cmschina.com.cn

S1090519080004

周靖明

86-21-68407742

zhoujingming@cmschina.com.cn

S1090519080007

### 风险提示

本报告模型及结论全部基于对历史数据的分析，当市场环境变化时，存在模型失效风险。

敬请阅读末页的重要说明

“慧博资讯”专业的投资研究大数据分享平台

点击进入  <http://www.hibor.com.cn>

## 正文目录

一、引言 .....	3
二、指数成分股的定期调整 .....	5
2.1、成分股调整规则 .....	5
2.2、2020 年 12 月定期调整预测 .....	6
三、事件效应分析及增强策略 .....	9
3.1、事件分析 .....	9
3.2、成分股调整事件的增强策略 .....	10
3.3、数据准备 .....	10
3.4、模型的构建 .....	13
3.5、模型结果展示 .....	15
四、结语 .....	18

## 图表目录

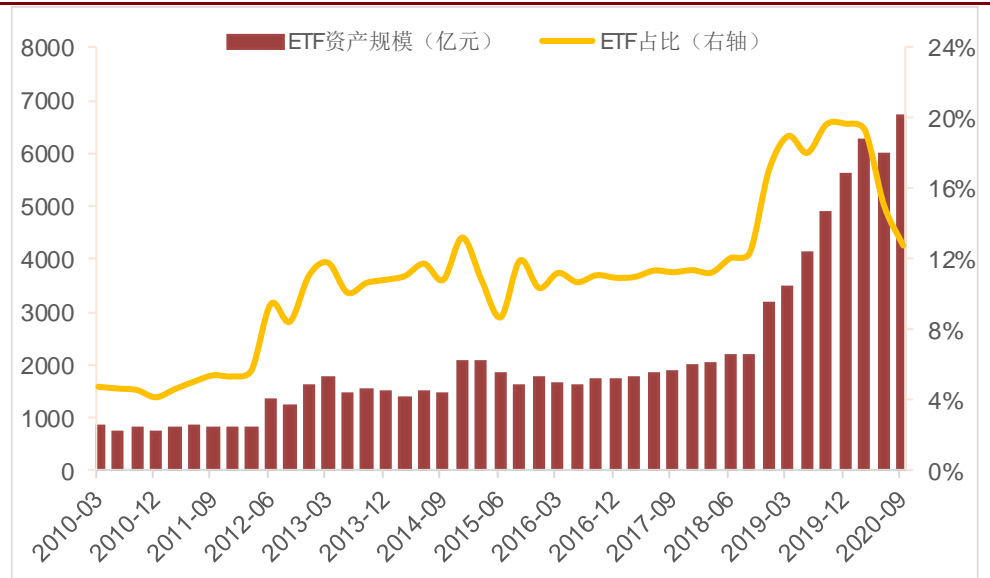
图 1: 我国 ETF 基金规模 .....	3
图 2: 我国被动指数基金规模 .....	4
图 3: 基金跟踪标的分布 .....	4
图 4: 沪深 300 历史纳入效应的超额收益表现 .....	9
图 5: 中证 500 历史纳入效应的超额收益表现 .....	9
图 6: 沪深 300 历史剔除效应的超额收益表现 .....	10
图 7: 中证 500 历史剔除效应的超额收益表现 .....	10
图 8: 被动指数基金调整持仓引发事件冲击逻辑图 .....	11
图 9: 被中证 500 指数剔除后是否进入沪深 300 的分组超额收益情况 .....	12
图 10: 净资金流量参考值重要性和被动指数基金规模的变化 .....	15
图 11: 沪深 300 纳入效应的分组超额收益表现 .....	16
图 12: 中证 500 纳入效应的分组超额收益表现 .....	16
图 13: 沪深 300 剔除效应的分组超额收益表现 .....	17
图 14: 中证 500 剔除效应的分组超额收益表现 .....	17
表 1: 沪深 300 和中证 500 的定期调整的筛选过程 .....	6
表 2: 沪深 300 指数成分股调整预测 .....	6
表 3: 中证 500 指数成分股调整预测 .....	7
表 4: 所选特征值信息 .....	12
表 5: IC 测试结果 .....	13
表 6: 时间序列上各特征重要度次序 .....	14
表 7: 纳入效应的详细分组表现 .....	16
表 8: 剔除效应的详细分组表现 .....	18

## 一、引言

每年 5 月和 11 月的下旬，中证指数公司都会对以沪深 300 和中证 500 为代表的指数样本股进行重新审核，并根据指数编制方案对指数成分股进行一定的调整。中证指数公司将于 6 月和 12 月初公布调整名单，并于 6 月和 12 月中旬正式生效。从对个股的影响来说，那些以对应指数为跟踪标的的被动指数基金将会买入新纳入的成分股，同时卖出被调出的老的成分股。因此，指数成分股的定期调整在某种程度上来讲，是利好新纳入的成分股而利空被调出的老成分股的。在当前我国以 ETF 为代表的被动指数基金规模大幅增长的背景下，指数成分股定期调整所带来的上述事件效应则越发明显。

近年来国内的 ETF 市场进入成长爆发期。截止 2020 年 11 月 23 日，ETF 基金合计规模达到 6700.87 亿元，相较于十年前增长了 712.19%。而其在股票型基金<sup>1</sup>中的规模占比也从 4.74% 提升到了 11.86%。近年来海外 ETF 规模的增长则同样迅猛。根据美国投资公司协会 ICI 统计，截止 2020 年 9 月，美国境内交易所 ETF 股票基金规模达到 2.71 万亿美元，同期股票共同基金的规模共计 13.99 万亿美元，占比为 19.37%。由此可见，国内 ETF 的发展方兴未艾，仍然还具有广阔的发展空间。

图 1：我国 ETF 基金规模



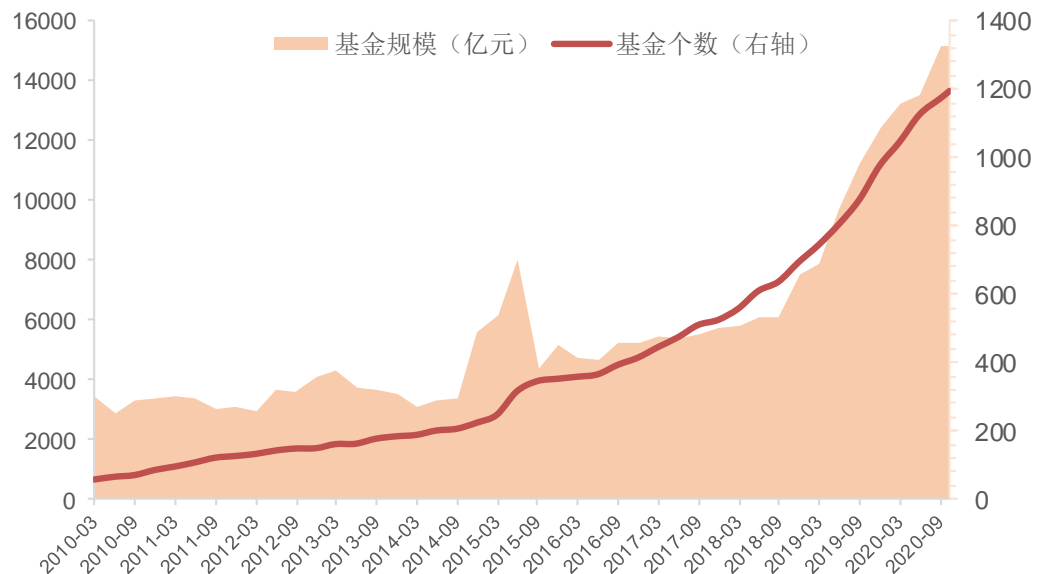
资料来源：Wind，招商证券，统计日期 20100101-20200930

除了 ETF 基金以外，市场上还有许多被动的投资形式是以指数型基金和指数增强型基金存在。在国内市场上，以指数型和指数增强型基金为代表的被动型基金(含 ETF)，不论是规模和基金数量在近年来都维持了较高的增长。截止 2020 年 11 月 23 日，被动型投资基金合计规模达到 15210.27 亿元，在股票型基金中规模占比为 26.88%。

<sup>1</sup> 本文中股票型基金包括 Wind 基金分类下的股票型基金和偏股混合型基金

敬请阅读末页的重要说明

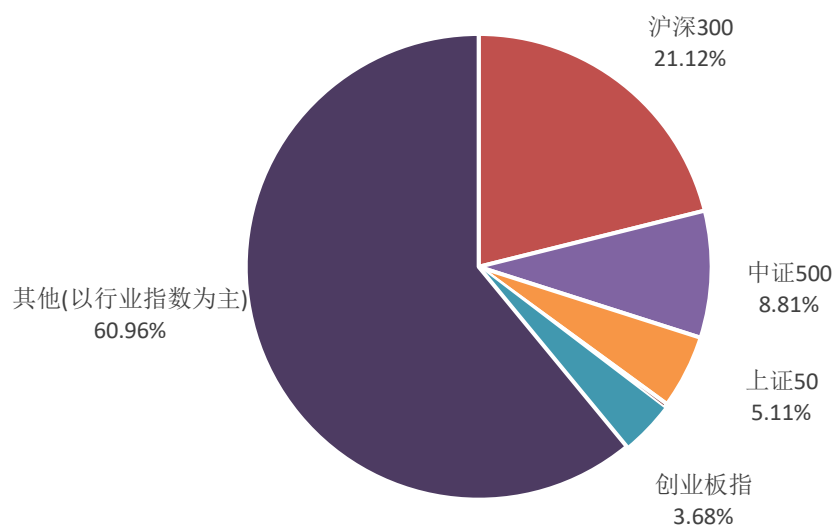
图 2：我国被动指数基金规模



资料来源：Wind，招商证券，统计日期 20100101-20201123

在众多的被动指数基金跟踪的标的中，沪深 300 和中证 500 无疑是非常最具代表性的标的指数。截止 2020 年 11 月 23 日，以沪深 300 和中证 500 指数为跟踪标的的被动指数基金（含指数增强基金）合计规模分别为 3208.62 亿和 1349.22 亿，在所有被动指数基金中的规模占比分别为 21.12% 和 8.81%，在宽基指数中分别排在前两名。

图 3：以不同宽基指数为跟踪标的的被动基金规模分布



资料来源：Wind，招商证券，统计日期 20201123

在此背景下，沪深 300 和中证 500 指数成分股的定期调整无疑会给相关个股带来越来越大的影响。同时也促使大家去根据指数编制规则来提前去预测指数成分股调整结

果，并通过事件驱动交易来获利。随着越来越多的投资者参与到这类事件驱动交易中，指数成分股调整对相关样本股的影响变得更加复杂化，也大大降低了此类事件的收益。一方面，本文将对指数的定期调整规则进行系统的梳理，给出 2020 年 12 月指数定期调整的预测名单。另一方面，我们将通过机器学习的方法，综合各类因素来刻画样本股受到指数成分股调整的影响，并以此来提升这类事件的收益。

## 二、指数成分股的定期调整

首先，我们对指数成分股定期调整的规则进行梳理，并给出我们对 2020 年 12 月沪深 300 指数和中证 500 指数成分股调整的预测结果。

### 2.1、成分股调整规则

指数定期调整的流程大致可以分为三个步骤，即数据准备，确定样本空间，筛选样本股。

#### 数据准备

指数公司一般在每年 5 月和 11 月的下旬审核沪深 300 和中证 500 指数样本股，样本调整时间分为是每年 6 月和 12 月的第二个星期五的下一个交易日。每年 6 月的样本定期调整的数据考察区间为上一年度 5 月 1 日至审核年度 4 月 30 日（期间新上市股票为上市第四个交易日以来）；每年 12 月的数据考察区间为上一年度 11 月 1 日至审核年度 10 月 31 日（期间新上市股票为上市第四个交易日以来）。

#### 确定样本空间

沪深 300 和中证 500 的指数样本空间是一致的，由同时满足以下条件的沪深 A 股组成：

- 非创业板股票：上市超过一个季度，除非该股票自上市以来日均 A 股总市值在全部沪深 A 股（非创业板股票）中排在前 30 位；
- 创业板股票：上市超过三年
- 非 ST，\*ST 股票，非暂停上市股票

值得注意的是，公开的指数编制方案没有对科创板股票进行详细的考察规则，实际上的指数编制时也暂从未纳入过科创板股票。考虑到科创板的股票投资风险并不亚于创业板股票，其考察规则不应弱于创业板股票，故当前指数定期调整也暂不考虑科创板股票。

在此样本空间上，指数样本的审核还需考虑公司的经营状况、是否存在违法违规事件以及是否存在市场操纵行为。此外，沪深 300 原则上不将财务亏损的公司纳入指数。指数公司关于这一部分的考察规则并没有给出明细的量化考察指标，因此相关指标选择的差异将有可能影响到预测成分股的准确性。

## 筛选样本股

表 1: 沪深 300 和中证 500 的定期调整的筛选过程

具体步骤	沪深 300	中证 500
数据处理	计算样本空间内股票最近一年（新股为上市第四个交易日以来）的 A 股日均成交金额与 A 股日均总市值	
流动性筛选	选取样本空间日均成交额前 <b>50%</b> 的股票（原样本日均成交额前 60% 即可）	选取样本空间日均成交额前 <b>80%</b> 的股票（原样本日均成交额前 90% 即可）
市值规模选取	将通过流动性筛选的样本股票按照日均总市值的从大到小排序，排名前 <b>240</b> 名的新样本优先进入指数，排名前 <b>360</b> 名的老样本优先保留	将通过流动性筛选的样本股票按照日均总市值的从大到小排序，排名前 <b>400</b> 名的新样本优先进入指数，排名前 <b>600</b> 名的老样本优先保留
入选规定	若入选名单超过样本调整上限，则按照总市值顺序选取上限数量的股票入选（沪深 300 为 30 只，中证 500 为 50 只，但沪深 300 往期通常只保留优先进入指数的样本，数量均在 15 只-25 只之间）	
剔除规定	剔除数量取决于按照上述规则入选的股票数量，剔除顺序为上述规则中自上而下被剔除出考察空间的老样本	

资料来源：中证指数公司官网，招商证券

## 2.2、2020 年 12 月定期调整预测

根据上述规定，我们依据市场上能获得的数据对本次定期调整做出如下预测：

表 2: 沪深 300 指数成分股调整预测

纳入股票代码	纳入股票简称	上市日期	剔除股票代码	剔除股票简称	上市日期
002049.SZ	紫光国微	2005-6-6	000627.SZ	天茂集团	1996-11-12
002384.SZ	东山精密	2010-4-9	000709.SZ	河钢股份	1997-04-16
002414.SZ	高德红外	2010-7-16	002468.SZ	中通快递	2010-9-08
002600.SZ	领益智造	2011-7-15	600038.SH	中直股份	2000-12-18
002812.SZ	恩捷股份	2016-9-14	600089.SH	特变电工	1997-6-18
002821.SZ	凯莱英	2016-11-18	600188.SH	兖州煤业	1998-7-1
300308.SZ	中际旭创	2012-4-10	600170.SH	上海建工	1998-6-23
300529.SZ	健帆生物	2016-8-2	600219.SH	南山铝业	1999-12-23
300676.SZ	华大基因	2017-7-14	600372.SH	中航电子	2001-7-6
600161.SH	天坛生物	1998-6-16	600398.SH	海澜之家	2000-12-28
600584.SH	长电科技	2003-6-3	600516.SH	方大炭素	2002-8-30
600600.SH	青岛啤酒	1993-8-27	600583.SH	海油工程	2002-2-5



纳入股票代码	纳入股票简称	上市日期	剔除股票代码	剔除股票简称	上市日期
600763.SH	通策医疗	1996-10-30	600674.SH	川投能源	1993-9-24
600845.SH	宝信软件	1994-3-11	600867.SH	通化东宝	1994-8-24
600872.SH	中炬高新	1995-1-24	600928.SH	西安银行	2019-3-1
600918.SH	中泰证券	2020-6-3	600968.SH	海油发展	2019-6-26
601696.SH	中银证券	2020-2-26	601018.SH	宁波港	2010-9-28
601872.SH	招商轮船	2006-12-1	601298.SH	青岛港	2019-1-21
601990.SH	南京证券	2018-6-13	601577.SH	长沙银行	2018-9-26
603087.SH	甘李药业	2020-6-29	601828.SH	美凯龙	2018-1-17
603195.SH	公牛集团	2020-2-6	601898.SH	中煤能源	2008-2-1
603233.SH	大参林	2017-7-31	601997.SH	贵阳银行	2016-8-16
603392.SH	万泰生物	2020-4-29	603260.SH	合盛硅业	2017-10-30

资料来源：Wind，招商证券

表 3：中证 500 指数成分股调整预测

纳入股票代码	纳入股票简称	上市日期	剔除股票代码	剔除股票简称	上市日期
000540.SZ	中天金融	1994-2-2	601898.SH	中煤能源	2008-2-1
000709.SZ	河钢股份	1997-4-16	601828.SH	美凯龙	2018-1-17
000958.SZ	东方能源	1999-12-23	601298.SH	青岛港	2019-1-21
002299.SZ	圣农发展	2009-10-21	600674.SH	川投能源	1993-9-24
002409.SZ	雅克科技	2010-5-25	600398.SH	海澜之家	2000-12-28
002459.SZ	晶澳科技	2010-8-10	000025.SZ	特力 A	1993-6-21
002468.SZ	申通快递	2010-9-8	000712.SZ	锦龙股份	1997-4-15
002568.SZ	百润股份	2011-3-25	000785.SZ	居然之家	1997-7-11
002683.SZ	宏大爆破	2012-6-12	000800.SZ	一汽解放	1997-6-18
002705.SZ	新宝股份	2014-1-21	000848.SZ	承德露露	1997-11-13
002791.SZ	坚朗五金	2016-3-29	000987.SZ	越秀金控	2000-7-18
002966.SZ	苏州银行	2019-8-2	002118.SZ	紫鑫药业	2007-3-2
002985.SZ	北摩高科	2020-4-29	002127.SZ	南极电商	2007-4-18
300223.SZ	北京君正	2011-5-31	002285.SZ	世联行	2009-8-28
300463.SZ	迈克生物	2015-5-28	002385.SZ	大北农	2010-4-9
300699.SZ	光威复材	2017-9-1	002407.SZ	多氟多	2010-5-18
600038.SH	中直股份	2000-12-18	002414.SZ	高德红外	2010-7-16
600089.SH	特变电工	1997-6-18	002439.SZ	启明星辰	2010-6-23
600131.SH	国网信通	1998-4-2	002491.SZ	通鼎互联	2010-10-21

敬请阅读末页的重要说明

Page 7

“慧博资讯”专业的投资研究大数据分享平台

点击进入  <http://www.hibor.com.cn>

600188.SH	兖州煤业	1998-7-1	002665.SZ	首航高科	2012-3-27
600219.SH	南山铝业	1999-12-23	002821.SZ	凯莱英	2016-11-18
600372.SH	中航电子	2001-7-6	300197.SZ	铁汉生态	2011-3-29
600398.SH	海澜之家	2000-12-28	300251.SZ	光线传媒	2011-8-3
600516.SH	方大炭素	2002-8-30	300253.SZ	卫宁健康	2011-8-18
600517.SH	国网英大	2003-10-10	300558.SZ	贝达药业	2016-11-7
600546.SH	山煤国际	2003-7-31	600058.SH	五矿发展	1997-5-28
600556.SH	天下秀	2001-8-7	600125.SH	铁龙物流	1998-5-11
600764.SH	中国海防	1996-11-4	600138.SH	中青旅	1997-12-3
600867.SH	通化东宝	1994-8-24	600161.SH	天坛生物	1998-6-16
600871.SH	石化油服	1995-4-11	600298.SH	安琪酵母	2000-8-18
600685.SH	中船防务	1993-10-28	600335.SH	国机汽车	2001-3-5
600928.SH	西安银行	2019-3-1	600428.SH	中远海特	2002-4-18
600956.SH	新天绿能	2020-6-29	600478.SH	科力远	2003-9-18
600968.SH	海油发展	2019-6-26	600521.SH	华海药业	2003-3-4
601298.SH	青岛港	2019-1-21	600565.SH	迪马股份	2002-7-23
601456.SH	国联证券	2020-7-31	600694.SH	大商股份	1993-11-22
601778.SH	晶科科技	2020-5-19	600751.SH	海航科技	1996-9-9
601828.SH	美凯龙	2018-1-17	600757.SH	长江传媒	1996-10-3
601992.SH	金隅集团	2011-3-1	600770.SH	综艺股份	1996-11-20
601997.SH	贵阳银行	2016-8-16	600777.SH	新潮能源	1996-11-21
603005.SH	晶方科技	2014-2-10	600845.SH	宝信软件	1994-3-11
603260.SH	合盛硅业	2017-10-30	600874.SH	创业环保	1995-6-30
603290.SH	斯达半导	2020-2-4	600996.SH	贵广网络	2016-12-26
603345.SH	安井食品	2017-2-22	601019.SH	山东出版	2017-11-22
603589.SH	口子窖	2015-6-29	601678.SH	滨化股份	2010-2-23
603638.SH	艾迪精密	2017-1-20	601811.SH	新华文轩	2016-8-8
603719.SH	良品铺子	2020-2-24	603025.SH	大豪科技	2015-4-22
603737.SH	三棵树	2016-6-3	603233.SH	大参林	2017-7-31
603893.SH	瑞芯微	2020-2-7	603517.SH	绝味食品	2017-3-17
603927.SH	中科软	2019-9-9	603939.SH	益丰药房	2015-2-17

资料来源：Wind，招商证券



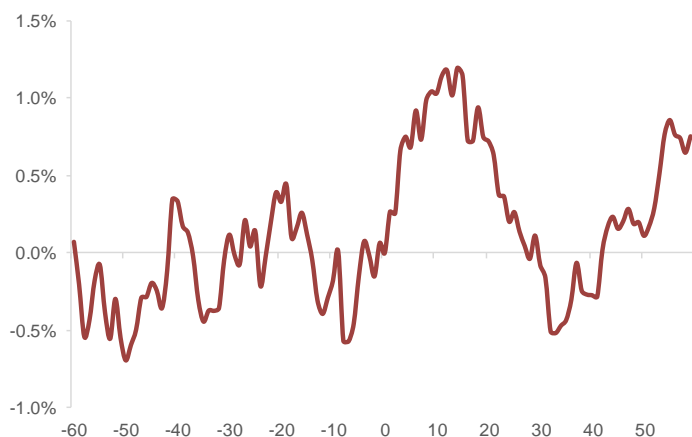
### 三、事件效应分析及增强策略

在前言中我们提到，指数成分股定期调整会对相关样本股带来一定的影响。通常来说，当新的成分股被纳入指数时（除了沪深 300 老样本被纳入至中证 500 新样本以外），定期调整对于新成分股是一个短期利好事件。相反地，当老样本股被剔除时（除了中证 500 老样本被纳入至沪深 300 新样本以外），定期调整对于老成分股是一个短期利空事件。那么指数成分股定期调整对相关成分股的影响究竟有多大呢？下面，我们先详细地分析指数成分股定期调整的事件效应。

#### 3.1、事件分析

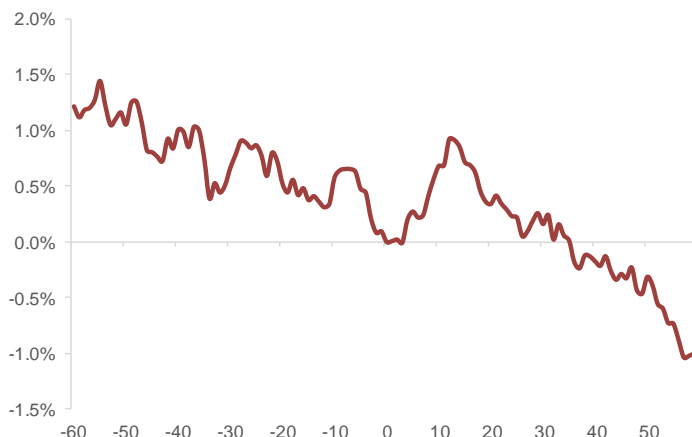
首先，我们先对过去指数成分股定期调整事件的冲击进行统计分析，结果如下图 4 和图 5 所示。其中， $t$  时刻为每一期定期调整的公告日，我们将  $t$  时刻设为收益零点。可以看到，样本股在被纳入沪深 300 和中证 500 后的 10 日之内存在明显的超额收益，而在之后的十多个交易日后超额收益会被逐渐消化。短期冲击使成分股价格呈现先上升后回落的走势。公告日未来 10 个交易日这一期间的样本股的超额收益在一定程度上代表了股票受到纳入效应的短期影响。

图 4：沪深 300 历史纳入效应的超额收益表现



资料来源：Wind，招商证券，统计日期 20070101-20201031

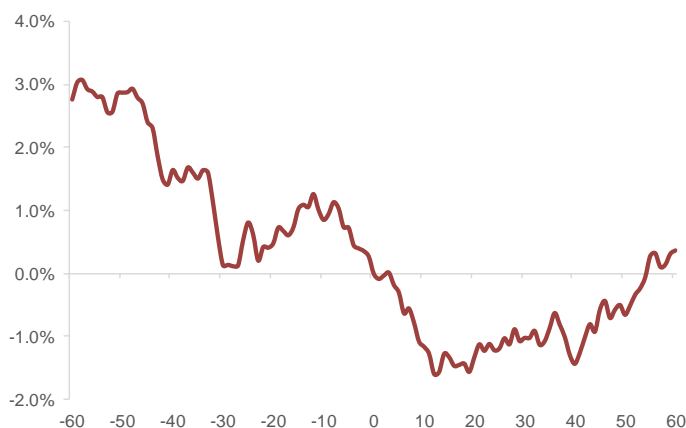
图 5：中证 500 历史纳入效应的超额收益表现



资料来源：Wind，招商证券，统计日期 20070101-20201031

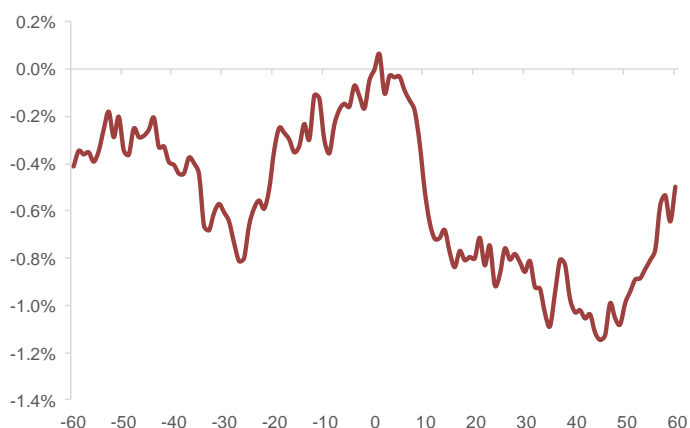
同样地，我们对指数成分股剔除的冲击进行统计，结果如图 6 和图 7 所示。从结果来看，剔除效应导致的收益表现变化则更加显著。从指数成分股中剔除后，个股超额收益公告日未来 15 日内呈现出了明显的下滑趋势，这一期间的样本超额收益在一定程度上代表了股票受到剔除效应的短期影响。

图 6: 沪深 300 历史剔除效应的超额收益表现



资料来源: Wind, 招商证券, 统计日期 20070101-20201031

图 7: 中证 500 历史剔除效应的超额收益表现



资料来源: Wind, 招商证券, 统计日期 20070101-20201031

### 3.2、成分股调整事件的增强策略

事实上, 由于受到总市值、占指数成分股的权重和流动性等方面的影响, 不同的样本股受到定期调整事件的影响存在一定的差异。并且由于影响因素的多维化及非线性化, 导致我们很难用传统的单一指标来线性判断样本股受冲击的影响程度的大小。因此, 我们采用了能够有效处理复合问题的随机森林模型来预测成分股调整结果发布后相关个股的超额收益。

股票价格是多方面因素所共同决定的。这意味着从某一个角度去考察股价将一定会充满大量的噪音信息。若使用单一机器学习模型进行拟合, 容易过多地学习到训练样本中的噪音, 出现过拟合的情况。随机森林作为一种有监督的集成学习方法, 其集合了多个弱分类器的投票结果, 具备较强的抗噪声能力。随机森林算法保证了较好的精准度, 对于非线性的复杂问题拥有较好的泛化能力。

### 3.3、数据准备

#### ● 样本空间

2007 年以来每年 6 月和 12 月定期调整的新旧样本股。

#### ● 模型设定

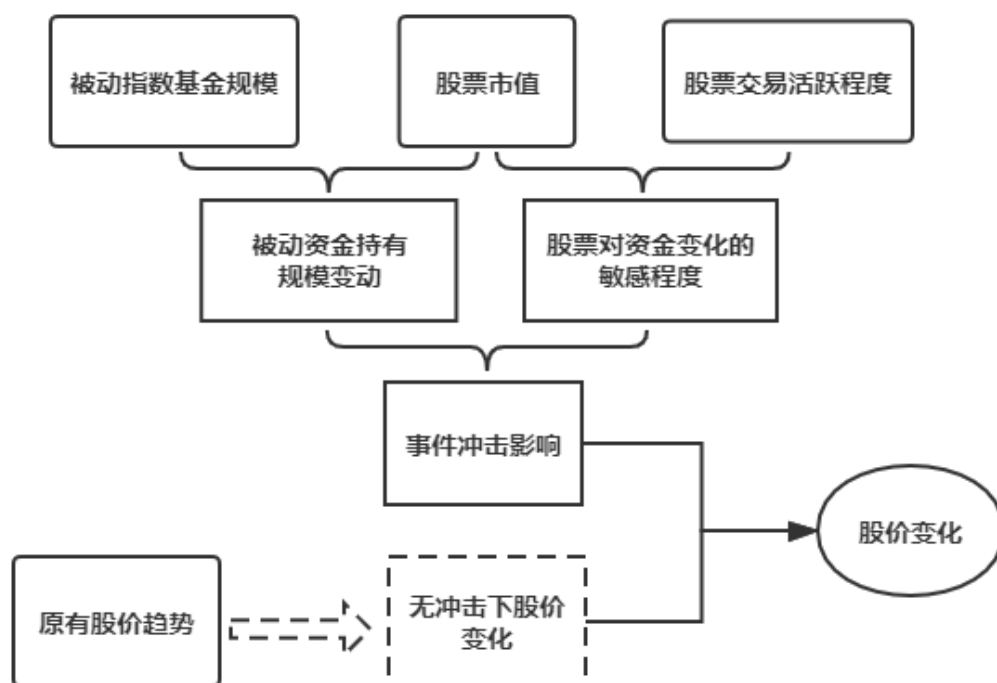
因为样本调整事件学习样本有限, 故采用扩展窗口利用随机森林方法拟合模型。即在每一期定期调整的历史截面上, 利用所有过去定期调整成分股作为训练集建立模型, 用以预测当期成分股调整的影响。

#### ● 特征的选取

对于机器学习模型而言, 特征值的选择的合理有效性直接决定了模型的泛化能力。我们可以从很多方面考虑样本股调整的冲击, 一个典型的思考逻辑是被动指数基金将根据指数调整结果相应地改变成分股权重, 进而对相关样本股产生影响。本文我们将根据这一

逻辑考察事件冲击影响，并从该逻辑的三个维度去选取特征值：（1）被动资金持有规模变动（2）股票交易流动性（3）公司市值规模（4）股票价格趋势

图 8：被动指数基金调整持仓引发事件冲击逻辑图

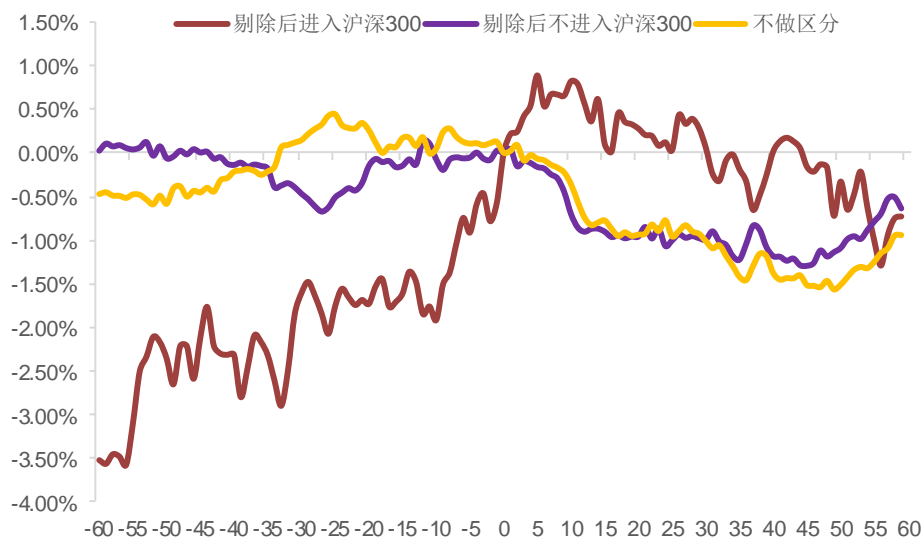


资料来源：Wind，招商证券

#### ➤ 被动资金持有规模变动

资金增减持规模是指数调整事件影响成分股股价的最重要的因素之一，具体来说，被动指数基金（主要包括指数型基金和指数增强型基金）会根据样本股调整结果，增持新纳入的新成分股，减持被剔除的老成分股。非常直观地，在这一过程中被动指数基金增减持导致的基金持有规模的变化将影响成分股的价格。这里值得提醒的是，应当注意股票在被纳入前是否属于其他指数，以及被剔除后是否被编入另一指数。例如，当股票被中证 500 指数剔除后，其被纳入沪深 300 指数与其不被纳入沪深 300 指数的股票相比，受到事件冲击的影响是存在着的较大的差异，如图 9 所示。

图 9：被中证 500 指数剔除后是否进入沪深 300 的分组超额收益情况



资料来源：Wind，招商证券

### ➤ 公司市值规模

公司市值规模大致从两个方面影响指数调整事件冲击，一方面，公司市值能够体现股票承受被动资金增减持的能力。另一方面，公司市值的顺序也体现了股票入选和剔除指数的次序，在一定程度上代表了股票入选指数的前期可预测性。

### ➤ 股票交易活跃度

股票的成交额体现了该股票在过去一段时间成交的活跃度。当股票的成交额较小，其受到资金增减持时股价变动将会较大。

### ➤ 股票的价格走势

股票自身的价格走势也会在一定程度上起到影响。股票自身价格走势叠加短期事件冲击影响最终形成了公告后价格的走势。此外，随着越来越多短期进行事件交易资金的介入，将短期价格走势纳入进行也能在一定程度上刻画这一方面的影响因素。

此外，以上特征值除了从单个指标来看，特征值之间还会形成复合作用，以非线性的方式进一步影响事件冲击，这也是选择了随机森林这一类机器学习模型来进行事件研究的主要原因之一。按照上述叙述，我们选取如下特征值：

表 4：所选特征值信息

名称	简称	含义	选取原因
净现金流量参考值	net_flow	通过权重得到的资金变化	被动指数基金持有规模变动
总市值次序	mkt_rank	当期总市值逆序排序次序	影响个股权重及对资金的承载能力
20 日均成交额	amt20	过去 20 日成交额简单平均	股价近期交易的活跃度
过去 5 日收益率	revs5	过去 5 日股票的涨跌幅	短期趋势
过去 60 日收益率	rves60	过去 60 日股票的涨跌幅	中期趋势

资料来源：招商证券

其中，净资金变化参考值通过如下步骤计算：

- (1) 将对应指数的指数型基金和指数增强型基金规模加和得到被动指数基金规模值。
- (2) 根据事件影响得到被动资金变化，即当股票被纳入指数时，其被动资金变化为相应指数的被动指数基金规模（正值），当其被剔除时，其被动资金变化为负的对应的被动指数基金规模。
- (3) 用流通股占比和总市值估计股票占指数权重，并将权重和相应的被动资金变化相乘，得到初始资金流量值。
- (4) 根据股票在纳入之前是否属于另一指数，以及剔除后是否属于另一指数来调整初始净资金变化值。例如，当新股被纳入中证 500 时，若其原先是沪深 300 样本股，以初始值减去按照(1)、(2)步得出的被剔除沪深 300 的净资金变化，得到该股票的最终净资金流量参考值。其余同理，若不属于上述情况则不进行调整，以初始资金变化值作为参考值。

### 3.4、模型的构建

#### 建模与 IC 测试

为了保证有一定数量的训练样本，我们从 2011 年 6 月调整开始，分别对沪深 300 纳入效应、中证 500 纳入效应、沪深 300 剔除效应和中证 500 剔除效应在每一个历史截面进行了模型训练和预测<sup>2</sup>。其中，将 t 日（公告日）截面的总市值次序、过去 5 日收益率、过去 60 日收益率、20 日均成交额以及净资金流量参考值作为自变量，将公告日至公告日后 10 日的区间累计超额收益作为模型拟合的因变量。

接下来为了初步检验模型的有效性，我们将模型拟合值作为因子值与实际超额收益值来进行 IC 测试，进一步考察其预测效果。如表 5 所示，纵向来看，各模型的预测准确率尚可，预计能够在一定程度上对指数成分股调整事件的冲击进行量化性的预测。横向来看，沪深 300 指数的纳入事件 IC 结果测试要稍好于剔除事件。究其市场原因，可能是由于沪深 300 指数的纳入事件受到的市场关注程度最大，通过建模对样本股纳入事件效应的提升较为显著。

表 5：IC 测试结果

	300 纳入	300 剔除	500 纳入	500 剔除
201106	-21.47%	-20.18%	33.36%	-10.67%
201112	49.39%	9.06%	12.69%	18.56%
201206	40.60%	-44.65%	4.04%	-6.74%
201212	0.15%	-0.99%	31.64%	33.97%
201306	30.03%	37.00%	30.84%	24.87%
201312	28.42%	-6.61%	9.56%	-4.96%
201406	-16.62%	12.81%	25.69%	4.94%
201412	38.00%	20.24%	25.51%	24.55%
201506	-49.14%	-51.02%	-24.06%	37.29%
201512	45.59%	5.19%	7.83%	12.35%

<sup>2</sup> 考虑到 2009 年 9 月才成立首只中证 500 指数型基金，故其相关模型的训练集样本是从 2010 年后开始的。

敬请阅读末页的重要说明



	300 纳入	300 剔除	500 纳入	500 剔除
201606	-9.92%	-40.07%	-24.65%	10.34%
201612	1.26%	10.74%	2.03%	3.04%
201706	-18.16%	3.13%	-13.97%	-6.48%
201712	15.90%	17.97%	-16.09%	18.26%
201806	-2.83%	1.61%	12.22%	-29.63%
201812	44.63%	-16.81%	12.85%	-4.69%
201906	9.02%	30.77%	4.33%	-22.73%
201912	39.32%	39.26%	-31.58%	10.80%
202006	-14.71%	57.61%	10.77%	-1.79%
IC 均值	11.03%	3.42%	5.95%	5.86%
IC 胜率	70%	70%	75%	75%
IC_IR 值	15.75%	4.89%	7.93%	7.81%

资料来源：Wind，招商证券，统计日期 20070101-20201031

## 特征值排序

进一步地，我们对各截面上训练模型的特征重要度进行了梳理，简单来说，特征重要度代表着这一指标在模型学习中对拟合值拟合过程中起到的重要程度。当特征重要度越大时，说明该指标对于模型的贡献越大。表 6 是沪深 300 指数纳入效应自 2011 年 6 月以来，在每个公告日截面上建模的特征重要度次序。接下来，我们针对几个特征值的重要度排序进行进一步分析。

表 6：时间序列上各特征重要度次序

指标	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
net_value	4	3	5	5	4	3	4	3	3	4	4	4	5	5	5	2	2	2	2
mkt	5	5	4	4	5	4	3	4	4	3	3	5	4	4	4	4	5	5	5
amt20	3	4	3	3	3	5	5	5	5	5	5	3	3	3	3	3	4	4	4
revs5	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
revs60	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	5	3	3	3

资料来源：Wind，招商证券，统计日期 20130601-20200601，次序越小说明该特征值越重要

### ● 5 日收益率（revs5）和 60 日收益率（revs60）

5 日收益率在 2015 年后的特征重要度次序排到了第一名，而 60 日收益率同期降到了第二名，且在后续的时间内排名仍有下降。事实上，在早些年份，相关被动基金的整体规模不高，指数成分股调整受关注度不高。股票受到的事件效应影响较小，股票自身中期走势更为重要。60 日收益率能够在一定程度上代表股票过去趋势。近年来，由于部分投资者开始逐渐关注指数调整可能带来的收益，并在公告日前提前布局成分股调整，导致近年来股票在公告日前就开始出现异动。这一部分的信息将会被过去 5 日的收益率捕捉。过去 5 日收益率在近年呈现出重要性增强的现象，是模型根据实际市场数据训练学习得到的结果，符合实际市场情况，利于提升模型泛化能力。

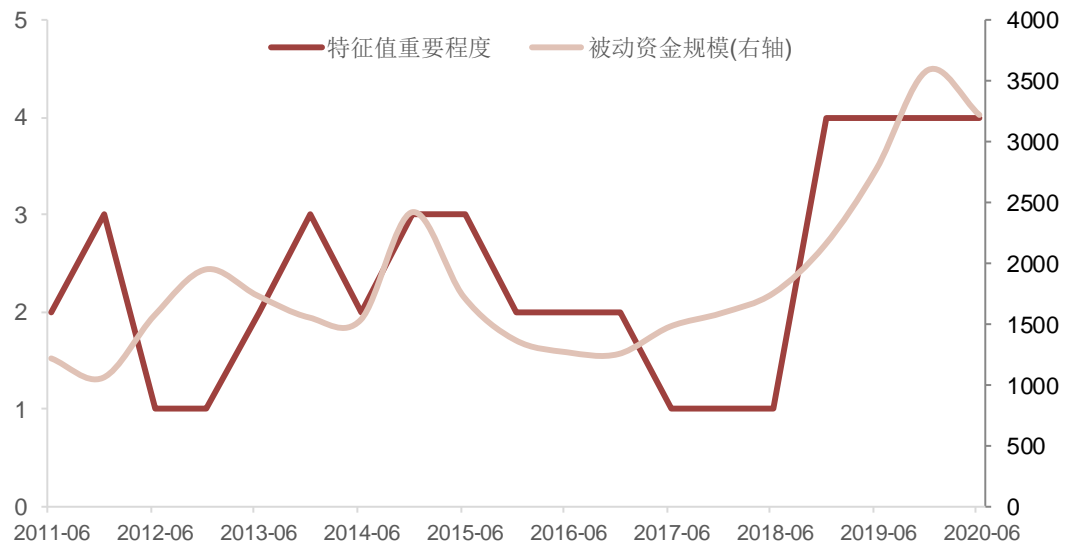
### ● 净资金流量参考值（net\_value）

净资金流量参考值特征重要度与被动指数基金规模在时间序列上呈现出一定的相关性。如图



10 所示, net\_value 在 2013 年特征重要度上升, 到 2015 年之后开始下降, 2019 年开始出现反弹, 并在最近几次指数调整中都排在较高的水平。这与实际市场中, 沪深 300 的被动式投资基金规模在时间序列上的趋势表现较为相似, 即表现出当被动资金较大时, 模型中被动资金对纳入效应的影响也越大。这也说明, 利用该随机森林方法建立模型, 是能够有效的捕捉来自被动资金对纳入效应的影响。

图 10: 净资金流量参考值重要性和被动指数基金规模的变化

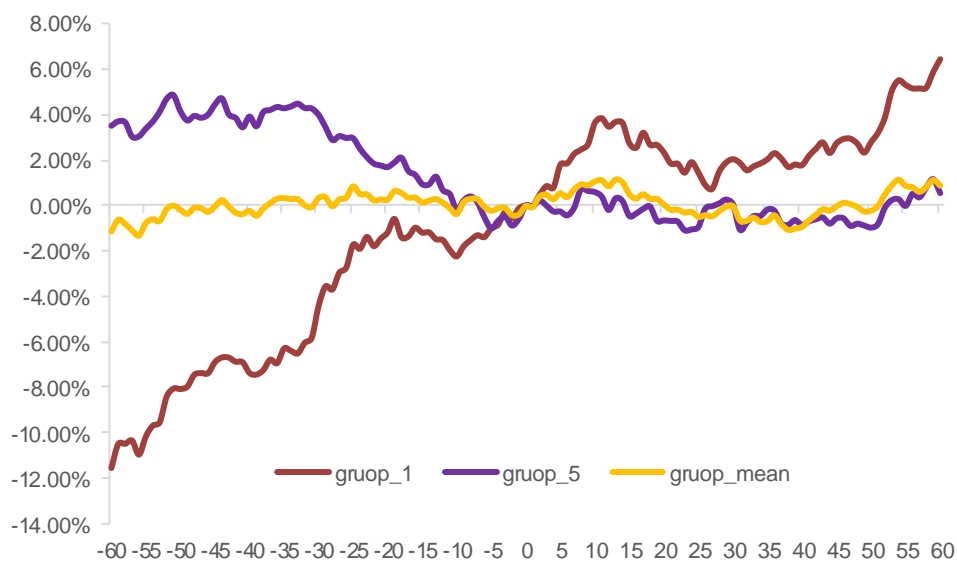


资料来源: Wind, 招商证券, 统计日期 20110601-20200601, 为方便展示, 图中特征值重要程度为逆序次序, 即纵轴值越高, 代表特征值此时越重要

### 3.5、模型结果展示

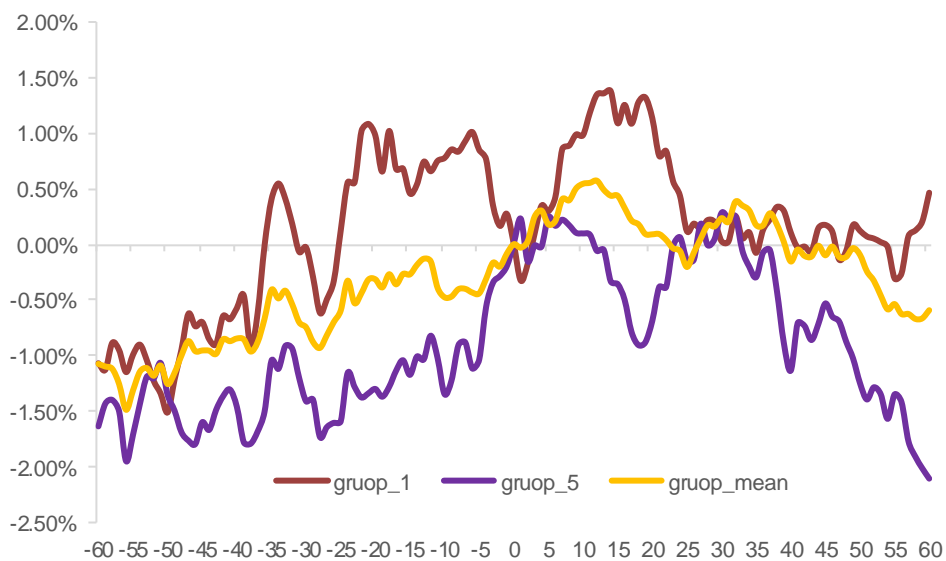
接下来, 我们在最新截面上得到收益预测估计后, 按照模型预测值从大到小分为五个组, 分别考察在公告日前后 60 个交易日内超额收益表现情况, 进一步地评估模型预测的市场效果。如图 11 和图 12 所示, 在沪深 300 中, 呈现出预测受到纳入效应影响最大的第一组收益表现要高于平均水平和第五组, 而在中证 500 中, 第一组在公告日后的 30 个交易日内, 收益表现同样优于平均水平和第五组。其中, 沪深 300 中的第一组在公告日后 10 日、20 日和 60 日的收益率达到 3.64%、2.33% 和 6.45%, 均表现出了显著的相对超额收益。中证 500 中的第一组则在公告日后的 10 日和 20 日的超额收益更为显著。

图 11: 沪深 300 纳入效应的分组超额收益表现



资料来源: Wind, 招商证券, 统计日期 20070101-20201031

图 12: 中证 500 纳入效应的分组超额收益表现



资料来源: Wind, 招商证券, 统计日期 20100101-20201031

表 7: 纳入效应的详细分组表现

指标	t+10 收益率	T 统计量	t+20 收益率	T 统计量	t+60 收益率	T 统计量
沪深 300 group1	3.64%	4.22 ***	2.33%	1.80 *	6.45%	1.93 *
沪深 300 group5	0.60%	0.83	-0.66%	-0.57	0.55%	-0.02
中证 500 group1	0.98%	1.62	1.14%	1.31	0.46%	0.30
中证 500 group5	0.09%	0.22	-0.70%	-1.13	-2.11%	-2.03 **

资料来源: Wind, 招商证券, 统计日期 20070101-20201031, \*代表通过 10%的显著性检验, \*\*代表通过 5%的显著性检验, \*\*\*代表通过 1%的显著性检验

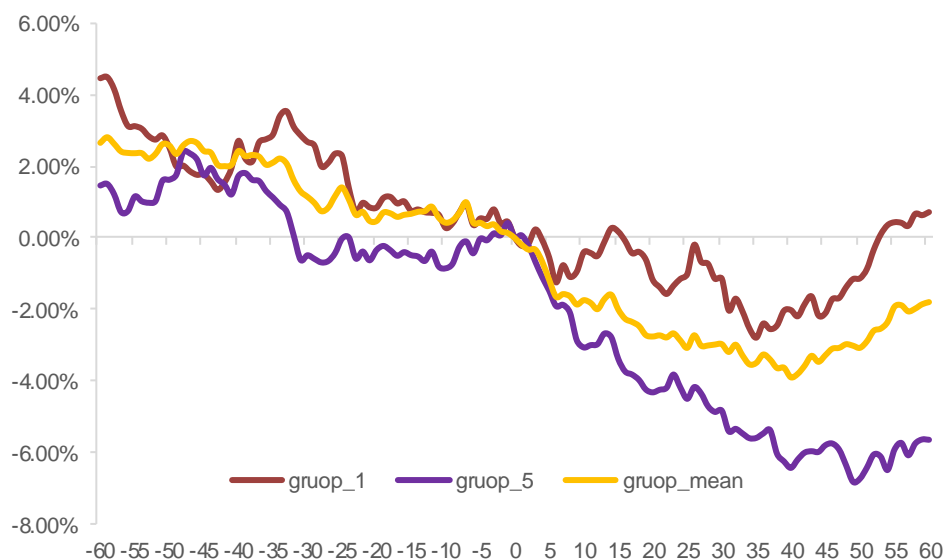
敬请阅读末页的重要说明

“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

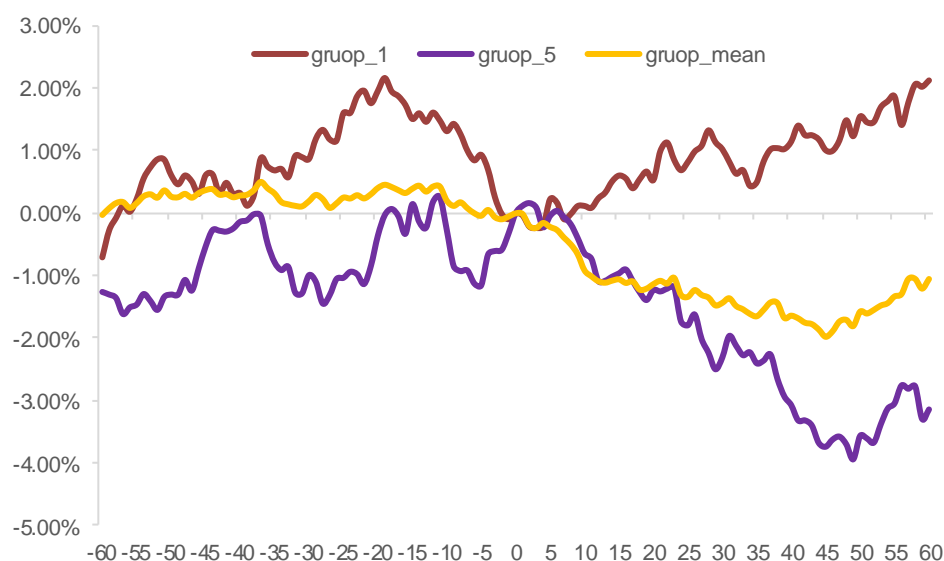
图 13 和图 14 所展示的模型对剔除效应的预测。总体而言，两个指数根据预测估计得到的第一组都明显地优于平均水平，而第五组相对基准都呈现出较差的表现。其中，中证 500 的剔除效应分组表现分化更为显著，第一组在公告日后 10 日、20 日和 60 日收益分别达到了 0.12%、0.53% 和 2.11%，所有个股的平均收益分别为 -0.91%、-1.22% 和 -1.05%，而第五组的收益分别为 -0.64%、-1.23% 和 -3.13%。

图 13: 沪深 300 剔除效应的分组超额收益表现



资料来源: Wind, 招商证券, 统计日期 20070101-20201031

图 14: 中证 500 剔除效应的分组超额收益表现



资料来源: Wind, 招商证券, 统计日期 20100101-20201031

表 8: 剔除效应的详细分组表现

指标	t+10 收益率	T 统计量	t+20 收益率	T 统计量	t+60 收益率	T 统计量
沪深 300 group1	-1.63%	-1.80	-1.79%	-1.51	0.71%	0.35
沪深 300 group5	-2.94%	-3.85 ***	-5.30%	-4.93 ***	-5.65%	-3.03 ***
中证 500 group1	0.12%	0.15	0.53%	0.46	2.11%	1.21
中证 500 group5	-0.64%	-1.11	-1.23%	-1.51	-3.13%	-2.68 ***

资料来源: Wind, 招商证券, 统计日期 20070101-20201031, \*代表通过 10%的显著性检验, \*\*代表通过 5%的显著性检验, \*\*\*代表通过 1%的显著性检验

## 四、结语

随着以 ETF 为代表的被动指数产品规模的不断壮大, 指数成分股调整事件给相关个股带来的冲击越来越大。与此同时, 越来越多的资金开始参与到这类事件中来获取收益。本文一方面根据沪深 300 和中证 500 这两个重要指数的编制规则, 对 2020 年 12 月指数成分股定期调整的结果进行了预测。另外一方面, 我们综合了个股总市值、被动资金持有规模变化、流动性、中期走势、短期走势等诸多因素, 并基于机器学习的方法对调整公告后个股的收益进行进一步建模, 并显著提升了原有事件的收益。

后续我们将会定期发布指数成分股调整的预测结果, 也请各位投资者继续关注我们的后续的研究!

### 风险提示

本文的实证是基于历史数据得到的, 当市场环境发生变化的时候, 存在模型失效的风险, 文章结论不构成投资建议。

## 分析师承诺

负责本研究报告全部或部分内容的每一位证券分析师，在此申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告中的具体推荐或观点直接或间接相关。

**任瞳：**首席分析师，定量研究团队负责人，管理学硕士，16 年证券研究经验，2010 年、2015 年、2016、2017 年新财富最佳分析师。在量化选股择时、基金研究以及衍生品投资方面均有深入独到的见解。

**周靖明：**高级量化分析师，武汉大学金融工程硕士，5 年量化策略研究开发经验。研究方向是量化选股与创新基金产品研究，在多因子选股，SmartBeta 产品等方面有长期深入的研究。

## 投资评级定义

### 公司短期评级

以报告日起 6 个月内，公司股价相对同期市场基准（沪深 300 指数）的表现为标准：

强烈推荐：公司股价涨幅超基准指数 20%以上

审慎推荐：公司股价涨幅超基准指数 5-20%之间

中性：公司股价变动幅度相对基准指数介于±5%之间

回避：公司股价表现弱于基准指数 5%以上

### 公司长期评级

A：公司长期竞争力高于行业平均水平

B：公司长期竞争力与行业平均水平一致

C：公司长期竞争力低于行业平均水平

### 行业投资评级

以报告日起 6 个月内，行业指数相对于同期市场基准（沪深 300 指数）的表现为标准：

推荐：行业基本面向好，行业指数将跑赢基准指数

中性：行业基本面稳定，行业指数跟随基准指数

## 重要声明

本报告由招商证券股份有限公司（以下简称“本公司”）编制。本公司具有中国证监会许可的证券投资咨询业务资格。本报告基于合法取得的信息，但本公司对这些信息的准确性和完整性不作任何保证。本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价，在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。除法律或规则规定必须承担的责任外，本公司及其雇员不对使用本报告及其内容所引发的任何直接或间接损失负任何责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突。

本报告版权归本公司所有。本公司保留所有权利。未经本公司事先书面许可，任何机构和个人不得以任何形式翻版、复制、引用或转载，否则，本公司将保留随时追究其法律责任的权利。

敬请阅读末页的重要说明

Page 19

“慧博资讯”专业的投资研究大数据分享平台

点击进入  <http://www.hibor.com.cn>