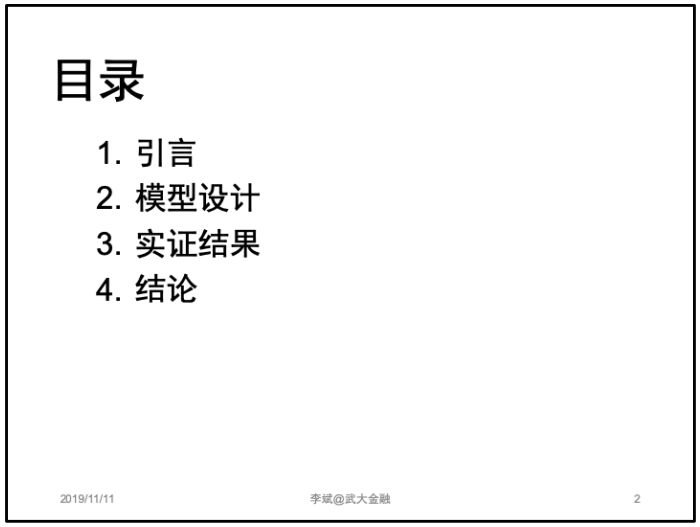
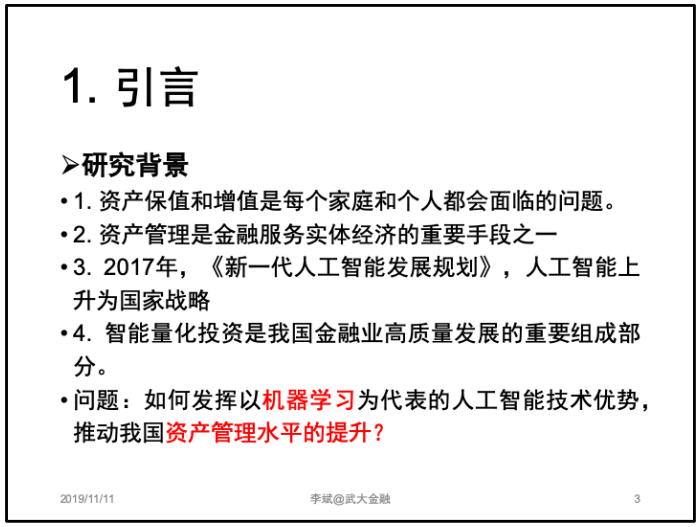
第一页，说明展示的文献名称、作者、期刊（工作论文写 “Working Paper”）、年份等；下方是解读者和解读日期



目录的结构：

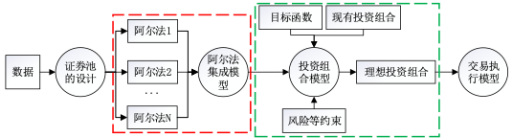
1. 引言：介绍背景、研究动机、研究问题、研究内容概述、创新点和贡献



背景从现实出发，研究通常从现实问题出发。条目化列出来，不要大段拷贝问题，字体大小最小为20号字体，建议正文采用24号字体（强迫自己总结，利用简洁的文字表达）。

1. 引言

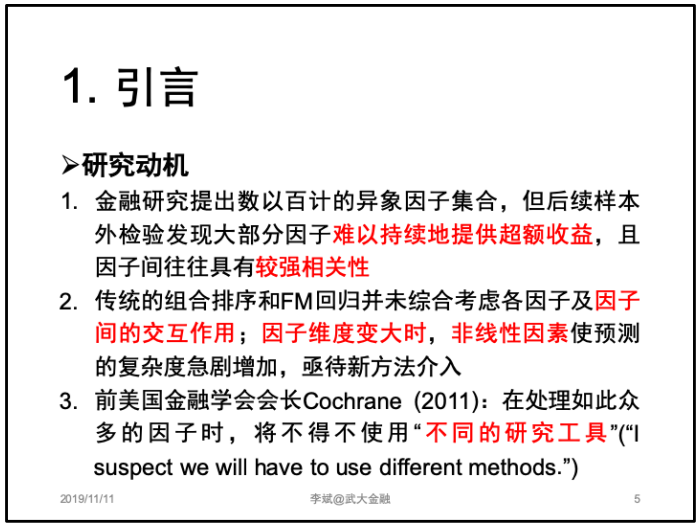
➢研究背景



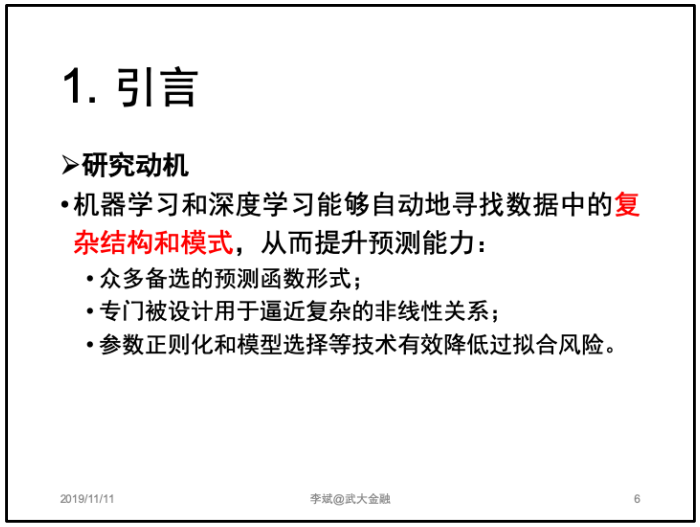
其中，红色框内为资产定价模块；绿色框内为投资组合模块。

- **基本面量化投资融合了量化投资（算法驱动）与基本面投资（人为驱动），从异象因子出发集成能够提供超额收益的阿尔法信号并构建投资组合。**
- **部分投资者表现出“算法厌恶”，根源在于对量化投资的机理和决策机制缺乏深入的了解，亟待开展对智能量化投资方面的研究。**

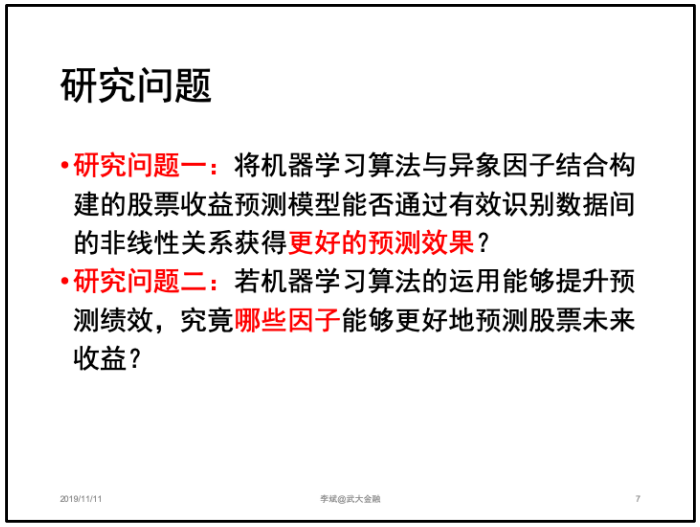
逐步从现实问题步入研究问题。从大到小，研究问题来源于现实问题，同时逐渐抽象化。



研究动机从研究问题所面临的挑战出发，突出为什么要去研究某一个具体的点？通过动机，你能否说服自己去解决所面临的研究挑战？说服不了自己，认真想和读。



研究动机2，用新方法解决老问题通常需要说明新方法相对老方法的优势。



在说明研究的动机之后，提出研究问题。一篇文章有1-2个研究问题。后面的实证/理论都是为了解决研究问题而服务，要围绕着研究问题展开。本文有2个研究问题，所以后面分了两章。如果只有1个，分一章也可以。

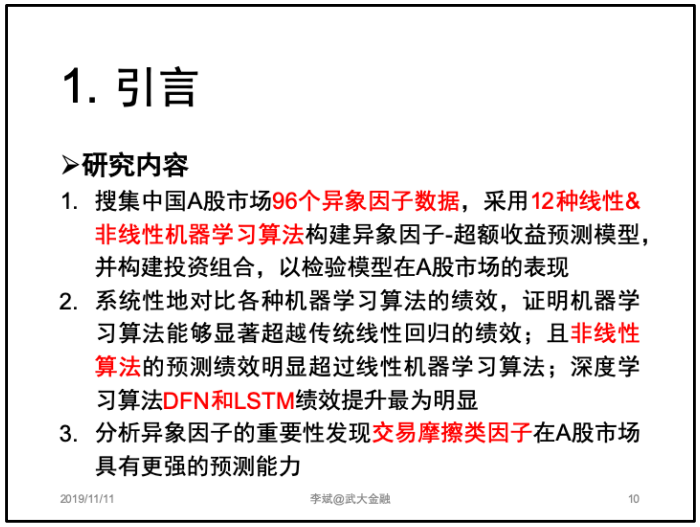
现有研究是如何解决研究问题，主要阅读文章的文献综述部分。尽量总结。

1. 引言

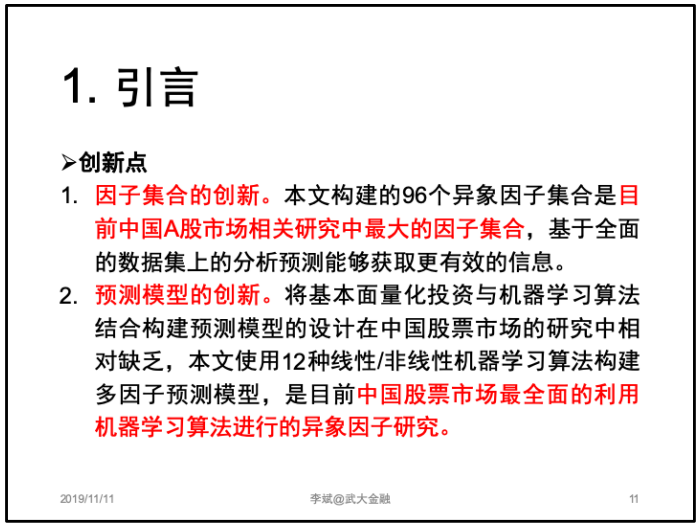
➢现有研究

2. 异象因子研究

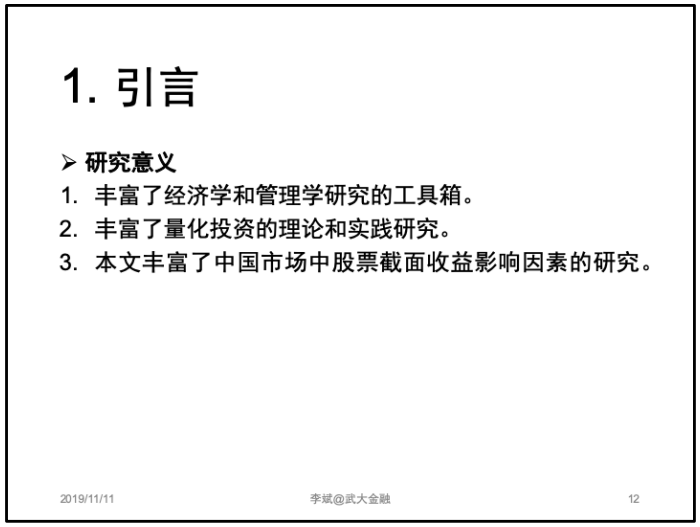
- Green，Hand和Zhang（2017)采用Fama-MacBeth回归检验了1980-2014年美国股票市场的94项异象因子，发现12个因子对股票月度收益具有显著的预测效果。
- Light，Maslov，and Rytchkov（2017)采用“偏最小二乘法”（PLS）来检验公司特征对期望收益的预测能力；
- Hou et al. (2019)检验了447个文献中的异象因子，发现大部分并不能预测股票未来收益；
- Jiang等（2019)分别采用Fama-MacBeth回归、PCA、PLS和FC方法整合A股市场中的75个异象因子，证明上述线性方法能够从因子中提取出有助于预测的信息。



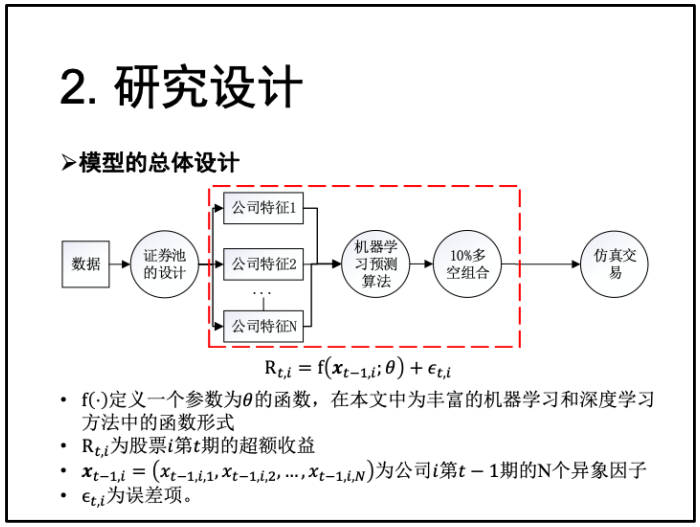
1. 所读文献是如何解决研究问题？概述研究内容，不要超过1页。用大字，强迫自己总结。
2. 解决研究问题所得最重要的结论是什么？概述，1-2个条目结课，用24号以上的大字。



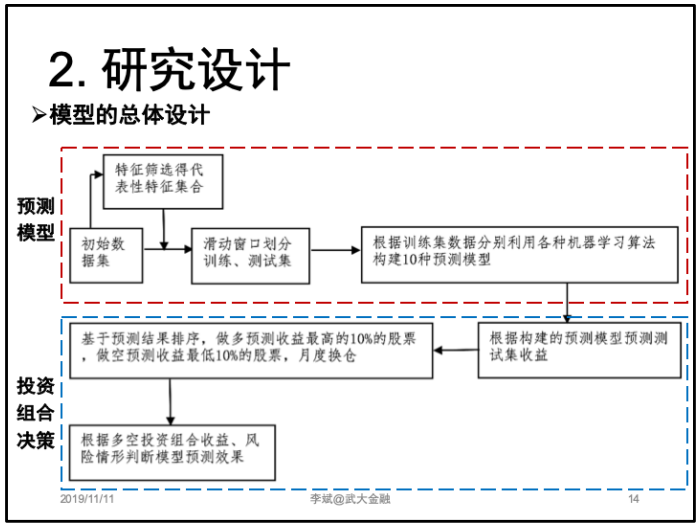
总结文章的创新点。文章都会讲，提取出来即可。



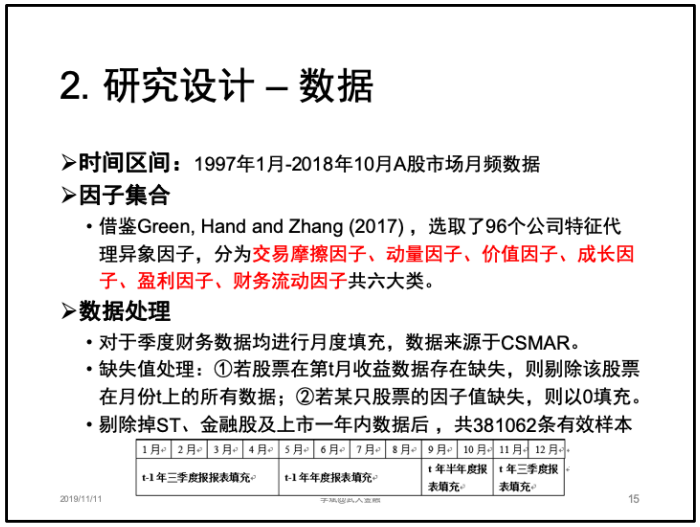
这里是研究意义或者贡献，主要是从理论上来讲。同样要大字。



第一部分“引言”讲完后，基本上小同行专家都已经了解本文做什么了。
第二部分讲“研究设计”，主要告诉听众，本文解决研究问题的主要途径，通过这一套研究设计检验的结果在第3部分实证结果中。比如通过实证验证的话，说明研究假设、实证研究方法和实证数据等等。本文是讲整套的研究设计。这里需要更加详细化，需要用学术的语言讲。不要随意发挥，很容易被专家看出来。
多想想为什么：为什么推导出本文的假设？是否可能还有其他的推导路径？解决研究问题为什么需要采用本文的研究方法？是否还能采用其他更好或更合适的研究方法？为什么采用本文的实证数据？X和y变量是否会有更好的代理变量？等等。


2. 研究设计
➢模型的总体设计
预测模型
特征筛选得代表性特征集合
初始数据集
滑动窗口划分训练、测试集
根据训练集数据分别利用各种机器学习算法构建10种预测模型
投资组合决策
基于预测结果排序，做多预测收益最高的10%的股票，做空预测收益最低10%的股票，月度换仓
根据构建的预测模型预测测试集收益
根据多空投资组合收益、风险情形判断模型预测效果
2019/11/11
李斌@武大金融
14

继续总体设计。



说明数据的选择：1. 数据的区间、频率、市场等；2. 选取数据的维度；3. 关键的数据处理步骤，本文作了填充，需要特别说明；4. 最后所得数据的统计（样本数等）。同样，每一步多想想为什么？常见的选择原因：1. 最长可用的数据区间、全股票集合等；2. 参考别人的做法（文献）。

2. 研究设计 – 数据

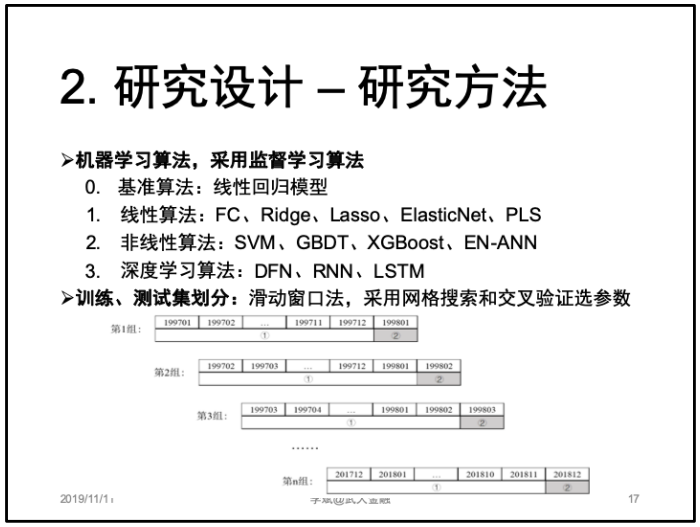
➢ 单因子检验中显著的因子集合（22项）

异象因子	size	std_dvol	LM	lagretn	rd_mve	size_ia	volumed
年化收益率	21.26%	18.18%	16.71%	15.17%	13.93%	12.13%	11.65%
t-statistics	3.8701	8.7984	7.7101	3.8841	5.4255	2.911	5.9414

std_turn	illq	mom36	momchg	chfeps	depr	aeavol	retnmax
11.62%	9.21%	9.07%	9.01%	8.44%	8.09%	6.67%	5.99%
6.5113	4.2123	2.563	2.8739	4.7377	2.1071	3.6105	4.5582

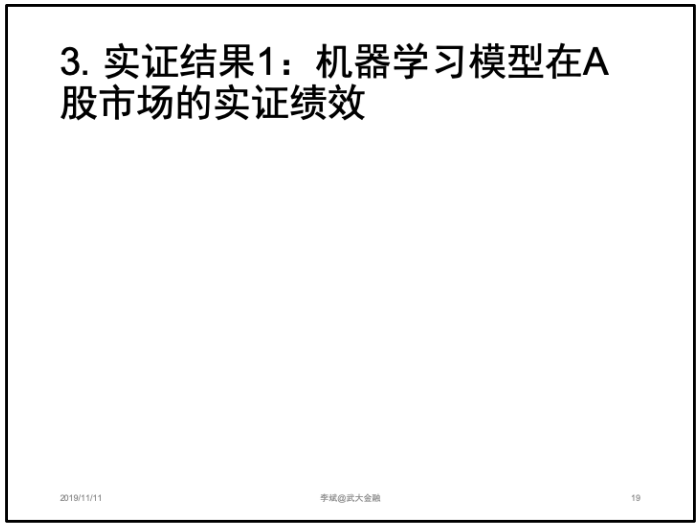
SP	vol	idvol	skew	pchsaleinv	SgINVg	CRG
5.23%	5.00%	4.74%	4.54%	3.70%	3.63%	2.91%
2.0128	2.3778	3.2926	3.774	2.8627	2.924	2.1851

数据的描述性统计，不要超过1页。没啥特殊说明，可以不展示。



1. 说明在前述研究框架中采用的方法。这里只列出了方法，口述说明选取理由，本文的预期结果（或想验证的假设）有三点：1. 线性的表现>基准; 2. 非线性>线性; 3. 深度学习>非线性。
2. 框架中比较重要的滑动窗口方法。

1. 该页说明了第二个研究问题的研究方法，可以放到研究设计中去（论文中是这样写的）。
2. 研究问题2事实上是研究问题1进一步检验的结果，研究问题1的建议在总体设计中已经说明。当然也可以将检验研究问题1和2的方式显式说明，会更好。



“实证结果”是为了通过“研究设计”去解决“研究问题”所给出，通过对实证结果的解读，回答研究问题。注意：论文通常有很多表格，但最重要的表格通常1个研究问题1-2张，其他稳健性等等，都是分支，不是最重要的，文献解读时可以忽略。

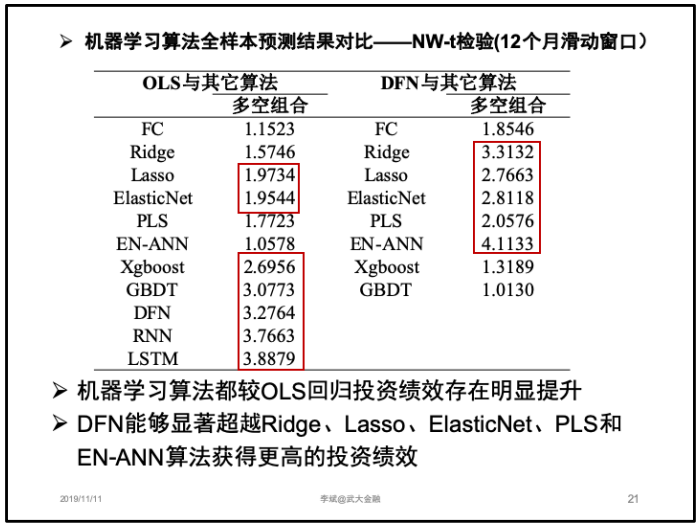
回到本文的实证结果的解读从两个角度展开。
1. 机器学习模型在A股市场的实证绩效
2. 在机器学习的视角下，因子的重要性。

因为下一页的表格太大，因此这边用了一个空白页，以显示这是研究问题1的实证结果。大家做的话，可以将这一页和下一页合并。

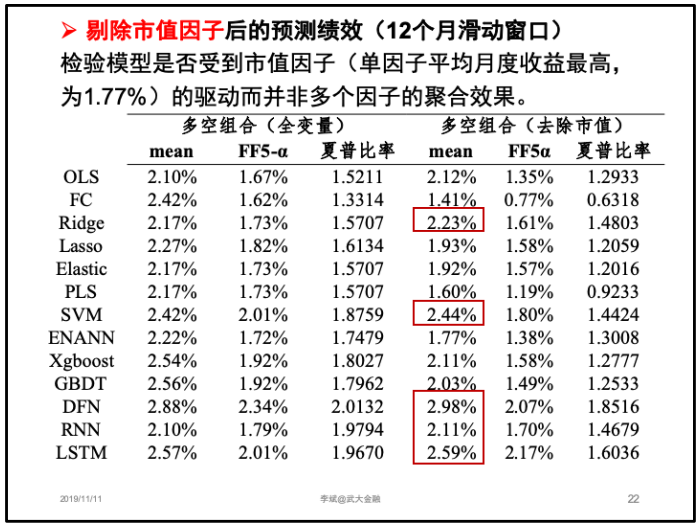
机器学习算法全变量预测结果对比(12月滑动窗口）

	多头组合			多空组合		
	mean	FF5-α	夏普比率	mean	FF5-α	夏普比率
OLS	2.47%	0.90%	0.73	2.10%	1.67%	1.52
FC	2.71%	1.01%	0.80	2.42%	1.62%	1.33
Ridge	2.52%	0.95%	0.75	2.17%	1.73%	1.57
Lasso	2.55%	0.96%	0.76	2.27%	1.82%	1.61
Elastic	2.52%	0.95%	0.75	2.17%	1.73%	1.57
PLS	2.52%	0.95%	0.75	2.17%	1.73%	1.57
SVM	2.60%	1.07%	0.79	2.42%	2.01%	1.87
EN-ANN	2.51%	0.94%	0.76	2.22%	1.72%	1.74
Xgboost	2.69%	0.98%	0.81	2.54%	1.92%	1.80
GBDT	2.71%	1.00%	0.82	2.56%	1.92%	1.79
DFN	2.94%	1.34%	0.88	2.88%	2.34%	2.01
RNN	2.52%	1.10%	0.78	2.10%	1.79%	1.97
LSTM	2.86%	1.18%	0.85	2.57%	2.01%	1.96
SIZE	2.45%	0.51%	0.70	1.73%	0.27%	0.68
MKT	0.61%	-0.01%	0.1795			

1. 解决研究问题1所需要的实证结果，有2个panel，1个是数值，1个是统计检验。
2. 表格比较大，因此分成2张ppt。有必要的话，只截取其中重要的行/列即可，展示结果的含义最重要，细枝末节可以忽略。
3. 严格注意表格中字体的大小，保证可读性。不要小于20号字体，截图的话，将其中的文字与20号字体比较下。
4. 通过不同颜色的框图，将重点突出。



在讲完之后，一定要用文字总结下该表最主要的观察。注意用大字，提炼总结。



➢ **剔除市值因子后的预测绩效（12个月滑动窗口）**

检验模型是否受到市值因子（单因子平均月度收益最高，为1.77%）的驱动而并非多个因子的聚合效果。

	多空组合（全变量）			多空组合（去除市值）		
	mean	FF5-α	夏普比率	mean	FF5α	夏普比率
OLS	2.10%	1.67%	1.5211	2.12%	1.35%	1.2933
FC	2.42%	1.62%	1.3314	1.41%	0.77%	0.6318
Ridge	2.17%	1.73%	1.5707	2.23%	1.61%	1.4803
Lasso	2.27%	1.82%	1.6134	1.93%	1.58%	1.2059
Elastic	2.17%	1.73%	1.5707	1.92%	1.57%	1.2016
PLS	2.17%	1.73%	1.5707	1.60%	1.19%	0.9233
SVM	2.42%	2.01%	1.8759	2.44%	1.80%	1.4424
ENANN	2.22%	1.72%	1.7479	1.77%	1.38%	1.3008
Xgboost	2.54%	1.92%	1.8027	2.11%	1.58%	1.2777
GBDT	2.56%	1.92%	1.7962	2.03%	1.49%	1.2533
DFN	2.88%	2.34%	2.0132	2.98%	2.07%	1.8516
RNN	2.10%	1.79%	1.9794	2.11%	1.70%	1.4679
LSTM	2.57%	2.01%	1.9670	2.59%	2.17%	1.6036

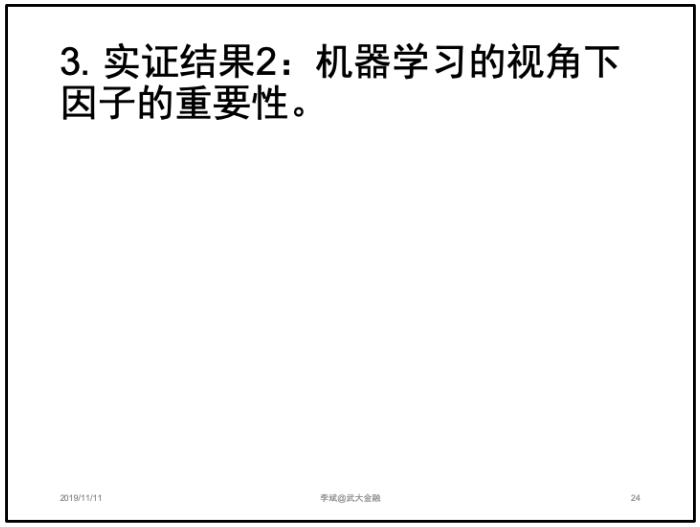
1. 稳健性检验，时间紧可以不要。
2. Ppt上一定要用1-2句话说明做什么和结果说明了什么。

➢ **集成各类算法预测结果**

为了更直观的说明机器学习算法相对于传统线性模型的绩效提升，本文简单加权11种机器学习算法（其中未包括线性组合的FC）构建集成预测模型：$R_{t,i}^{ensemble} = \frac{1}{11}\sum_{j=1}^{11} R_{t,i}^{j}$

	3个月滑动窗口			12个月滑动窗口		
	多空组合	多头组合	空头组合	多空组合	多头组合	空头组合
mean	2.56%	2.60%	0.04%	2.98%	2.89%	-0.09%
FF3-α	2.38%	1.34%	-1.20%	2.67%	1.41%	-1.46%
FF5-α	2.18%	1.20%	-1.21%	2.55%	1.28%	-1.48%
夏普比率	1.4698	0.7715	-0.0651	2.1736	0.8720	-0.1085
	24个月滑动窗口			**36个月滑动窗口**		
	多空组合	多头组合	空头组合	多空组合	多头组合	空头组合
mean	2.56%	2.60%	0.04%	2.98%	2.89%	-0.09%
FF3-α	2.38%	1.34%	-1.20%	2.67%	1.41%	-1.46%
FF5-α	2.18%	1.20%	-1.21%	2.55%	1.28%	-1.48%
夏普比率	1.4698	0.7715	-0.0651	2.1736	0.8720	-0.1085

稳健性说明。说明做了什么？建议再加个结果说明。



同样

为了证明研究问题2所给出的实证结果表。

同时，并非是交易摩擦因子本身占总体比例较大导致

因子类别	因子总数	重要因子数	占比
交易摩擦因子	21	10	48%
财务流动性因子	10	2	20%
动量因子	6	1	17%
盈利因子	14	1	7%
成长因子	35	1	3%
价值因子	10	0	0%

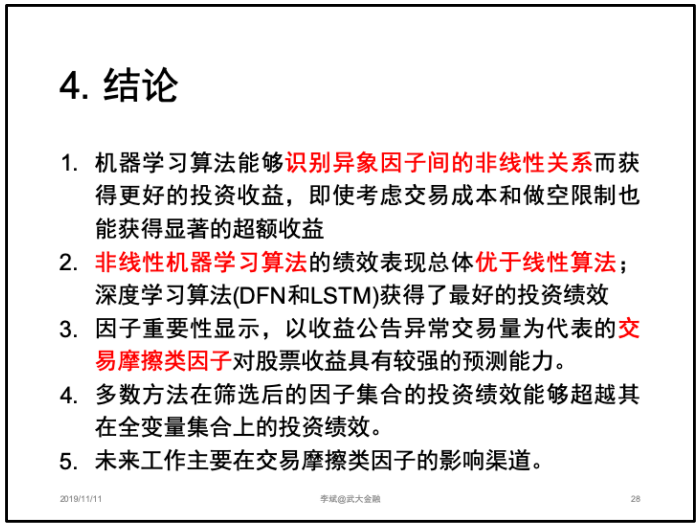
交易摩擦因子具有较强预测能力的原因？有待继续探索。

1. 很可能会有人猜测，是因为交易摩擦类因子比例过大所导致。为了排除这一猜测，继续检验。
2. 标题直接说明了验证的假设。
3. 说明后续并未被解释。红色高亮

➢ 集成结果筛选出代表性因子12个月滑动窗口预测结果

	多空组合（全变量）			多空组合（筛选变量）		
	mean	FF5-α	夏普比率	mean	FF5-α	夏普比率
OLS	2.10%	1.67%	1.52	2.51%	1.78%	1.65
FC	2.42%	1.62%	1.33	3.03%	2.14%	1.47
Ridge	2.17%	1.73%	1.57	2.53%	1.83%	1.59
Lasso	2.27%	1.82%	1.61	2.60%	1.89%	1.60
Elastic	2.17%	1.73%	1.57	2.61%	1.93%	1.61
PLS	2.17%	1.73%	1.57	2.63%	1.92%	1.63
SVM	2.42%	2.01%	1.87	2.75%	2.03%	2.26
EN-ANN	2.22%	1.72%	1.74	2.33%	1.88%	1.84
Xgboost	2.54%	1.92%	1.80	2.61%	2.37%	2.01
GBDT	2.56%	1.92%	1.79	2.65%	2.34%	2.04
DFN	2.88%	2.34%	2.01	3.51%	2.99%	2.68
RNN	2.10%	1.79%	1.97	3.23%	2.68%	2.71
LSTM	2.57%	2.01%	1.96	3.39%	2.61%	2.69

结合研究问题1和研究问题2所衍生的稳健性检验。



总结做了什么；和未来要做什么。

标准动作。