

# 江南智投

无论美国还是中国，在金融市场取得成功的基础是相同的，即对未来进行准确的估计。有人利用对宏观经济形势的估计来投资和交易；有人利用对行业发展预期的估计来投资和交易；有人利用对一些技术指标的估计来投资和交易；KDS 利用 UBX™ 人工智能系统来投资和交易。

在过去 30 年中,在统一量子场论 (QED) 的指导下, 通过在金融市场不断积累与实践, KDS 在 UBX™ 人工智能系统中建立并发展了一系列强有力的金融建模方法。从功能角度看, 可以划分为收益模型、风险模型和优化模型；从交易频率看, 可以划分为高频模型和低频模型。UBX™ 高频收益模型可以对债券或标的物在下一时刻的价格 (间隔不大于 1 分钟) 进行预测；UBX™ 低频收益模型可以对债券或标的物未来一个月的收益进行预测；UBX™ 风险模型可以对债券或者标的物在未来一段时间的波动进行预测。UBX™ 优化模型综合考虑收益和风险, 确定最优的投资组合, 得益于 KDS 独有的反褶积算法, UBX™ 优化模型可以处理百万级别的数据。在 UBX™ 人工智能系统的帮助下, KDS 在美国金融市场取得了巨大的成功。

江南智投是 KDS 根据江南农村商业银行的收益预期和风险偏好, 结合中国 A 股市场的实际情况, 利用 UBX™ 人工智能系统为江南农村商业银行量身定做的一款 A 股基金产品

## 1. 模型概述

由于中国股票市场是 T+1 交易且不可以买空卖空, 所以我们用 UBX™ 低频收益模型预测每支股票未来一个月的收益, 用 UBX™ 风险模型预测每支股票未来一个月收益率序列的方差和任意两支股票收益率序列的相关系数 (方差和相关系数代表了风险), 结合对收益和风险的预测, 用 UBX™ 优化模型确定风险可控下的最优投资组合。模型建立完成后, 在每个交易日结束, 会根据最新的截面数据, 计算下下一个交易日的最优投资组合。市场开盘后, 如果某一时刻满足这两个条件就换仓: a) 当前交易日投资组合的预期收益大于前一个交易日投资组合的预期收益; b) 由前一个交易日的投资组合换仓到当前交易日投资组合的收益大于 0; 否则保持上前一个交易日投资组合不变。

UBX™ 人工智能系统对中国 A 股市场建立的投资组合模型有以下优点:

- 收益率高

随着交易数据积累和交易结果的滚动反馈, UBX™ 人工智能系统对 A 股的分析 and 预测会越来越好, 从而在保证预期收益率的前提下, 收益率会逐步提高, 达到甚至超过 36%。

- 风险可控

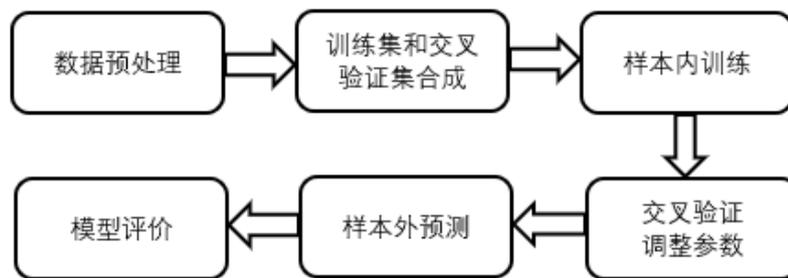
模型的最终输出结果是一个最优投资组合集合，集合中的每个元素是一个投资组合。每个组合都对应着相应的风险，且是在这个风险下收益最高的组合。江南农村商业银行可以根据自己的预期收益和风险偏好，选择适合自己的投资组合。

- 交易自动化

得益于 UBX™ 人工智能系统强大的计算能力，每个交易日都可以为下一个交易日计算出最优投资组合。根据上面提到换仓条件 a) 和 b)，一个简单的交易程序就可以实现自动化交易。自动化交易可以降低基金的运营成本还可以避免人为操作的失误。

## 2. 建模过程

这个部分我们会简要介绍下整个建模过程。下面是建模的流程图。



建模流程

### 2.1. 数据准备

我们从对外部变化的响应、行业因素、季节因素和风险指数这四个作为模型输入。外部变化的响应包括：10 年期国债收益率、存贷款基准利率、中国工业生产指数、国际石油价格、人民币对美元汇率、S&P500 指数。按照中国证券监督管理委员会公布的《上市公司行业分类指引》（2012 年修订），将整个 A 股上市公司分成农林牧渔业、采矿业、制造业等 20 个大类。我们在建模时会考虑建模数据截面所处的月份。风险因素包括价值、成长、财务质量、财务杠杆、市值、动能、股东、交易数据等方面。

我们只关注基本数据，对输入模型的数据不做过多的人为处理，如计算波动率，通过对市场指数收益的拟合计算贝塔以及其它常用的技术分析指标。模型本身会对这些基本数据做分析，提取归纳出不同因素对收益和风险的影响，这是一个非常复杂的非线性过程。

### 2.2. 训练过程

整个训练过程包括两个阶段，第一阶段是对收益和风险进行预测，第二阶段是在第一阶段工作的基础上建立优化模型，确定最优的投资组合。我们的建模从数据预处理部分开始，数据预处理包括：缺失值填充，异常值分析以及离散变量向量化。接下来通过交叉验证确定

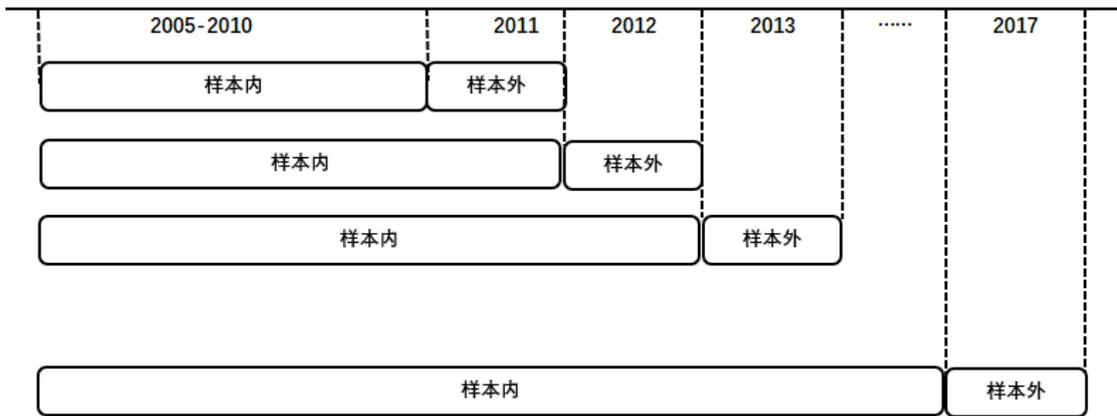
模型的超参数，最后用所有的数据进行训练，建立预测模型。

假设 A 股有 3000 支股票（实际大于 3000），要对这些股票之间的相关系数进行预测，那么输入模型中的记录是  $3000 \times 2999 / 2$  条，大约是 4500000 条。UBX™ 人工智能发动机会把这些记录分布到 36 子系统上同时运算，从而大大加快了建模速度。

建立优化模型时面临着两个问题：1) 所有股票协方差矩阵进行准确的估计；2) 求解最优投资组合集合的过程复杂，且通常要求协方差矩阵非奇异。得益于 UBX™ 人工智能发动机的帮助，我们可以在 QED 框架下对协方差矩阵进行非线性估计，相比于常用的线性模型有更好的准确度，因为非线性模型更符合实际情况。利用 UBX™ 人工智能发动机的反褶积算法，可以快速准确求解最优投资组合集合。

### 2.3. 回溯测试

我们用“滚动测试法”进行回溯测试，如图所示：



回溯测试

样本内的数据用来建模，样本外的数据用来测试。回测区间为 2010-01-31 到 2017-12-31，每个月进行一次换仓，即对给定的风险，在每个月最后一个交易日计算下个月的最优投资组合，按照下个月第一个交易日的收盘价换仓。我们会从收益率，夏普比率，信息率，最大回撤，相对沪深 300 的超额收益，相对于中证 500 的超额收益，相对于沪深 300 的超额收益最大回撤，相对于中证 500 的超额收益最大回撤等方面考察和评价模型的效果。

### 2.4. 模拟交易

模拟交易是在真实交易平台的模拟交易系统中进行的。交易策略就是在模型概述中提到两个交易触发条件 a) 和 b)。

## 3. 合作方式

KDS 会在整个建模过程中对江南农村商业银行完全开放，防止“黑箱操作”的发生，消除

不必要的误会。在建模过程的每个阶段，都会给出详细建模报告，包括数据处理与分析，数学原理，模型输出及评价以及下阶段工作计划。为了完成整个建模工作，需要江南农村商业银行提供必要的硬件和数据方面的支持。数据、模型以及应用都是在江南农村商业银行的“江南一号”（具体配置参见第三部分-必要条件）上进行的。KDS 的技术总监会带领相关技术人员在江南农村商业银行搭建起建模所需的软硬件环境，这个过程大约需要 3 到 4 周，KDS 的建模人员通过远程登陆江南一号的方式完成后续的模式开发工作。

## 4. 必要条件

### 4.1. 数据需求

- i. 所有 A 股上市公司历年财务报表。
- ii. 所有 A 股上市公司的历史交易数据，包括起始时间节点价格，持续期内的最高价，持续期内的最低价，持续期内的换手率，持续期内的交易量。所谓的持续期，也就是更新频率，为 1 分钟。

### 4.2. 硬件需求

所有的数据，建模以及应用都是在江南一号上进行的，具体配置要求如下。

2 个主控系统和 36 个分散式运算子系统

UBX 主控系统

CPU: 2\*12 核 Intel(R) Xeon(R) E5 26 系

RAM: 256GB DDR4 ECC

HD: 8\*2TB SATA 6GB/s

4 个服务器, 1000w 冗余电源供应

UBX 分散式运算子系统

CPU: 2\*4 核 Intel(R) Xeon(R) E3 12 系

RAM: 4\*4GB DDR4 ECC, 可增加至 32GB

HD: 4\*2TB SATA 6GB/s, Raid 1

2 个服务器,冗余电源供应

UBX 网站服务器

2 个网站服务器, 其中 1 个用于内部开发, 另一个用于产品线

CPU: 2\*4 核 Intel(R) Xeon(R) E3 12 系

RAM: 4\*4GB DDR4 ECC,可增加至 32GB

HD: 2\*500GB SSD, Raid 1

1 个服务器,冗余电源供应

机架

8\* 42 个服务器机架

外围宽度: 600mm

高度: 2,000mm – 机架数: 42 个

深度: 1,000mm & 1,050mm

所有服务器必须是机架式的

保留最底层机架空间给 6 个不断电电源系统

网络

每个系统都配有 2 个千兆网卡

8 \* 48 个用于局域网的端口千兆交换机

可配置式防火墙, 路由器和域名系统

操作系统

Centos 7 Linux 发行版最新稳定版

软件

Github

Python

Java C++

TomCat

Apache,etc

#### 4.3. 组织需求

为了建模的公开, 透明以及相互沟通, 需要有江南农村商业银行的同事加入到团队中,

包括 IT 部门的同事，风险控制部门的同事和交易部门的同事。

#### 4.4. 所需时间

整个过程需要 6 个月时间，其中建模需要 3 个月，回溯测试需要 1 个月，模拟交易 2 个月。