

2023 年第四届“大湾区杯”粤港澳 金融数学建模竞赛

题目

跨境 ETF 套利策略设计

摘 要：

本题旨在研究 ETF 的交易规则，制定不同市场下 ETF 的套利策略。依据套利策略的套利条件我们制定了选取 ETF 模型。

针对问题一，我们分析了 ETF 的市场组成成分、交易规则，并依据 ETF 对应成分股上市的公司划分为三个类别：单市场 ETF、跨市场 ETF、跨境 ETF。针对单市场，依据其交易规则我们设计了瞬时折溢价套利策略和基于 ARIMA 时间序列的延时套利策略。**溢价套利策略为：**低价买入成份股、申购 ETF、高价卖出 ETF。**折价套利策略为：**低价买入 ETF、赎回成份股、高价卖出成份股。**延时套利策略**依据 ARIMA 预测是否有套利机会提前完成折溢价套利。我们考虑三要素（盈利率、市场流动性、最大回撤率）建立单市场 ETF 选择模型，通过历史数据计算单市场 ETF 的三要素指标值并进行均一化处理，最后根据三个数据的加权求和得到的套利综合指标选取最适合套利的前 10 只股票见表 1。

针对问题二，我们套利市场属于跨境 ETF。在该市场下对比问题一市场，市场交易的规则有所变动。在申购时，T+0 使用现金提交申购申请，T+1 以 IOPV 收盘价确认 ETF 份额，T+2 进行 ETF 卖出或赎回。在赎回时，T+0 使用 ETF 赎回现金，T+10 内取得现金。在该市场规则下更我们修改溢价套利策略，我们为更准确的预测非线性市场，依据历史汇率、ETF 价格分布、一级市场价格构建神经网络模型预测 T+1 一级市场的价格与 T+2ETF 价格分布。**基于神经网络跨境 ETF 溢价套利策略为：**基于神经网络输出结果进行选择申购时机，以期望在 T+1 天低价申购 ETF、T+2 天上高价卖出 ETF。在跨境 ETF 的选取中应考虑溢价套利的计算的是 T+1 天的 IOPV 和 T+1 天的 ETF 价格，再结合问题一的单市场 ETF 选取模型可选取出最适合套利的前十只股票见表 3。

针对第三问，我们要利用跨境 ETF 对上证 50、沪深 300、中证 500、中证 1000 进行跨市场套利模型。基于期货交易规则我们制定了两种期现套利策略。**期现套利策略一：**低价买入跨境 ETF、高价卖出期货合约。**期现套利策略二：**向基金公司借入跨境 ETF、高价卖出跨境 ETF、买入低价期货、买入低价跨境 ETF 归还基金公司。由于股指期货与股指间具有一致性与收敛性，所以我们选择跨境 ETF 充当现货时要经过跟踪偏离度检验。依据标准差公式对 99 支跨境 ETF 对 4 支股指进行检验。得出最适合跨市场套利的跨境 ETF 和股指期货的组合见表 5。

关键词：ARIMA 模型 神经网络 溢折价套利 ETF 选择策略 跨境 ETF 套利策略

一、问题重述

1.1 引言

交易型开放式指数基金(Exchange Traded Fund, 简称 ETF), 又称交易所交易基金, 是一种跟踪“标的指数”变化的基金。投资人可以如买卖股票那么简单地去买卖“标的指数”的 ETF, 获得与该指数基本相同的报酬率。根据投资标的类型或运作模式不同, 分为股票 ETF、债券 ETF、跨境 ETF、商品 ETF、交易型货币基金几类。

ETF 属于开放式基金的一种特殊类型, 结合了封闭式基金和开放式基金的运作特点, 投资者可以向基金管理公司申购或赎回基金份额, 又能像封闭式基金一样在二级市场上按市场价格买卖 ETF 份额。不过, 申购赎回必须以一篮子股票换取基金份额或者以基金份额换回一篮子股票。由于同时存在二级市场交易和申购赎回机制, 投资者可以在 ETF 基金一级市场基金份额净值和二级市场买卖价格之间存在差价时进行套利交易。ETF 套利的基本形式有两种, 分别是折价套利和溢价套利。作为基本形式的延伸, ETF 的套利还有事件套利、期现套利。而跨境 ETF 的投资者套利模式可以借助 T+0 获取盘中价的套利方法、实时折溢价套利方法、ETF 事件套利方法。其他, 投资者还可以根据实际情况进行“与境内 ETF 进行配对交易”、“期现套利”等套利操作。

1.2 问题提出

本题目主要是讨论 ETF 基金的套利策略设计, 请你运用数学建模方法, 完成如下任务(数据时间为 2023 年 8 月 1 日-10 月 31 日)。

任务一: 针对在上海交易所和深圳交易所上市的 ETF 基金(参赛者可使用但不限于附件给出的 ETF 基金交易数据), 请建立一二级市场间套利模型, 选出最适合进行套利的前 10 只。

任务二: 目前与 A 股股票相关的 ETF 基金只能 T+1, 而跨境 ETF(软件显示有 100 多只)可以进行 T+0 交易, 针对交易制度的不同, 修改任务一的策略, 建立跨境 ETF 套利模型, 选出最适合进行套利的前 10 只。

任务三: 目前中金所有上证 50、沪深 300、中证 500、中证 1000, 四个指数的股指期货。利用跨境 ETF 与国际市场的联动性, 以及国际市场对 A 股的影响, 建立跨境 ETF 与股指期货的跨市场套利模型。并选出最适合跨市场套利的跨境 ETF 和股指期货组合。

任务四: 请以任务二建立的模型对恒生科技 ETF(513130), 以及任务三的跨市场套利模型对跨境 ETF 和股指期货组合进行实测, 并提交实测报告。

参数设置: 初始投资为 150 万元, 期货为 100 万元, 现货为 50 万元, 期货保证金比率为 10%, 交易费率期货 0.005%, 股票为 0.01%, 价格滑点为 0.01。

二、问题分析

本题主要对 ETF 在不同交易市场不同的交易规则建立套利模型,对于本题主要的交易市场有三个,分别是单市场 ETF 交易、跨市场 ETF 交易、跨境 ETF 交易,基于这三个交易市场,我们需要对附件 A862 只 ETF 进行对应交易市场的分类,从而建立不同市场的套利模型。

2.1 问题1分析

针对问题一,通过上海证券交易所和深圳证券交易所上市的 ETF 要给出建立一二级市场间 ETF 的套利模型,并且选择出最合适套利的前10只。首先对于本问,我们建立的是单市场交易套利模型,再根据 ETF 两种基本套利模型:折价、溢价套利模型入手,建立折溢价瞬时套利和延时套利,但两者的最终目的都是使得套利获得的差价最大。对于瞬时套利模型,买和卖几乎是在同一时刻进行的,因此我们不需要预判,而对于延时套利模型,我们需要对未来 ETF 的走势进行预判,为此,我们需要引入时间序列算法,对其未来几分钟 ETF 走势进行预判,如下图:

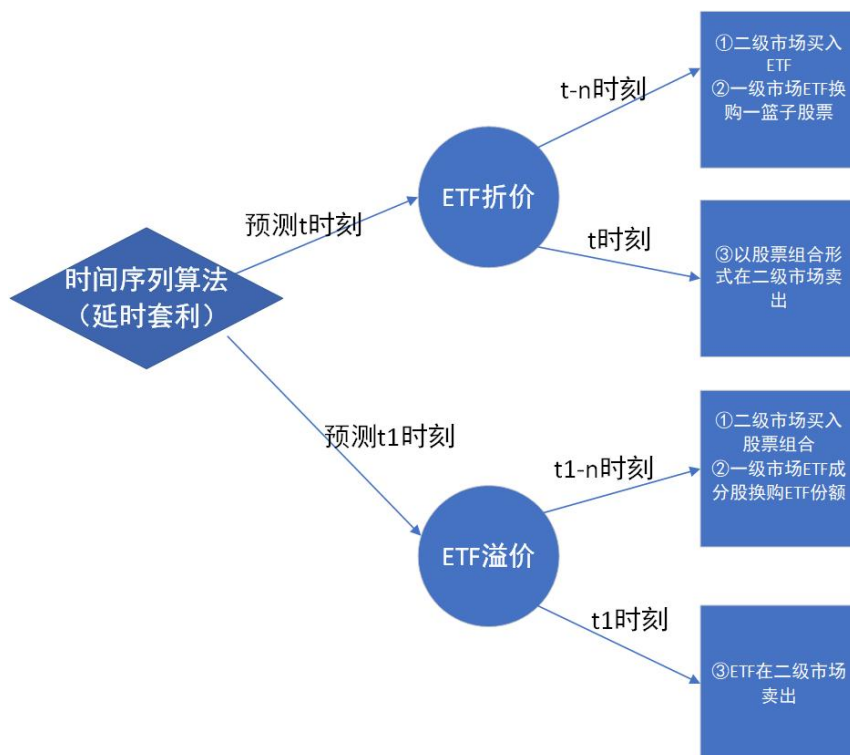


图 1：延时套利步骤流程

我们可通过设置三个指标来选取最适合套利的前10只 ETF；第一个指标是 ETF 的平均收益率是否最大,第二个指标是 ETF 的回撤率大小,第三个指标是 ETF 流动性测度大小。对于选取最适合单市场套利的前10只 ETF: 我们主要通过对三个指标赋予权重,根据选股模型(具体见模型建立5.1.2)结果选取排名前10只 ETF 进行套利。利用同花顺软件平台,我们可以抓取得到98只单市场交易的 ETF 近三个月的单位净价,利用题目给出862只 ETF 在二级单市场的买卖交易数据信

息，通过 *matlab* 编程可以计算得到每只单市场 ETF 权重综合值。

另外，由于不同交易市场的交易制度不同，我们需要对于单市场的交易制度进行分析，如下图：

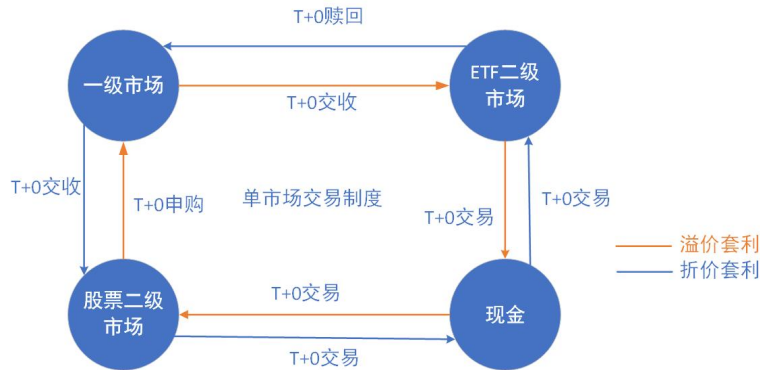


图 2：单市场交易制度

单市场一二级市场之间的交易遵守 T+0 交易制度、单市场二级市场之间的交易遵守 T+1 交易制度。

2.2 问题 2 分析

针对问题二，根据交易制度的不同要对问题一的策略进行修改并建立跨境 ETF 套利模型。首先，由于跨境市场 ETF 的交易制度与国内的交易制度不同，我们需要先分析出跨境 ETF 的交易制度，根据上交所跨境 ETF 的交易制度，我们得到如下图的跨境 ETF 交易：

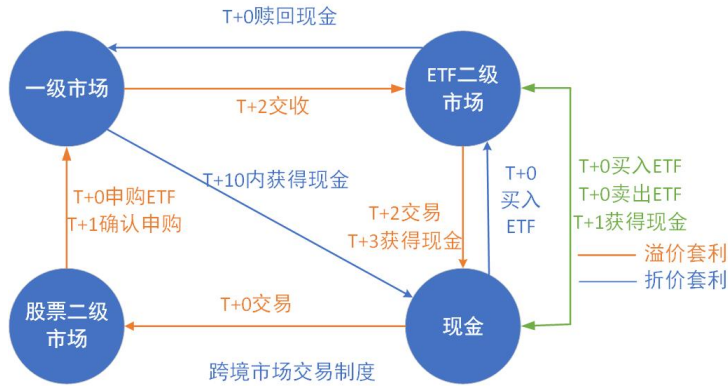


图 3：跨境市场交易制度

不同于国内单市场 ETF 的折溢价套利在一二级市场的交易是 T+0，跨境折价套利交易制度：可以 T+0 买入 ETF，T+0 赎回现金、T+10 内现金到账，不同于国内可以赎回股票和现金；跨境溢价套利交易制度可以 T+0 申购 ETF、T+1 确认申购 ETF，等到 T+2 才能在 ETF 二级市场卖出 ETF、T+3 现金才能到账；跨境 ETF 套利盘中差价遵守 T+0 交易、T+1 获得现金。

接着我们利用同花顺平台从附件 A 中筛选出 99 只跨境 ETF，对于要选取适合跨境套利的前十只 ETF，我们仍然使用 ETF 的平均收益、ETF 的流动性测度、ETF

的回撤率作为选取参考指标，由于跨境 ETF 折溢价套利的交易制度不同，我们需要对其进行分类讨论，对于跨境 ETF 的折价套利，由于它实行 T+0 交易，国内单市场折价套利对其仍然适用；对于跨境 ETF 的溢价套利，由于它实行 T+1 申购、T+2 交收，由于投资者是在 T+0 时看到跨境 ETF 溢价才进行请求申购的行为，但是实际上投资者真正将跨境 ETF 拿在自己手里进行买卖交易的时候，此时已经是 T+2 天了，那么 T+2 天跨境 ETF 是折价还是溢价是在 T+0 天的时候还不能知道的，因此我们通过神经网络算法来预测 T+2 天的均价，从而来判断 T+2 天后的 ETF 是否适合溢价套利。

2.3 问题3分析

针对问题三，我们需要建立由于需要跨境 ETF 与股指期货的跨市场套利模型。考虑跨境 ETF 与国际市场的联动性，以及国际市场对 A 股的影响，所以可以利用跨境 ETF 的走势跟踪上证 50、沪深 300、中证 500、中证 1000 这四个指数，选择跨境 ETF 充当现货进行进行套利。由于股指期货与股指间具有一致性与收敛性，所以我们选择跨境 ETF 充当现货时要经过跟踪偏离度检验，再选取最适合跨市场套利的跨境 ETF 和股指期货的组合进行套利。

三、符号说明

符号	说明
$A_1(i, j)$	第 j 个 ETF 在第 i 天是否进行折价套利
$A_2(i, j)$	第 j 个 ETF 在第 i 天是否进行溢价套利
a_j	第 j 个单笔 ETF 在总交易天数内平均单位套利收益
$iopv(i, j)$	第 j 个单笔 ETF 在第 i 天的参考单位基金净值 (IOPV)
$q(i, j)$	第 j 个单笔 ETF 在第 i 天的实时净值均值
b_j	第 j 个 ETF 的流动性测度
$amount(i, j)$	第 i 天第 j 只 ETF 的交易金额
c_j	第 j 只 ETF 在 n 日交易天数的最大回撤率
W	折溢价套利策略盈利金额
φ_i	AR 自回归模型参数
μ_i	MA 移动平均模型参数

四、模型假设

- 1、假设 ETF 基金成分股对应公司无重大事件，如公司破产、公司被收购。
- 2、假设 ETF 再外汇的管理制度、交易规则没有发生变动。
- 3、假设基金公司能在规定时间内及时补券。
- 4、假设股票市场流动性不会枯竭。
- 5、假设溢价套利、折价套利需要的手续费成本分别为 0.1%、0.2%。

五、模型建立与求解

5.1 ETF 溢折套利分析

依我们认识，在市场交易中通过差价买卖商品赚取差价称为套利。一个简单的套利模型为批发商、零售商、消费者。零售商以批发价（低价）向批发商购买商品，然后以零售价（高价）把商品卖给消费者。

同理，在股市交易中我们套利的方法不外乎低价购买基金高价出售基金。本题我们要研究 ETF 的套利空间。ETF 基本的市场第一是面向基金管理公司（一级市场），第二是面向 ETF 交易市场（二级市场）。

我们注意到当 ETF 二级市场交易价格高于一级市场 IOPV 价格时，可以进行溢价套利。具体操作如下：第一步从股票市场买入一篮子股票，第二步用一篮子股票和少量现金向基金管理公司申购 ETF，第三步在 ETF 交易市场出售 ETF 换取差价。具体如下图：

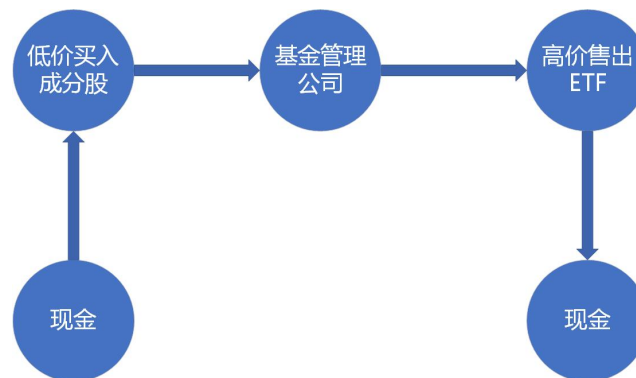


图 4：溢价套利

而当 ETF 二级市场交易价格低于一级市场 IOPV 价格时，可以进行折价套利。具体操作如下：第一步从 ETF 交易市场买入 ETF，第二步用 ETF 向基金管理公司赎回一篮子股票和少量现金，第三步在股票市场出售股票换取差价。具体如下图：

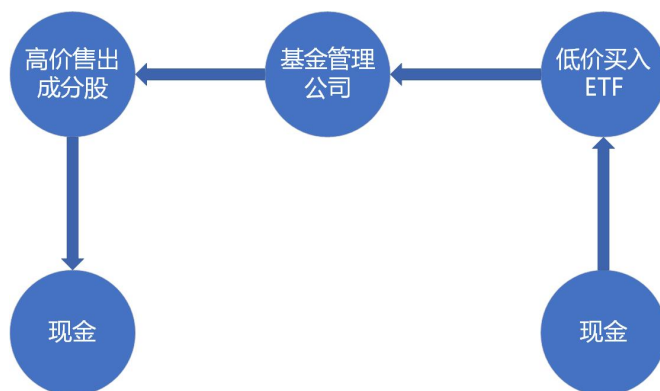


图 5：折价套利

在上述两种套利中，我们注意到套利成功的关键要素。第一为市场存在差价，第二为操作成本控制，第三为市场上的流动性充足。市场存在差价是我们套利的出发点，而我们将 ETF 转换需要额外费用所以我们要控制成本，最后我们要保证市场上有充足的买卖需求。具体如下所示：

$$W = p_0 x - \sum_{i=1}^n p_i y_i - c$$

上述为溢价套利模型获利的计算公式。其中， W 获利部分； p_0 为单位 ETF 价格； x 为 ETF 售出数量； p_i 为购入 ETF 清单第 i 支股票的单位价格； y_i 为购入 ETF 清单第 i 支股票数量； c 为操作成本。同理我们依据折价套利模型也可得出获利的计算公式，具体如下所示：

$$W = \sum_{i=1}^n p_i y_i - p_0 x - c$$

其中， W 获利部分； p_0 为单位 ETF 价格； x 为 ETF 购入数量； p_i 为售出 ETF 清单第 i 支股票的单位价格； y_i 为售出 ETF 清单第 i 支股票数量； c 为操作成本。

5.1.1 ETF 延时套利分析

在上述中我们依赖于一二级市场交易规则完成瞬时套利，依据文献^[1]我们还可以对一二级市场进行延时套利。延时套利与瞬时套利的区别是整体操作的时间区间提前，从而比瞬时套利更快的完成在二级市场卖出 ETF 或股票。

由于我们要提前操作，故我们要有一个预判 ETF 溢折价。我们考虑使用时间序列 ARIMA 模型来预测 ETF 的走势。依据文献^[2]ARIMA 模型由两个基本模型组成具体如下所示：

(1) AR 自回归模型

AR 模型主要思想为当前的价格在一定程度上由历史价格所决定的。在数学上, 我们通过对 ETF 价格时间序列 $\{X_1, X_2, \dots, X_p\}$ 的线性组合可计算出预测序列 $\{X_{p+1}, X_{p+2}, \dots, X_{p+n}\}$, 具体如下所示:

$$X_{p+j} = \varphi_0 + \sum_{i=1}^p \varphi_i X_{p+j-i} \quad j=1 \dots n$$

其中, φ_i 表示自回归系数; $\{X_1, X_2, \dots, X_p, \dots, X_{p+n}\}$ 为时间序列; p 为选取历史序列的数量, 称为 p 阶数; n 为模型预测的步数。

(2) MA 移动平均模型

MA 模型主要思想为所有价格应是平稳的, 价格在时间上产生波动的原因是因为与过去预测误差有关。即对 ETF 预测误差的时间序列 $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q\}$ 的线性组合可计算出预测序列 $\{X_{q+1}, X_{q+2}, \dots, X_{q+n}\}$, 具体如下所示:

$$X_{q+j} = \mu + \varepsilon_{q+j} + \sum_{i=1}^q \theta_i \varepsilon_{q+j-i} \quad j=1 \dots n$$

其中, θ_i 表示移动平均系数; $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q, \dots, \varepsilon_{q+n}\}$ 为预测误差序列; $\{X_{q+1}, X_{q+2}, \dots, X_{q+n}\}$ 为 ETF 预测值序列; μ 为 ETF 价格时间序列的期望值; n 为模型预测的步数。

对此我们 ARIMA 结合了上述两种模型, 以 AR 模型计算得出的预测值充当 MA 模型的均值 μ , 具体如下所示:

$$X_{t+j} = \left(\varphi_0 + \sum_{i=1}^p \varphi_i X_{t+j-i} \right) + \varepsilon_{t+j} + \sum_{i=1}^q \theta_i \varepsilon_{t+j-i} \quad j=1 \dots n$$

其中, φ_i 表示自回归系数; θ_i 表示移动平均系数; X_{t+j} 表示未来第 j 个时间点的预测值; X_{t+j-i} 表示历史价格观察值; ε_{t+j-i} 表示历史价格误差值; n 为模型预测的步数。

我们对 IOPV 和 ETF 二级市场依据 ARIMA 模型建立两个预测模型, 当我们捕获到时刻 $t+j$ 会产生溢折价时, 我们可以在时刻 $t+j-1$ 提前做出反应 (买一篮子股票或 ETF), 当到达 $t+j$ 时刻完成溢折操作 (卖出 ETF 或一篮子股票)。

我们使用 ARIMA 模型进行预测时我们可以得出 IOPV 预测值的 95% 区间 $[X_{\min}, X_{\max}]$ 和 ETF 二级市场价格预测值的 95% 区间 $[Y_{\min}, Y_{\max}]$ 。当区间 $[X_{\min}, X_{\max}]$ 和区间 $[Y_{\min}, Y_{\max}]$ 交集区域越小说明一二级市场产生差价的准确性越大, 具体以某数字去界定进行延时套利则需要实测获取。具体如下:

$$\begin{aligned} D &= [X_{\min}, X_{\max}] \cap [Y_{\min}, Y_{\max}] \\ &= [D_{\min}, D_{\max}] \\ f(D) &= \begin{cases} 1 & D_{\max} - D_{\min} \leq d \\ 0 & D_{\max} - D_{\min} > d \end{cases} \end{aligned}$$

其中，运算 $X \cap Y$ 表示取区间 X 和区间 Y 的交集； $f(D)=1$ 表示进行延时套利操作， $f(D)=0$ 表示不进行延时套利操作； d 为判断边界值。

5.1.2 基于指标与套利空间的 ETF 的选取模型

在上述中我们分析了 ETF 的一二级市场间的瞬时与延时溢折套利模型。一次成功套利涉及到一二级市场的差价和套利成本控制、市场流动性、每只 ETF 在一个周期内的最大回撤率是多少，它的单位平均收益是多少，是选取 ETF 进行套利重要的参考指标。接下来我们依据上述条件建立套利最优化的选取模型，具体如下所示：

(1) ETF 的单市场套利价值收益

由于单市场一二级市场之间实行 T+0 交易，可通过需比较的 ETF 相同时刻在一级市场的参考单位基金净值 (IOPV) 和二级市场的实时净值，计算扣除套利手续费后的单位套利收益。具体方式如下：

① “0-1” 变量约束

同时我们需要统计在交易总天数里面每只 ETF 折价套利的天数和溢价套利的天数，因此需要对第 j 个 ETF 在第 i 天是否进行折价或溢价套利进行 “0-1” 变量的约束，具体约束如下：

$$\begin{cases} A_1(i, j) \in \{0,1\} & i=1 \dots n; j=1 \dots 98 \\ A_2(i, j) \in \{0,1\} & i=1 \dots n; j=1 \dots 98 \end{cases}$$

由于 ETF 在一级和二级市场进行套利时需要一定的套利手续费，所以需要 ETF 的价格与其 IOPV 的差价足以覆盖交易费用时才能进行套利。可分别得到折价套利和溢价套利的条件如下：

$$\begin{cases} A_1(i, j)(iopv(i, j) - q(i, j) - iopv(i, j)c_1) > 0 \\ A_2(i, j)(q(i, j) - iopv(i, j) - iopv(i, j)c_2) > 0 \\ c_1 = 0.002 \\ c_2 = 0.001 \end{cases}$$

其中 c_1 和 c_2 分别为赎回的手续费和申购的手续费， $iopv(i, j)$ 为第 j 个单笔 ETF 在第 i 天的参考单位基金净值 (IOPV)， $q(i, j)$ 为第 j 个单笔 ETF 在第 i 天的实时净值均值。

② 平均单位套利收益

为了套利价值的稳定性,需计算出各个 ETF 在一定天数内的平均单位套利收益。对于 ETF 的折价套利收益,我们假设折价套利的一个份额 ETF 的成本费占单位基金净价的 0.1%,那么第 i 天的一个份额的 ETF 折价收益为当天二级市场的 ETF 均价减去当天的单位净价再减去一个份额的成本费,而对于 ETF 的溢价套利一个份额 ETF 的成本费我们假设占单位基金净值的 0.2%,那么第 i 天的一个份额的 ETF 溢价收益为当天二级市场的 ETF 均价减去当天的单位净价再减去一个份额的成本费。因此设第 j 个单笔 ETF 在总交易天数内平均单位套利收益为 a_j ,可得:

$$\left\{ \begin{array}{l} a_{1j} = \frac{\sum_{i=1}^n q(i,j) - iopv(i,j) - iopv(i,j)c_1}{\sum_{i=1}^n A_1(i,j)} \quad j = 1 \dots 98 \\ a_{2j} = \frac{\sum_{i=1}^n iopv(i,j) - q(i,j) - iopv(i,j)c_2}{\sum_{i=1}^n A_2(i,j)} \quad j = 1 \dots 98 \\ a_j = \frac{a_{1j} + a_{2j}}{2} \quad j = 1 \dots 98 \end{array} \right.$$

其中 a_{1j} 和 a_{2j} 分别为第 j 个单笔 ETF 在总交易天数内的折价套利平均收益和溢价套利平均收益, c_1 和 c_2 分别为赎回的手续费和申购的手续费, n 为总交易天数, 每个 $iopv(i,j)$ 为第 j 个单笔 ETF 在第 i 天的参考单位基金净值 (IOPV), $q(i,j)$ 为第 j 个单笔 ETF 在第 i 天的实时净值均值。

(2) ETF 的流动性

同时在选取合适的 ETF 进行套利时。我们需要考虑该 ETF 在市场上的流通性大小,一般来说,一只 ETF 的流通性越大,说明它在市场上的行情是不错的、它的活力是越好的,买卖交易这只 ETF 的人多,说明交易便利越好。因此我们要对每一只 ETF 进行流通性测度的计算,从而从流通性维度来考量这只 ETF 的好坏,假设第 j 个 ETF 的流动性测度为 b_j ^[3]可得约束如下:

$$b_j = \frac{1}{n} \sum_{i=2}^n \frac{amount(i,j)}{|q(i,j) - q(i-1,j)|} \quad j = 1 \dots 98$$

其中 $amount(i,j)$ 为第 i 天第 j 只 ETF 的交易金额, $q(i,j)$ 为第 i 天第 j 只 ETF 在二级市场的均价。

(3) ETF 的最大回撤率

ETF 的最大回撤率是投资者在购买 ETF 时需要参考的比較重要的一个风险参考指标,其反应了 ETF 的净值在某一时间点之后可能出现的最糟糕的情况。对投资

者而言，最大回撤率越低越好，在选取 ETF 时应计算第 j 只 ETF 在 n 日交易天数的最大回撤率 c_j ^[4] 如下：

$$c_j = \frac{\min_{n > x > y} (iopv(x) - iopv(y))}{iopv(x)} \quad j = 1..98$$

其中 y 为 ETF 的净值最大的那一日， x 为 y 日之后 ETF 的净值最小的那一日， n 为 y 日与 x 日区间内交易的总天数。

(4) 单市场套利 ETF 选取

综上所述，将 ETF 的单市场套利价值收益、ETF 的流动性和 ETF 的回撤度均一化并附权重，可得到各 ETF 的适合套利程度指标 d_j 如下：

$$d_j = 0.5 \frac{a_j - a_{\min}}{a_{\max} - a_{\min}} + 0.2 \frac{b_j - b_{\min}}{b_{\max} - b_{\min}} - 0.3 \frac{c_j - c_{\min}}{c_{\max} - c_{\min}} \quad j = 1...98$$

其中 $(a_j - a_{\min}) / (a_{\max} - a_{\min})$ 表示每只 ETF 在 n 天交易的平均收益在 98 只 ETF 排名等级，同理， $(b_j - b_{\min}) / (b_{\max} - b_{\min})$ 表示每只 ETF 在 n 天交易的流通性测度在 98 只 ETF 排名等级， $(c_j - c_{\min}) / (c_{\max} - c_{\min})$ 表示每只 ETF 在 n 天交易的回撤率在 98 只 ETF 排名等级，根据各 ETF 的适合套利指标，选取排名前十的 ETF。

5.1.3 基于折溢价套利模型的算法设计

在上述中我们将折溢价套利模型分成瞬时套利以及延时套利，在套利的流程上两者并大的区别。延时套利相对于瞬时套利多出了对 ETF 价格波动的预测，导致在进行申赎的时机上有所不同。接下来我们先对 ARIMA 时间序列模型进行算法设计，在后对折溢价的操作进行算法设计。

Step1: 数据初始化。我们调用同花顺 iFinD 接口，获取历史三个月分钟级别的 ETF 交易数据。且在训练 ARIMA 模型前，设置 AR 模型的阶数 p 以及 MA 模型的阶数 q 以及预测的步数 n 。依据历史经验来说阶数 $p = 4$ 、阶数 $q = 4$ ，结合延时套利的风险要求设置步数 $n = 5$ 。

Step2: 数据平稳性检验及模型训练。在该步骤上我们需要对 ETF 交易数据进行平稳性检验，依据检验结果对 ETF 交易数据进行差分处理。将差分处理后的数据进行模型训练。依据文献^[5]我们调用自动化程序完成该一步骤。

Step3: 折溢价套利算法设计。依据套利公式实时计算一级市场 IOPV 价格与 ETF 二级市场价格的价格差，当差价能覆盖交易成本时进行折溢价套利操作。计算公式如下：

$$W_1 = p_0 x - \sum_{i=1}^n p_i y_i - c$$

$$W_2 = \sum_{i=1}^n p_i y_i - p_0 x - c$$

其中， W_1 为溢价获利金额当 W_1 大于0时我们选择溢价套利，买入成分股申购ETF卖出ETF。 W_2 为折价获利金额当 W_2 大于0时我们选择折价套利，买入ETF赎回成分股卖出成份股。

5.1.4 单市场选取ETF套利模型的具体结果

综上所述，根据三个指标模型及单市场ETF选取模型，通过同花顺软件平台抓取附件A有98只单市场的ETF，同时我们在同花顺抓取对应日期ETF的单位基金净价，通过matlab编程可以得到这98只单市场的ETF的综合指标值。下表是排名前十只的ETF具体结果（98只ETF的排名情况和对应的综合指标值见附录文件支撑材料）。

表 1：单市场排名前十 ETF

序号	ETF	综合指标值	序号	ETF	综合指标值
1	159814.SZ	0.420660851	6	588090.SH	0.223594104
2	588060.SH	0.379419902	7	159682.SZ	0.180995997
3	159969.SZ	0.31446366	8	159681.SZ	0.173219262
4	159903.SZ	0.275688811	9	510050.SH	0.162901322
5	159975.SZ	0.23285735	10	588180.SH	0.146214495

通过上表数据，我们可以看到属于上海的ETF有9只，而深圳的ETF只有1只。下面我们将对编程出来这98只ETF的三个指标进行分析，如下图：

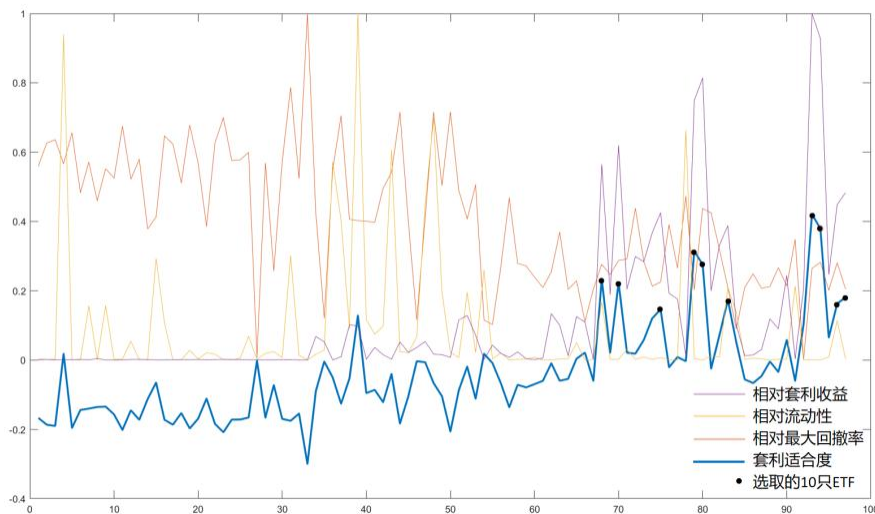


图 6：98 只 ETF 指标和综合指标（套利适合度）折线图

通过上面的折线图，我们可以清楚的看到选取套利的10只ETF它量化后的相对套利收益比起其他ETF是较可观的，大部分是分布在0.43~0.65的范围内，相对最大回撤率是叫其他ETF小，及对应的损失风险小，相对流动性是较其他ETF小很多，这10只ETF的先对流动性大部分在0.02上下波动。而综合指标就是各个ETF选取合适套利的综合排名情况。总体而言，选取套利的ETF是收益良好、亏损风险相对较小的。

5.2 跨境ETF溢折套利分析

在任务二中，我们选择跨境ETF进行套利分析。通过阅读文献^[6]可知跨境ETF的交易规则与国内A股ETF有些许差别。依据跨境ETF的规则我们制作了操作流程图具体如下所示：

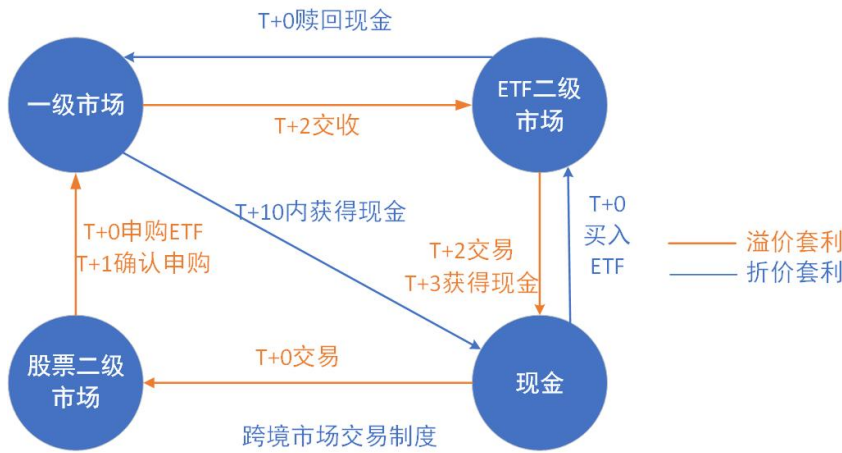


图 7：跨境 ETF 套利流程

在图中橙线为溢价套利流程，蓝色路线为折价套利流程。以溢价套利为例，我们走完一次溢价套利路线需要经过现金买股票，股票申购ETF，卖出ETF获取套利。由于我们不能直接购入境外国股票，我们需要让基金公司代购股票。走完整个申购流程我们可以在T+2（第二天）收到ETF，若我们T+2（第二天）卖出ETF则在T+3（第三天）现金回到我们手中。对于折价套利，我们可以当天在市场买入ETF且向基金公司赎回，赎回后基金公司在T+10（十天）内以现金回到我们手中。

在折价模型中大致流程与任务一折价模型相同，在收取差价上有10天的等待时间。在溢价套利上我们有两个影响套利的的时间点：第一为，T+1确认申购时IOPV价格的波动，第二为，T+2ETF二级市场价格波动。

5.2.1 神经网络预测溢价机会

由于申购的机制我们不能实时的进行套利，故我们以神经网络预测未来一二级市场价格决定我们是否进行溢价操作。

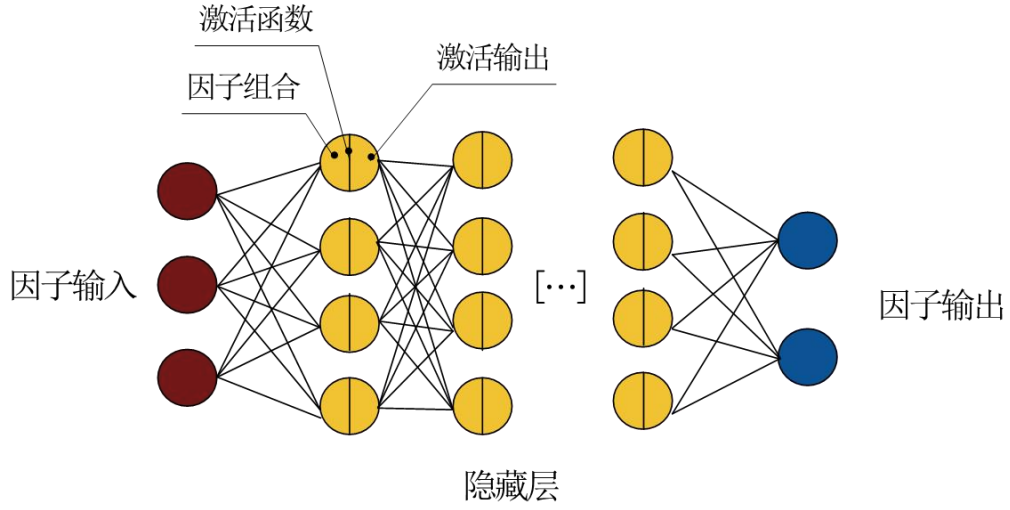


图 8：神经网络模型

神经网络由输入层、隐藏层、输出层构成。我们通过对输入层的因子线性组合以及添加激活函数构成我们的神经网络模型。在输入层我们选择以ETF标的股票价格、ETF一级市场价格、人民币汇率、ETF二级市场价格,作为因子输入。而在输出层我们选择未来一天ETF一级市场价格和未来两天ETF二级市场价格,作为因子输出。以单层隐藏层为例说明,具体如下所示:

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} = \begin{bmatrix} f(s_1) \\ f(s_2) \\ \vdots \\ f(s_m) \end{bmatrix}$$

其中, $f(x)$ 为激活函数; $a_{i,j}$ 为隐藏层第参数权重; x_i 为输入因子; s_i 为因子的线性组合; z_i 为激活输出。我们以激活输出 z_i 当做下一层的输入因子再进行一次计算可获取下一层的激活输出。我们通过多层隐藏层计算后我们可输出目标因子。具体如下所示:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

关于神经网络的训练我们需要将数据集划分三部分，70%训练集、15%测试集、15%验证集。训练集用于训练神经网络即 $a_{i,j}$ 和 $b_{i,j}$ ，而测试集用于调整神经网络的超参数如隐藏层的层数，验证集主要用于衡量神经网络训练效果的好坏。

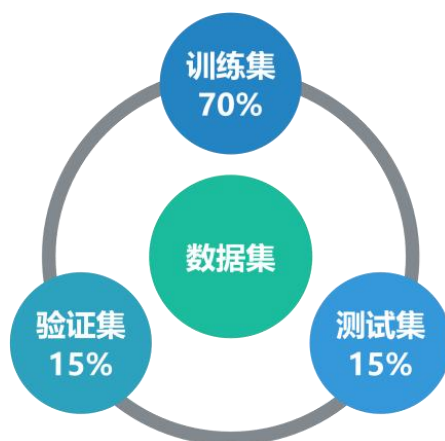


图 9: 数据集划分

在训练神经网络为防止不同量纲以及不同变量量级对训练结果的影响（数值问题、神经网络收敛速度、神经元饱和、大数吞小数），我们考虑对数据集进行归一化处理。具体公式如下所示：

$$X_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

其中， x_{\max} 为该数据集中特征因子的最大值； x_{\min} 为该数据集中特征因子的最小值； x_i 为特征因子； X_i 为归一化特征因子。

在基金走势中通常具有非线性，为使神经网络具有非线性所以我们给神经网络添加激活函数 $f(x)$ 。我们选择修正线性单元 ReLU 充当隐藏层激活函数 $f(x)$ 以及在输出层使用 tansig 当作激活输出函数。

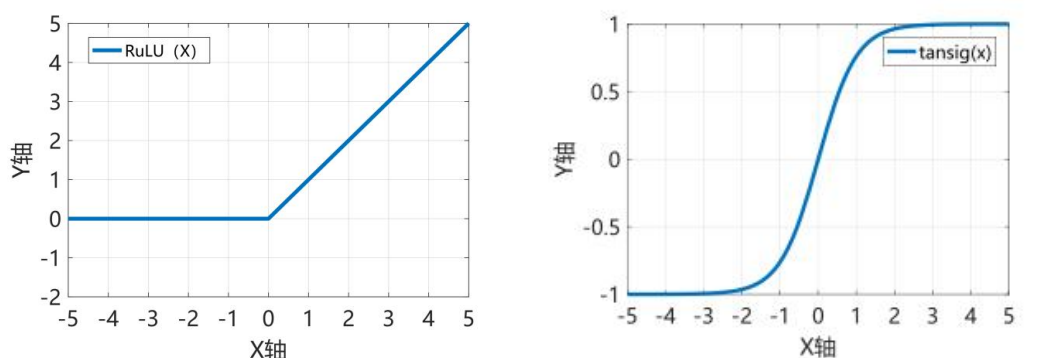


图 10: 激活函数

最后在超参数初始设置上，我们依据经验公式来设置隐藏层中神经元的个数，以及选取训练算法莱文贝格-马夸特方法，它可以提供非线性最小化的数值解。经验公式如下所示：

$$N_h = \frac{N_s}{\alpha(N_i + N_o)}$$

其中， N_i 为输入层因子的个数； N_o 为输出层因子的个数； N_s 为训练集样本的个数； α 为自取的参数取值范围 $[2,10]$ 。

5.2.2 基于神经网络的溢价算法设计

依据上述神经网络理论，结合套利模型，完成预测跨境ETF的神经网络模型具体如下所示：

Step1: 数据获取与数据预处理。以恒生科技ETF（513130）为例，我们利用同花顺iFinD插件获取恒生科技ETF每天（2021年6月1日至2023年10月31日）的历史数据（开盘价、最高价、最低价、收盘价、均价、单位净值），以及离岸汇率。由于ETF数据与离岸汇率数据的获取天数不同，故我们要筛选出同天的数据作输入样本以及输入样本未来1天ETF的单位净值与未来2天的均价。

Step2: 设置超参数。我们设置的超参数有隐藏层神经元的个数、数据集的划分、最大训练次数、训练目标误差、学习率。我们设置训练集为70%、测试集15%、验证集为15%、最大训练数为1000、训练目标误差为 10^{-3} 、学习率为0.1、训练目标误差为由经验公式可得神经元数为3，

$$N_h = \frac{N_s}{\alpha(N_i + N_o)}$$

其中， N_i 为输入层因子的个数； N_o 为输出层因子的个数； N_s 为训练集样本的个数； α 为自取的参数取值范围 $[2,10]$ 。

Step3: 训练神经网络。调用MATLAB神经网络的工具箱训练神经网络并显示训练成果。

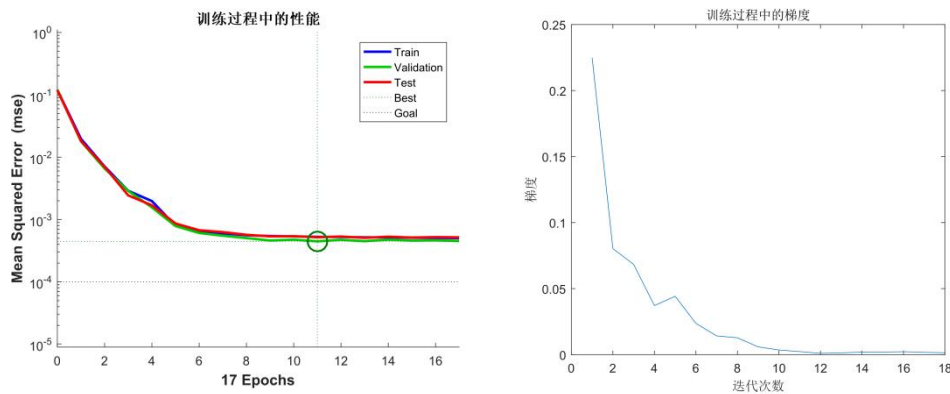


图 11：训练成果

由图中可知我们神经网络训练18次完成收敛，以及误差在 10^{-3} 。

Step4: 申购时机选择。依据文献^[1]我们设申购手续费为0.1%，依据差价公式我们选取能获取利润的申购时间。具体如下所示：

$$\begin{cases} W = p_1 x_1 - p_2 x_2 - \frac{p_2}{1000} \\ W > 0 \end{cases}$$

其中， p_1 为预测ETF卖出的均价； x_1 为ETF的数量； p_2 为预测ETF申购费用； x_2 申购ETF的份额，当利润大于0时既是申购ETF的时机。如下恒生科技ETF（513130）申购时机图。

表 2：申购时机

申购日期	2021/6/3	2021/6/4	2021/6/7	...	2023/10/31
预测获利 (元)	6820.348269	2687.365988	3549.741663	...	2140.423412

5.2.3 基于跨境 ETF 规则的 ETF 选择模型

在对计算 EFT 的跨境套利之前，因先通过 ETF 的选择模型选择出最适合进行套利的部分 ETF。根据上述对跨境 ETF 规则的分析可知，在实现跨境 ETF 的溢价套利时有天数上的延迟，而导致套利时选取该 ETF 的 IOPV 与 ETF 价格存在 1 天的延迟，因此需特殊处理，再结合任务一的 ETF 选择模型可建立任务二的 ETF 的选择模型如下：

(1) 跨境 ETF 的套利价值收益

由于跨境 ETF 折溢价套利交易制度不同，我们需要分开来讨论，对于跨境 ETF 的折价套利交易制度，它实现 T+0 在二级市场买入 ETF、在一级市场赎回现金，因此对于计算跨进 ETF 折价套利的收益，还是跟问题一的折价套利收益模型一致；而对于跨境 ETF 的溢价套利交易制度，它是不同于单市场的套利交易制度，它主要是实行 T+2 交易模式，具体交易流程是：投资者在 T+0 日提交申购跨境 ETF，在 T+1 日确认投资者的申购跨境 ETF，等到 T+2 日投资者才能拿着 T+0 日提交申购跨境 ETF 在 ETF 二级市场进行交易买卖。因此我们计算跨境 ETF 的溢价套利收益，假如投资者在 T+2 日卖出跨境 ETF，那么此时一份额跨境 ETF 的收益应该等于 T+2 日均价减去 T+1 申购的单位基金净价以及一份额跨境 ETF 的手续费，这里跨境 ETF 的溢价手续费仍然设为单位基金净价的 0.1%，折价为 0.2%。通过 iFinD 平台，我们可以抓取到 99 只附件 A 中跨境 ETF 那么跨境 ETF 折溢价套利平均收益的数学模型具体如下：

$$\left\{ \begin{array}{ll} a_{1j} = \frac{\sum_{i=1}^n q(i,j) - iopv(i,j) - iopv(i,j)c_1}{\sum_{i=1}^n A_1(i,j)} & A_1(i,j) = 1, j = 1 \dots 99, i = 1 \dots n \\ a_{2j} = \frac{\sum_{i=1}^n iopv(i-1,j) - q(i,j) - iopv(i-1,j)c_2}{\sum_{i=1}^n A_2(i,j)} & A_2(i,j) = 1, j = 1 \dots 99, i = 1 \dots n \\ a_j = \frac{a_{1j} + a_{2j}}{2} & j = 1 \dots 99 \\ A_1(i,j) \in \{0,1\} & i = 1 \dots n; j = 1 \dots 99 \\ A_2(i,j) \in \{0,1\} & i = 1 \dots n; j = 1 \dots 99 \\ c_1 = 0.002 \\ c_2 = 0.001 \end{array} \right.$$

其中 a_{1j} 和 a_{2j} 分别为跨境第 j 只单笔 ETF 在总交易天数内的折价套利平均收益和溢价套利平均收益， n 为总交易天数， $iopv(i,j)$ 为跨境第 j 个单笔 ETF 在第 i 天的参考单位基金净值 (IOPV)， $q(i,j)$ 为跨境第 j 个单笔 ETF 在第 i 天的实时净值均值。

(2) 跨境 ETF 选取模型

由于 ETF 的流动性测度、回车率的计算内容分别只涉及到 ETF 的交易金额和均价、单位基金净价，交易制度对其计算模型没有影响，因此它们的模型不需要改动。通过综合跨境 ETF 的平均收益、跨境 ETF 的流动性测度、跨境 ETF 的回车率三个指标来选取出最适合跨境套利的前十只跨境 ETF。具体的综合选取跨境 ETF 的数学模型如下：

$$\left\{ \begin{array}{ll} a_{1j} = \frac{\sum_{i=1}^n q(i,j) - iopv(i,j) - iopv(i,j)c_1}{\sum_{i=1}^n A_1(i,j)} & A_1(i,j) = 1, j = 1 \dots 99, i = 1 \dots n \\ a_{2j} = \frac{\sum_{i=1}^n iopv(i-1,j) - q(i,j) - iopv(i-1,j)c_2}{\sum_{i=1}^n A_2(i,j)} & A_2(i,j) = 1, j = 1 \dots 99, i = 1 \dots n \\ a_j = \frac{a_{1j} + a_{2j}}{2} & j = 1 \dots 99 \\ b_j = \frac{1}{n} \sum_{i=2}^n \frac{amount(i,j)}{|q(i,j) - q(i-1,j)|} & j = 1 \dots 99 \end{array} \right.$$

$$\left\{ \begin{array}{l} c_j = \frac{\min_{n>x>y} (iopv(x) - iopv(y))}{iopv(x)} \quad j = 1 \dots 99 \\ d_j = 0.5 \frac{a_j - a_{\min}}{a_{\max} - a_{\min}} + 0.2 \frac{b_j - b_{\min}}{b_{\max} - b_{\min}} - 0.3 \frac{c_j - c_{\min}}{c_{\max} - c_{\min}} \quad j = 1 \dots 99 \\ A_1(i, j) \in \{0, 1\} \quad i = 1 \dots n; j = 1 \dots 99 \\ A_2(i, j) \in \{0, 1\} \quad i = 1 \dots n; j = 1 \dots 99 \\ c_1 = 0.002 \\ c_2 = 0.001 \end{array} \right.$$

5.2.4 选取跨境 ETF 模型套利的具体结果

通过跨境 ETF 的选取模型，以及利用同花顺软件平台抓取到附件 A 数据中跨境 ETF 的有 99 只和对应跨境 ETF 交易日期的市场均价、单位基金份额的净价，通过 *matlab* 编程可以得到如下表排名前十的跨境 ETF（99 只跨境 ETF 的排名情况和对应的综合指标见文件支撑材料）：

表 3：跨境市场排名前十 ETF

序号	ETF	综合指标值	序号	ETF	综合指标值
1	159687. SZ	0.405707773	6	513330. SH	0.085845705
2	513690. SH	0.194422182	7	513060. SH	0.062649272
3	513900. SH	0.170628266	8	513980. SH	0.057775108
4	159788. SZ	0.129960096	9	513070. SH	0.053870267
5	510900. SH	0.114050616	10	513030. SH	0.047775475

从表中的数据可以看出深圳的跨境 ETF 只有 2 只，而上海的跨境 ETF 有 8 只。下面我们将对 99 只跨境 ETF 平均收益、流动性测度、回撤率三个指标进行分析，如下图所示：

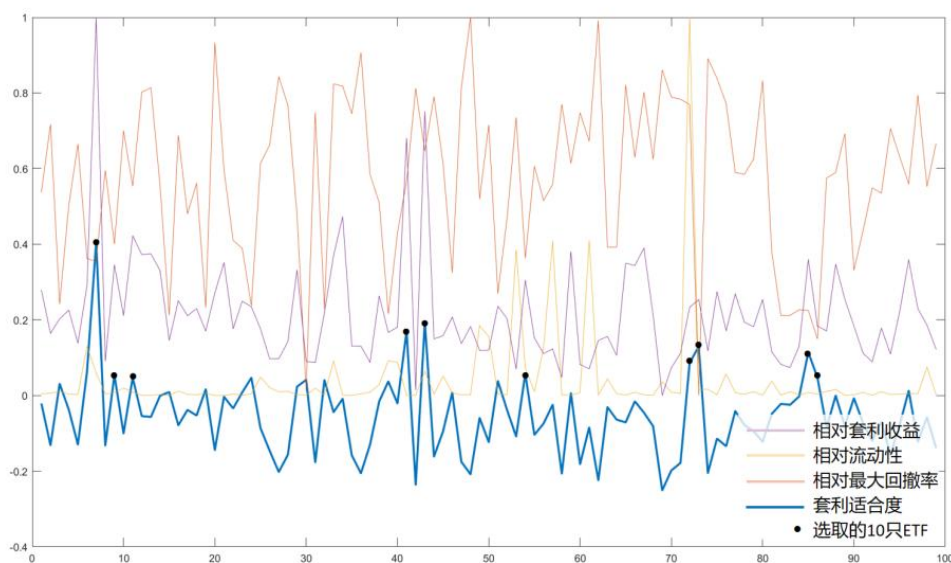


图 12：跨境 ETF 指标和综合指标（套利适合度）

通过上面的折线图，综合指标排名前10只跨境ETF对应其量化后的相对套利收益大部分是落在0.48~1的区间范围内，收益较为可观，它的相对流动性先对平稳，大部分在0.01~0.03范围波动，它的相对最大回撤率相对其他ETF是小很多的，主要落在0.24~0.4范围波动。总体来说排名前十ETF收益较好，亏损风险较小。

5.3 跨境ETF与股指期货的跨市场套利分析

在期现市场上，我们进行期现套利的做法有两种。第一种为：低价买入现货高价卖出期货。第二种为：高价卖出现货低价买入期货。通过两种做法我们赚取差价获取套利，但需要注意的是要控制好交易的成本（保证金）。具体期现套利分析图如下：

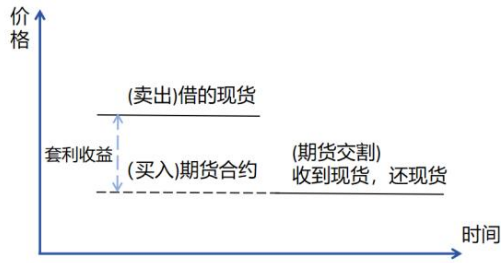


图 13：高卖现货低买期货

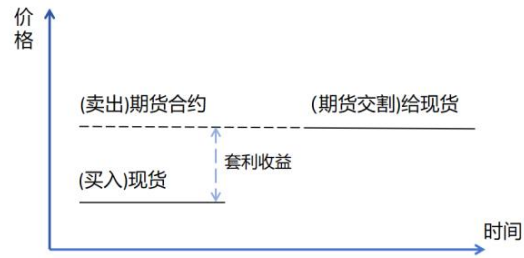


图 14：低买现货高卖期货

依据期货交易的规则，对于我们实现第一种情况套利的方式有：做空开仓做多平仓和做多开仓做空平仓。做空开仓的意思是我们先卖出高价期货合约，由于合约要求我们必须持有货物，所以我们在开仓后低价买入货物平仓。而对于做多开仓做空平仓则与之相反，我们先低价买入货物开仓，然后高价卖出期货合约平仓。

对于第二种情况套利方式有：当股指期货贴水时，我们向券商借现货然后高价卖出现货，以及低价买入股指期货。最后在未来时刻将股指期货换成现货还给券商。

5.3.1 跨境ETF选取模型

对于跨境ETF来说若想完成期现套利，需要跨境ETF复现现货即：在交割日前，跨境ETF走势与现货走势基本一致，在交割日时跨境ETF与期货收敛。我们以标准差法衡量跨境ETF的跟踪偏离度，具体公式如下所示：

$$\begin{cases} TD_{it} = R_i - R_m \\ TE_i = \sqrt{\frac{\sum_{t=1}^n (TD_{it} - \overline{TD}_i)^2}{n-1}} \end{cases}$$

其中， TD_{it} 为基金*i*在时间*t*内跟踪偏离度； R_i 为基金*i*的净值增长率； R_m 为

基准收益率； \overline{TD}_i 为基金*i*在跟踪时产生偏离度的样本均值；*n*为样本数； TE_i 为第*i*支跨境ETF的跟踪偏离度。

5.3.2 跨境ETF选取模型的算法步骤

Step1:数据获取。我们在同花顺iFinD软件下导出上证50、沪深500、中证500、中证1000日期为2023/1/1至2023/10/31日的现货价格。以及99支跨境ETF的收盘价。

Step2:计算跟踪偏离度。依据跟踪偏离度我们要先计算各个跨境ETF和现货的收益率。收益率计算为当天收盘价与前天的收盘价之差除以前天收盘价公式如下：

$$R = \frac{x_{t+1} - x_t}{x_t}$$

依据已有跟踪偏离度计算公式计算 TD_{ti} 以及对 TD_{ti} 求取平均值，由于某些跨境ETF的开通日期不在2023/1/1之前，故求取 TD_{ti} 时要添加判断数值是否为NAN或inf。最后跨境ETF的跟踪偏离度计算 TE_i ：

$$\begin{cases} TD_{ti} = R_i - R_m \\ TE_i = \sqrt{\frac{\sum_{t=1}^n (TD_{ti} - \overline{TD}_i)^2}{n-1}} \end{cases}$$

5.3.3 跨境ETF选取模型结果

依据算法我们对每支跨境ETF进行了与上证50、沪深300、中证500、中证1000现货的跟踪偏离度计算。具体如下所示：

表 4：跨境 ETF 跟踪偏离度

序号	跨境 ETF	沪深 300	中证 500	中证 1000	上证 50
1	159501.SZ	1.23%	1.21%	1.27%	1.28%
2	159506.SZ	1.47%	1.48%	1.52%	1.54%
3	159509.SZ	1.54%	1.58%	1.64%	1.58%
4	159513.SZ	1.22%	1.24%	1.31%	1.26%
5	159519.SZ	0.75%	0.87%	1.02%	0.70%
...
98	513980.SH	1.26%	1.36%	1.42%	1.31%
99	513990.SH	0.76%	0.91%	1.02%	0.77%

下图是对跨境ETF跟踪偏离度进行分析：

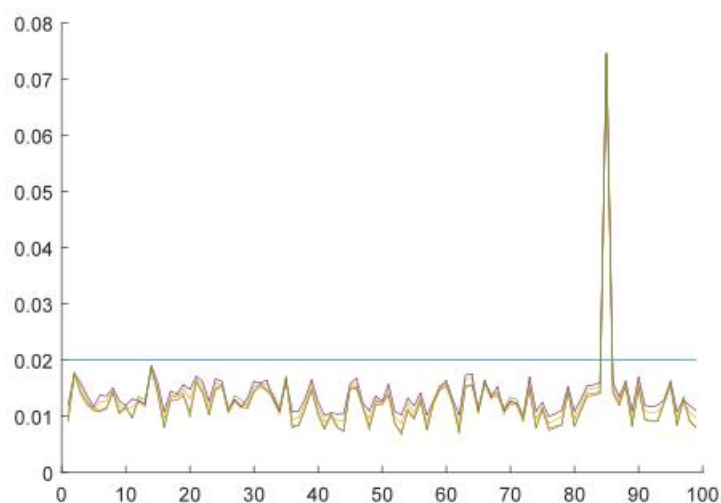


图 15: 跨境 ETF 跟踪偏离度

由图中可知大部分跨境 ETF 与上证 50、沪深 300、中证 500、中证 1000 现货的跟踪偏离度在 2%以内，特别的对于跨境 ETF159747. SZ 来说对于四支现货的跟踪偏离度高达 7.45%、7.45%、7.44%、7.46%。最后我们对每支现货我们选取最好的跨境 ETF 进行期限套利。具体如下所示：

表 5: 选择跨境 ETF

现货	跨境 ETF	跟踪偏离度
沪深 300	513950. SH	0.71%
中证 500	159519. SZ	0.87%
中证 1000	159699. SZ	1.00%
上证 50	514950. SH	0.67%

依据跟踪偏离度计算公式可知，跟踪偏离度越小表示选取的跨境 ETF 越契合。我们将上证 50、沪深 300、中证 500、中证 1000 四组数据中选取跟踪偏离度最小的跨境 ETF 进行与上证 50、沪深 300、中证 500、中证 1000 股指期货组合。组合如下：沪深 300 股指期货和 513950. SH、中证 500 股指期货和 159519. SZ、中证 1000 股指期货和 159699. SZ、上证 50 和 514950. SH。

六、模型评价与推广

6.1 模型优点

(1) 本文的模型充分考虑不同市场的交易制度，同时能够根据不同市场折溢价套利的不同能够设计时间序列算法、神经网络算法来预测未来 ETF 的价格，从而帮助投资者选择不同的套利方法达到获利目的。

(2) 通过对选取 ETF 设置三个参考指标，能够帮助投资者选取出更加全面、

科学、合适的 ETF 进行套利。

6.2 模型缺点

(1) 对于选取跨境 ETF 时, 本文没有考虑时差、外汇、舆论、投资者情绪等因素对跨境 ETF 套利的影响。

(2) 本文没有考虑 ETF 市场价格持续跌落的情况, 同时没有对 ETF 的套利进行对冲策略、套利风险方面做分析。

6.3 模型的推广

本文的模型充分考虑不同交易市场的交易制度, 可以帮助投资者快速了解不同市场的交易规则, 针对不同市场的套利, 本文能够对投资者给出选取最合适套利 ETF 的建议。

七、参考文献

[1][ETF 套利策略详解 套利的条件有两种, 第一, 存在两个市场, 同一产品在不同市场存在价格差; 第二, 同一市场存在两个价格。今天跟大家具体聊一聊 ET... - 雪球 \(xueqiu.com\)](#)

[2][时间序列模型\(四\): ARIMA 模型 - 知乎 \(zhihu.com\)](#)

[3][1] 吴偲立, 常峰源. ETF, 股票流动性与流动性同步性[J]. 经济学(季刊) (202102) [2023-11-06].

[4] 王超. 基于最大回撤的风险平价模型大类资产配置研究[D]. 青岛大学, 2019. DOI:10.27262/d.cnki.gqdau.2019.001887

[5][时间序列 \(ARIMA\) 模型及其 matlab 实现 - 知乎 \(zhihu.com\)](#)

[6][8、跨境 ETF | 上海证券交易所 | 基金网站 \(sse.com.cn\)](#)

[7] 吴玉霞, 温欣. 基于 ARIMA 模型的短期股票价格预测[J]. 统计与决策, 2016, (23):83-86. DOI:10.13546/j.cnki.tjyjc.2016.23.051

[8] 王满. Design of Programmed Trading Strategy for Futures Spot Arbitrage of Stock Index Futures[D]. 西北师范大学, 2022. DOI:10.27410/d.cnki.gxbfu.2022.002141

附录

一、支撑材料的文件列表

1. 问题一选取指标.xlsx
2. 神经网络预测恒生 ETF 申购时机与获利.xlsx
3. 问题二选取指标.xlsx
4. 跨境 ETF 跟踪偏离度.xlsx
5. findPQ.m
6. myARIMA.m
7. p2.m
8. p3.m
9. 问题一选取代码.m
10. 问题二选取代码.m

二、程序代码

1.任务一的 Matlab 程序代码

任务一的程序名 1: ARIMA 模型预测

```
% data 历史数据 stpe 预测步数 pmax 最高 p 阶 qmax 最高 q 阶
function [alldata, diffn, p, q] = myARIMA(data, stpe, pmax, qmax)
dcd=data;%预测数据
diffn=0;%差分次数
ddcd=dcd;
%平稳性检验且计算差分次数
while(1)
    %进行平稳性检验
    dcd_adf = adftest(ddcd);
    dcd_kpss = kpsstest(ddcd);
    if(dcd_adf==1&& dcd_kpss==0)
        %通过平稳性检验
        break;
    else
        %未通过平稳性检验对 dcd 取差分
        ddcd = diff(ddcd);
        diffn=diffn+1;
    end
end
%确定阶数
[p, q]=findPQ(ddcd, pmax, qmax);
%构建模型
Mdl = arima(p, diffn, q);
EstMdl = estimate(Mdl, dcd);
%模型预测
```



```
[forData, YMSE] = forecast(EstMdl, stpe, 'Y0', dcd);
lower = forData - 1.96*sqrt(YMSE); %95 置信区间下限
upper = forData + 1.96*sqrt(YMSE); %95 置信区间上限
alldata=[forData, lower, upper]; %预测数据
end
```

1.任务一的 Matlab 程序代码

任务一的程序名 2：单市场套利的 ETF 选取模型

```
iopv1=xlsread(' ?D:\桌面\大湾区\data1\iopv1.xlsx', 'i'); %ETF 的 iopv
averagel=xlsread(' C:\Users\XIAOAI\Documents\MATLAB\大湾区
\data1\averagel.xlsx', 'x'); %ETF 的价格日均价
turnover1=xlsread(' C:\Users\XIAOAI\Documents\MATLAB\大湾区
\data1\turnover1.xlsx', 'a'); %ETF 的日交易额
a=zeros(1,97); %各 ETF 套利收益
for j=1:1:97
    na=0;
    for i=1:1:60

if((iopv1(i,j)~=0)&&(averagel(i,j)~=0)&&(iopv1(i,j)~=averagel(i,j)))

if(iopv1(i,j)-averagel(i,j)-iopv1(i,j)*0.002>0) %折价套利
收益>0
        a(j)=a(j)+iopv1(i,j)-averagel(i,j)-iopv1(i,j)*0.002;
        na=na+1;
    end

if(averagel(i,j)-iopv1(i,j)-iopv1(i,j)*0.001>0) %溢价套利
收益>0
        a(j)=a(j)+averagel(i,j)-iopv1(i,j)-iopv1(i,j)*0.001;
        na=na+1;
    end
    end
    end
    a(j)=a(j)/na;
end
b=zeros(1,97); %各 ETF 流动性测度
for j=1:1:97
    n=0;
    s=1;
    iopv11=0.0;
    for i=1:1:60
        if(iopv1(i,j)~=0&&turnover1(i,j)~=0)
```

```

        if(s==0&&iopv1(i,j)~=iopv11)
b(j)=b(j)+turnover1(i,j)/(abs(iopv1(i,j)-iopv11)*1000000000);    %流动性
性测度公式
            iopv11=iopv1(i,j);
            n=n+1;
        end
        if(s==1)
            s=0;
            n=n+1;
            iopv11=iopv1(i,j);
        end
    end
end
if(b(j)~=0&&n~=0)
b(j)=(1/n)*b(j);
end
end
c=zeros(1,97);                %各 ETF 最大回撤率
iopvmax=max(iopv1);           %各 ETF 在 n 天交易日内的最大 iopv
iopvmin=zeros(1,97);
for j=1:1:97
    m=find(iopvmax(j));
    iopvmin(j)=iopvmax(j);
    for i=m:1:60                %iopv 最大交易日往后寻找最小 IOPV
        if(iopvmin(j)>iopv1(i,j)&&iopv1(i,j)~=0)
            iopvmin(j)=iopv1(i,j);
        end
    end
end
end
for j=1:1:97                    %最大回撤率计算
    c(j)=(iopvmin(j)-iopvmax(j))/iopvmin(j);
end
a1=zeros(1,97);
b1=zeros(1,97);
c1=zeros(1,97);
for j=1:1:97                    %均一化
    a1(j)=(a(j)-min(a))/(max(a)-min(a));
    b1(j)=(b(j)-min(b))/(max(b)-min(b));
    c1(j)=(c(j)-min(c))/(max(c)-min(c));
end
for j=1:1:97                    %赋权重求和得到选股指标
    d1(j)=0.5*a1(j)+0.2*b1(j)-0.3*c1(j);
end
for i = 1:97-1                    %将选股指标按照从小到达排序
    for j = 1:97-i
        if d1(j) < d1(j+1)

```

```

        temp = d1(j);
        d1(j) = d1(j+1);
        d1(j+1) = temp;
        s=num(j);
        num(j)=num(j+1);
        num(j+1)=s;
    end
end
end
xlswrite('D:\桌面\大湾区\data1\xuannum1.xlsx', num);
xlswrite('D:\桌面\大湾区\data1\xuandata1.xlsx', d1);

```

2.任务二的 Matlab 程序代码

任务二的程序名 1：神经网络训练

```

clear; close all; clc;
%提取历史数据
get_1=xlsread('D:\桌面\恒生科技 ETF(513130).xlsx','ETF 行情');
get_2=xlsread('D:\桌面\恒生科技 ETF(513130).xlsx','离岸汇率');
get_input=[];%输入因子
get_out=[];%输出因子
%% 遍历 ETF 行情数据，依据时间匹配离岸汇率
numb=[1,1];
for i=1:size(get_1,1)
    % 判断时间是否相同
    if get_2(numb(2),1)<get_1(i,1)
        % 翻阅汇率数据
        for j=numb(2):size(get_2,1)
            if get_2(j,1)<get_1(i,1)
                numb(2)=numb(2)+1;
            else
                break;
            end
        end
    end
    if get_2(numb(2),1)==get_1(i,1)
        get_input=[get_input;get_1(i,:),get_2(numb(2),2)];
        numb(1)=numb(1)+1;
    end
end
%% 遍历输入因子，依据时间匹配未来一天的 IOPV, 未来两天的 ETF 均价
n=size(get_input,2)+1;
for i=1:size(get_input,1)-2
    if get_input(i+1,1)-get_input(i,1)==1
        if get_input(i+2,1)-get_input(i,1)==2
            get_input(i,n)=1;

```

```

        get_out=[get_out;get_input(i+1,7),get_input(i+2,6)];
    end
end
end
%% 获取输入因子
get_input2=[];%输入因子
for i=1:size(get_input,1)
    if get_input(i,size(get_input,2))==1
        get_input2=[get_input2;get_input(i,2:8)];
    end
end
%% 建立神经网络
% 创建输入和目标数据
input_data = get_input2';
target_data = get_out';
% 创建神经网络
aerf=0.2;
% 经验公式：样本总数/ $\alpha$ （输入因子数+输出因子数）
N=size(get_input2,1)/(aerf*(size(get_input2,2)+size(get_out,1)));
N=ceil(N);
net = feedforwardnet(N);

% 设置训练参数
net.trainParam.epochs = 1000; % 最大训练轮次
net.trainParam.goal = 0.0001; % 训练目标误差
net.trainParam.lr = 0.1; % 学习率

% 划分数据集
% 将所有数据用于训练
net.divideParam.trainRatio = 0.7;
net.divideParam.valRatio = 0.15;
net.divideParam.testRatio = 0.15;
%%
% 训练神经网络
[trained_net, tr] = train(net, input_data, target_data);

% 测试神经网络
test_data = input_data;
test_output = trained_net(test_data);

% 将输出结果转换为二进制
binary_output = round(test_output);

% 显示结果
disp('测试数据: ');
disp(test_data');

```

```

disp(' 测试输出: ');
disp(binary_output);

% 可视化神经网络结构
view(trained_net);

% 可视化训练过程中的性能
figure;
plotperform(tr);
title(' 训练过程中的性能');

% 可视化训练过程中的梯度
figure;
plot(tr.gradient);
xlabel(' 迭代次数');
ylabel(' 梯度');
title(' 训练过程中的梯度');
%% 申购时机
% X 为输入因子, Y 为[净值、均价]
X=get_input(:,2:8)';
Y=sim(trained_net,X);
a=0.001;%手续费
cash=1000000;%资金
ETF_numb=1000000;%ETF 份额
date=[];
for i=1:size(Y,2)
    if cash>Y(1,i)*ETF_numb
        % ETF 均价*ETF 份额-申购成本-手续费
        W=Y(2,i)*ETF_numb-Y(1,i)*ETF_numb-Y(1,i)*ETF_numb*a;
        if W>0
            date=[date;get_input(i,1),W];
        end
    end
end
end

```

2.任务二的 Matlab 程序代码

任务二的程序名 2：跨境 ETF 套利的选取模型

```

iopv1=xlsread(' D:\桌面\大湾区\data2\iopv2.xlsx','i');
averagel=xlsread(' C:\Users\XIA0AQI\Documents\MATLAB\大湾区
\data2\average2.xlsx','x');
turnover1=xlsread(' C:\Users\XIA0AQI\Documents\MATLAB\大湾区
\data2\turnover2.xlsx','a');
a=zeros(1,99); %各 ETF 套利收益
for j=1:1:99

```

```

na=0;
for i=2:1:60

if((iopv2(i-1,j)~=0)&&(average2(i,j)~=0)&&(iopv2(i-1,j)~=average2(i,j)))

if(average2(i,j)-iopv2(i-1,j)-iopv2(i-1,j)*0.001>0) %溢价
套利收益>0

a(j)=a(j)+average2(i,j)-iopv2(i-1,j)-iopv2(i-1,j)*0.001;
na=na+1;
end
end

if((iopv2(i,j)~=0)&&(average2(i,j)~=0)&&(iopv2(i,j)~=average2(i,j)))

if(iopv2(i,j)-average2(i,j)-iopv2(i,j)*0.002>0) %折价
套利收益>0

a(j)=a(j)+iopv2(i,j)-average2(i,j)-iopv2(i,j)*0.002;
na=na+1;
end
end
end
a(j)=a(j)/na;
end
b=zeros(1,99); %各 ETF 流动性测度
for j=1:1:99
n=0;
s=1;
iopv22=0.0;
for i=1:1:60
if(iopv2(i,j)~=0&&turnover2(i,j)~=0)
if(s==0&&iopv2(i,j)~=iopv22)

b(j)=b(j)+turnover2(i,j)/(abs(iopv2(i,j)-iopv22)*10000000000); %流动
性测度公式

iopv22=iopv2(i,j);
n=n+1;
end
if(s==1)
s=0;
n=n+1;
iopv11=iopv2(i,j);
end
end
end
end
end

```

```

        if(b(j)~=0&&n~=0)
            b(j)=(1/n)*b(j);
        end
    end
    c=zeros(1,99); %各 ETF 最大回撤率
    iopvmax=max(iopv2); %各 ETF 在 n 个交易日内的最大 iopv
    iopvmin=zeros(1,99);
    for j=1:1:99
        m=find(iopvmax(j));
        iopvmin(j)=iopvmax(j);
        for i=m:1:60 %iopv 最大交易日往后寻找最小 IOPV
            if(iopvmin(j)>iopv2(i,j)&&iopv2(i,j)~=0)
                iopvmin(j)=iopv2(i,j);
            end
        end
    end
    end
    for j=1:1:99 %最大回撤率计算
        c(j)=(iopvmin(j)-iopvmax(j))/iopvmin(j);
    end
    a1=zeros(1,99);
    b1=zeros(1,99);
    c1=zeros(1,99);
    for j=1:1:99 %均一化
        a1(j)=(a(j)-min(a))/(max(a)-min(a));
        b1(j)=(b(j)-min(b))/(max(b)-min(b));
        c1(j)=(c(j)-min(c))/(max(c)-min(c));
    end
    for j=1:1:99 %赋权重求和得到选股指标
        d1(j)=0.5*a1(j)+0.2*b1(j)-0.3*c1(j);
    end
    for i = 1:99-1 %将选股指标按照从小到达排序
        for j = 1:99-i
            if d1(j) < d1(j+1)
                temp = d1(j);
                d1(j) = d1(j+1);
                d1(j+1) = temp;
                s=num2(j);
                num2(j)=num2(j+1);
                num2(j+1)=s;
            end
        end
    end
    end
    xlswrite('D:\桌面\大湾区\data2\xuannum2.xlsx',num2);
    xlswrite('D:\桌面\大湾区\data2\xuandata2.xlsx',d1);

```

3.任务三的 Matlab 程序代码

任务三的程序名 1：跟踪偏移度

```
%% 计算跟踪偏离度
% 读取数据
get_1=[];get_2=[];
get_1=xlsread(' ?D:\桌面\现货价格.xlsx');
get_2=xlsread(' ??D:\桌面\跨境 ETF 收盘价.xlsx');
% 计算收益率 R
R_1=[];R_2=[];
% 遍历现货价格数据, 计算收益率
for j=1:size(get_1,2)
    for i=1:size(get_1,1)-1
        R_1(i,j)=(get_1(i+1,j)-get_1(i,j))/get_1(i,j);
    end
end
% 遍历跨境 ETF 数据, 计算收益率
for j=1:size(get_2,2)
    for i=1:size(get_2,1)-1
        R_2(i,j)=(get_2(i+1,j)-get_2(i,j))/get_2(i,j);
    end
end
% 计算每天偏离度
TD=cell(4,1);
for k=1:size(R_1,2)
    for j=1:size(R_2,2)
        for i=1:size(R_2,1)
            TD{k}(i,j)=R_2(i,j)-R_1(i,k);
        end
    end
end
% 计算平均数
TD_Avg=zeros(size(R_2,2),size(R_1,2));
n=zeros(size(R_2,2),size(R_1,2));
for i=1:size(TD,1)
    for j=1:size(TD{i},2)
        for k=1:size(TD{i},1)
            if ~isnan(TD{i}(k,j))&&TD{i}(k,j)~=inf
                TD_Avg(j,i)=TD_Avg(j,i)+TD{i}(k,j);
                n(j,i)=n(j,i)+1;
            end
        end
        TD_Avg(j,i)=TD_Avg(j,i)/n(j,i);
    end
end
end
```



```

% 计算偏离度
TE=zeros(size(R_2,2),size(R_1,2));
for i=1:size(TD,1)
    for j=1:size(TD{i},2)
        for k=1:size(TD{i},1)
            if ~isnan(TD{i}(k,j))&&TD{i}(k,j)~=inf
                TE(j,i)=TE(j,i)+(TD{i}(k,j)-TD_Avg(j,i))^2;
            end
        end
        TE(j,i)=(TE(j,i)/n(j,i))^0.5;
    end
end

hold on
plot([1 99],[0.02 0.02])
plot(TE)

```