

量化投资大数据处理 ——基于MATLAB实践

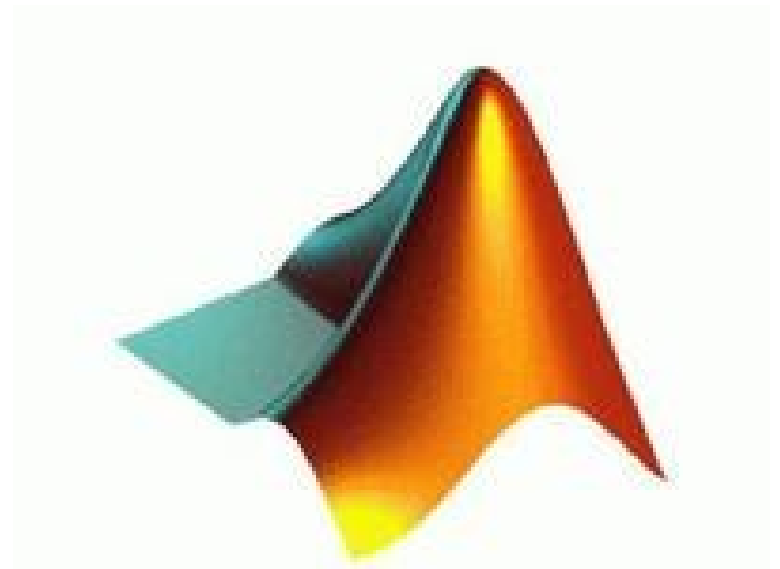
李洋（faruto）

faruto@163.com

<http://www.matlabsky.com>

<http://www.faruto.com>

MATLAB技术论坛联合创始人



内容目录

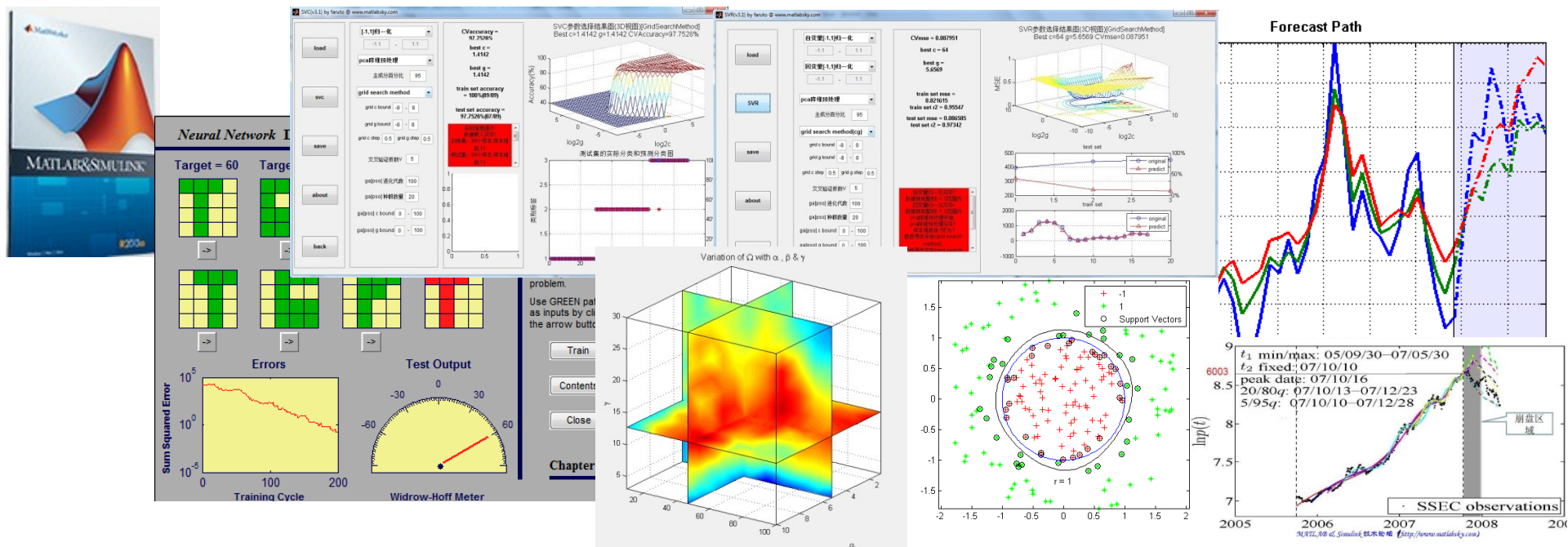
- MATLAB简介
- 量化投资的概念及简介
- 量化投资中的各类数据
- 量化投资大数据处理**MATLAB**实践

内容目录

- **MATLAB**简介
- 量化投资的概念及简介
- 量化投资中的各类数据
- 量化投资大数据处理**MATLAB**实践

MATLAB简介

- MATLAB的全称是矩阵实验室（Matrix Laboratory），是美国MathWorks公司出品的商业数学软件，用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境，主要包括MATLAB和Simulink两大部分。
- 在金融领域，使用MATLAB可以加速产品研究，减少开发时间，提高模型的仿真速度和控制项目成本，利用MATLAB以及相关产品，可以进行分析数据，评估风险、开发并优化策略等一系列金融建模工作。



MATLAB简介

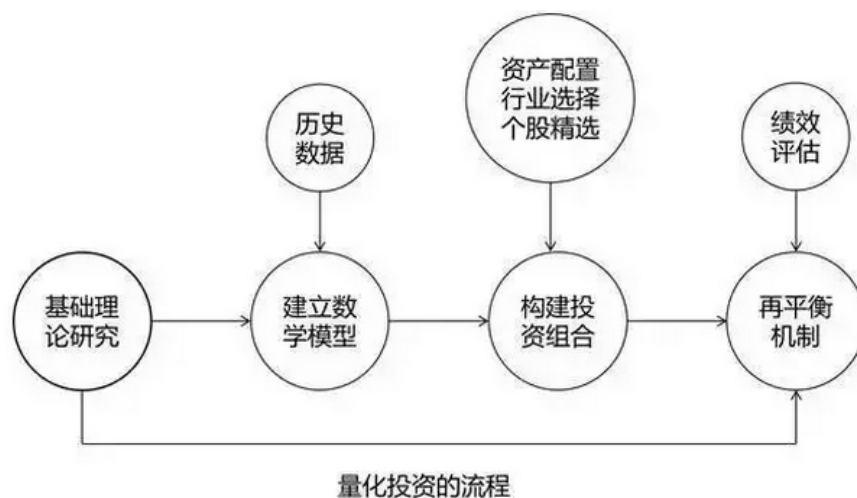
- **MATLAB**包括拥有数百个内部函数的**主工具箱**和**三十几种工具箱**。工具箱又可以分为**功能性工具箱**和**学科性工具箱**。功能性工具箱用来扩充**MATLAB**的符号计算，可视化建模仿真，文字处理及实时控制等功能。学科性工具箱是专业性较强的工具箱，其中金融领域相关的工具箱主要有：
 - **Datafeed Toolbox**: 金融数据工具箱
 - **Econometrics Toolbox**: 计量经济学工具箱
 - **Financial Derivatives Toolbox**: 金融衍生品工具箱
 - **Fixed-Income Toolbox**: 固定收益工具箱
 - **Optimization Toolbox**: 优化工具箱
 - **Statistics Toolbox**: 统计工具箱
- 通过这些工具箱，用户可以利用**MATLAB**进行交易策略实现和回测、投资组合优化和分析、资产分配、金融时序分析、期权价格和敏感度分析、现金流分析、风险管理、预测和模拟、利率曲线拟合和期限结构建模、**Monte Carlo**模拟、基于**GARCH**的波动性分析等。

内容目录

- MATLAB简介
- 量化投资的概念及简介
- 量化投资中的各类数据
- 量化投资大数据处理**MATLAB**实践

量化投资的概念及简介

- 量化投资的定义：量化投资是将人们总结出的投资思想，利用统计学、数学的方法，形成数学模型，借助计算机在海量历史数据中对模型进行验证，寻找能够带来超额收益的多种“大概率”模型，严格按照这些策略所构建的量化模型运算结果来指导投资。
 - 量化投资是一种方法论。
 - 量化投资通常与基本面分析相结合。
 - 量化投资是以定量方法进行投资的各种技术综合。



量化投资的概念及简介

- 量化投资的五大优势
 - 纪律性：严格执行量化投资模型投资建议。
 - 系统性：包括多层次的量化模型、多角度的观察及海量数据的观察等。
 - 及时性：及时跟踪市场变化，不断发现新的提供超额收益的交易机会。
 - 准确性：准确客观评价交易机会，克服主观情绪偏差，妥善运用套利的思想。
 - 分散化：严格控制风险，充当准确地实现分散化投资目标。



量化投资的概念及简介

- 量化投资从业人员到底每天做什么？

某量化交易员(Quant Trader)的一天

8:30-8:55 晨会	13:00-15:15 观察市场， 监控程序和持仓情况	贯穿全天：国内外研报、paper研读、新策略开发、新策略回测； 定期、不定期内部会议； 内部组会、头脑风暴； 参加外部量化、金工会议；
8:55-9:00 交易服务器 检查	15:00-16:00 盘后分析， 计算盈亏，持仓检查， 冲击成本统计	
9:00-11:30 观察市场， 监控程序和持仓情况	20:55-21:00 夜盘开盘 前，交易服务器检查	
	21:00-02:30 夜盘交易 时段	

量化投资的概念及简介

- 传统投资VS量化投资
 - 相同点：本质相同，都是基于市场非有效或是弱有效的理论基础。
 - 不同点：传统投资依赖公司调研和个人经验及主观判断；量化投资依靠数理模型实现投资理念。
- 传统投资名人：巴菲特、索罗斯
 - 沃伦.巴菲特
 - 1989-2008平均年回报约20%；
 - 2008年回报率约-15%。
 - BerkshireHathaway: <http://www.berkshirehathaway.com>
- 量化投资名人：西蒙斯、David Shaw（D.E. Shaw创始人）
 - 詹姆斯.西蒙斯
 - 1989-2008平均年回报约38.5%
 - 2008年回报率约80%
 - Renaissance Technologies: <https://www.rentec.com>

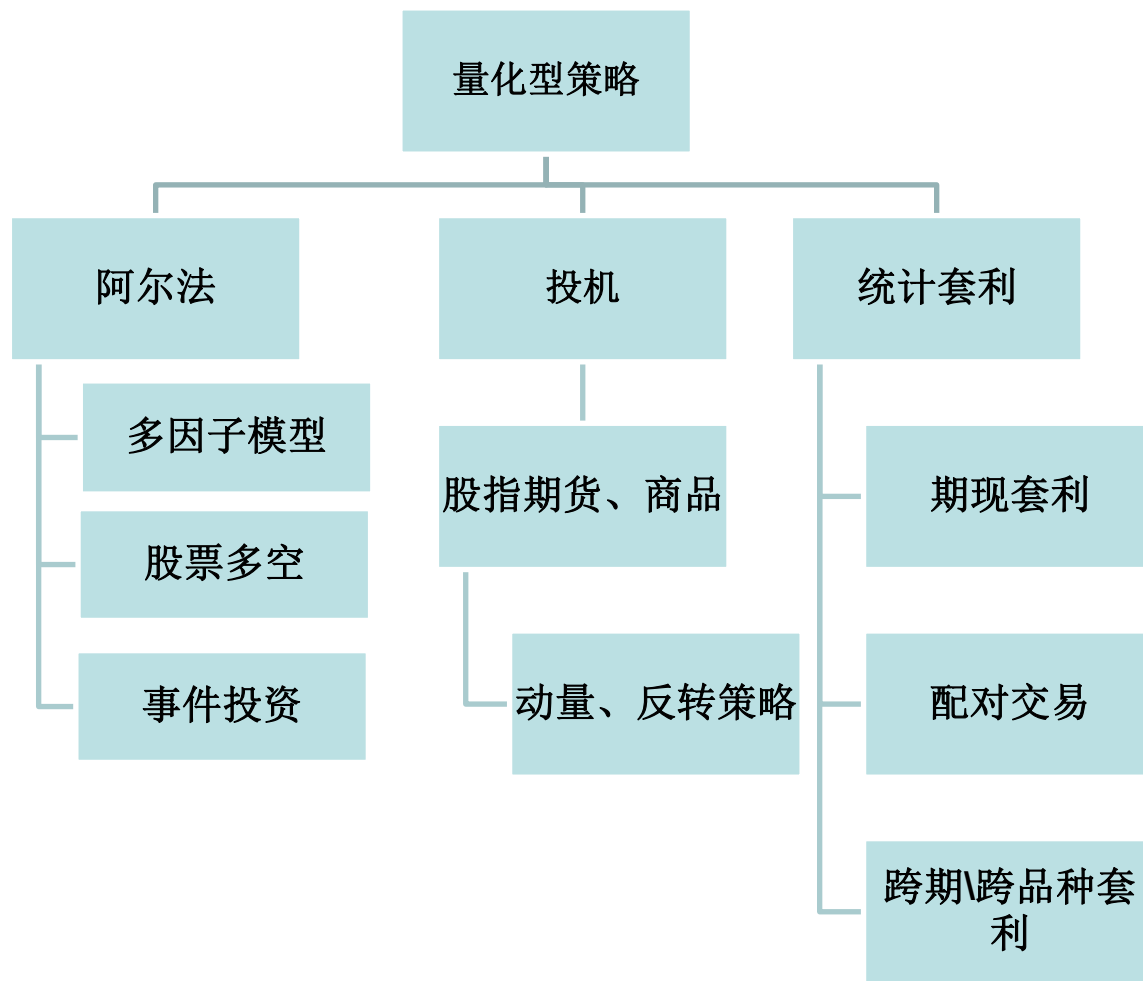


量化投资的概念及简介

	主动投资	量化投资
业绩可重复性	不容易重复	大概率可重复
源于什么样的投资思想	系统的、体系化的、基于基本面和盈利能力变化的、逻辑的、	局部的、针对某类型细节特征的、外科诊断式的、
纪律性	容易受到投资经理的情绪、性格特征、短期交易心态影响	可以克服人性的弱点，如贪婪、恐惧、侥幸心理，也可以克服认知偏差
净值稳定性	过程不一定，结果可预期程度不高	过程有相对清晰的数量特征，未来结果可预期
收益率分布	集中投资、偏离基准、20%的时间和品种贡献80%的收益	收益获取均匀
对全市场的观察力	局部为主	全局观察
执行速度	可快可慢，快不是必要因素	“天下武功，唯快不破”、求快
对系统的要求	知难行易	知难行难

量化投资的概念及简介

- 量化投资的应用



内容目录

- MATLAB简介
- 量化投资的概念及简介
- **量化投资中的各类数据**
- 量化投资大数据处理**MATLAB**实践

量化投资中的各类数据

- 行情类数据（股票、期货、期权）
 - 日线数据大小：MB级别（全市场股票所有历史日线前复权数据.mat文件192MB）
 - 分钟数据大小：MB-GB级别
 - Tick数据大小：GB级别（全市场股票2005至今所有tick数据.mat文件86.8GB）
- 基本面类数据（股票、期货、期权）
 - 财务指标数据大小：GB级别（全市场股票所有财务指标数据.mat文件4.75GB）
 - 三张表数据大小：GB级别（全市场股票所有三张表数据.mat文件8.27GB）
- 舆情类数据（股票、期货、期权）
 - GB-TB级别
 - 百度PC端、移动端搜索数据
 - 新浪微博数据
 - 股吧数据
 - 网易、和讯、东方财富等各大财经网站相关文章数据

内容目录

- MATLAB简介
- 量化投资的概念及简介
- 量化投资中的各类数据
- **量化投资大数据处理MATLAB实践**

量化投资大数据处理MATLAB实践

- 面对量化投资中的大数据，MATLAB中提高效率的方法。
- 并行计算
 - parfor: parallel for
 - SPMD: single program, multiple data
- matfile类
 - Access and change variables directly in MAT-files, without loading into memory
 - 把你的硬盘当作一块大的ROM使用
- GPU加速
 - CUDA: Compute Unified Device Architecture

量化投资大数据处理MATLAB实践

- 量化投资模型的参数优化
- 参数寻优考虑并行计算节省时间。
 - matlabpool, parfor
- 目标函数的选择
 - 常见指标：夏普比率、收益风险比、胜率、最大回撤比例、年化收益率
 - 自定义指标
- 图形展示单独用函数实现，不与原始参数寻优计算耦合。

模型参数优化模块

原始参数寻优数据
计算记录（并行化考虑）



排序处理&统一
图形展示



各种数据图形记录
保存

量化投资大数据处理MATLAB实践

- 量化投资模型的参数优化
- 代码掠影

```
else % 1 == ParallelMode_OnOff
MPS = matlabpool('size');
if 0 == MPS
matlabpool;
end

parfor k = 1:M*N

[i,j] = ind2sub([M,N],k);
Para1Value = ParaMat1(k);
Para2Value = ParaMat2(k);

str = ['===Para Sweep ParallelModeOn==='];
disp(str);
str = ['Total Loop Num:',num2str(M*N),',CurrentLoop:',num2str(k)];
disp(str);
str = ['CurrentIndex:',num2str(i),',',...
num2str(j),',Para1:',num2str(Para1Value),',Para2:',num2str(Para2Value),'];
disp(str);

SReport{k} = fun(F,Para,Para1Value,Para2Value);
end
%matlabpool close
end
```

```
%% Plot Block

% matrixplot
% matrixplot(x,'FigShap','s','FigSize','Full','ColorBar','on','FigStyle','Auto',...
% 'XVarNames',XVarNames,'YVarNames',YVarNames);

x = PSReport.RetdRisk;
XVarNames = num2cell(ParaBound{2,1});
YVarNames = num2cell(ParaBound{1,1});
matrixplot(x,'FigShap','s','FigSize','Full','ColorBar','on','FigStyle','Auto',...
'XVarNames',XVarNames,'YVarNames',YVarNames);
title({'目标函数：收益风险比;'},'FontWeight','Bold');

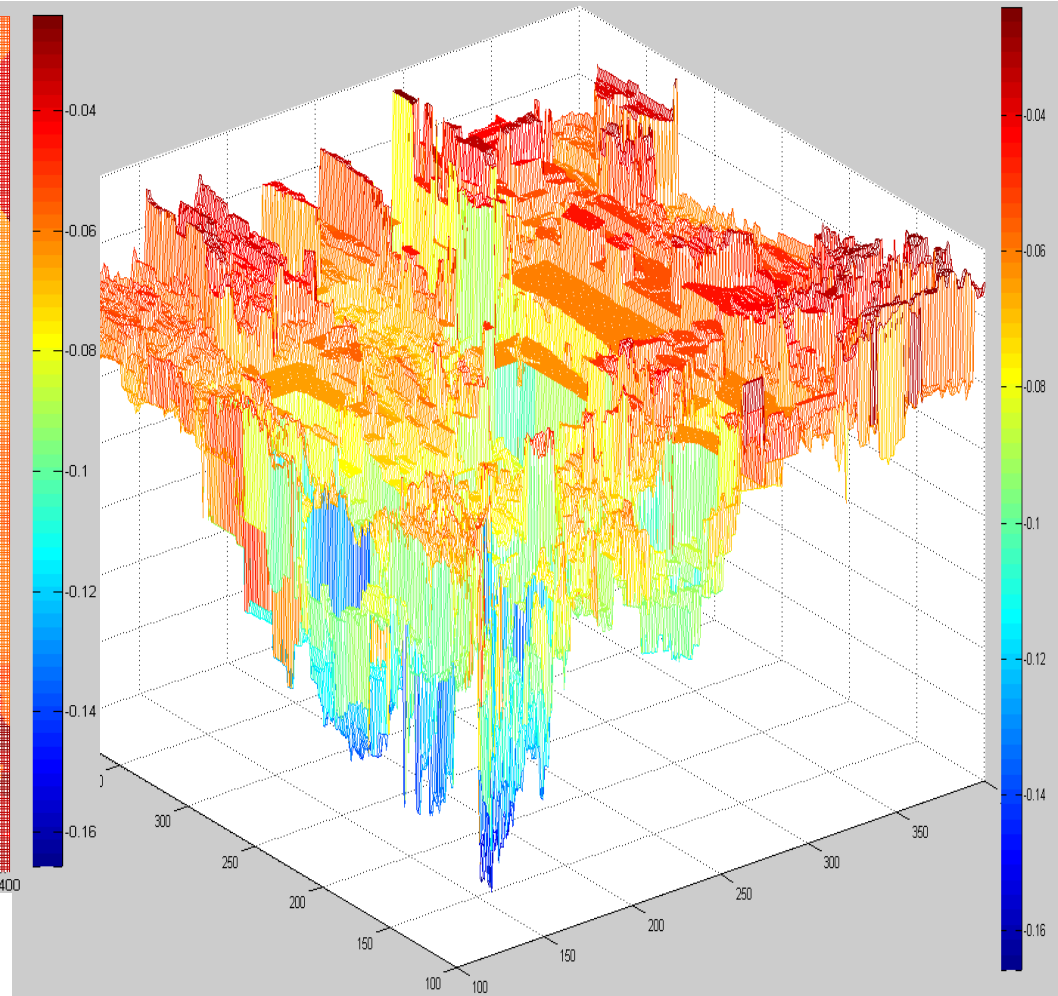
x = PSReport.WinRatio*100;
XVarNames = num2cell(ParaBound{2,1});
YVarNames = num2cell(ParaBound{1,1});
matrixplot(x,'FigShap','s','FigSize','Full','ColorBar','on','FigStyle','Auto',...
'XVarNames',XVarNames,'YVarNames',YVarNames);
title({'目标函数：胜率(%)';'},'FontWeight','Bold');

x = -PSReport.MDD*100;
XVarNames = num2cell(ParaBound{2,1});
YVarNames = num2cell(ParaBound{1,1});
matrixplot(x,'FigShap','s','FigSize','Full','ColorBar','on','FigStyle','Auto',...
'XVarNames',XVarNames,'YVarNames',YVarNames);
title({'目标函数：最大回撤比例(%)';'},'FontWeight','Bold');

x = PSReport.AnnualRet*100;
XVarNames = num2cell(ParaBound{2,1});
YVarNames = num2cell(ParaBound{1,1});
matrixplot(x,'FigShap','s','FigSize','Full','ColorBar','on','FigStyle','Auto',...
'XVarNames',XVarNames,'YVarNames',YVarNames);
title({'目标函数：年化收益率(%)';'},'FontWeight','Bold');
```

量化投资大数据处理MATLAB实践

- 量化投资模型的参数优化



Thank you

