

承 诺 书

我们已阅读 2013 广东高校金融建模邀请赛竞赛章程。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们大赛网站公布论文，以供网友之间学习交流，大赛网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队名为：

参赛队员（签名）：

队员 1：

队员 2：

队员 3：

参赛选题组别：【普通组】 ☐

【高频组】 ☐

编 号 专 用 页

竞赛统一编号（由竞赛组委会编号）：

竞赛评阅编号（由竞赛评委会编号）：

评阅记录（由 2-3 位评委评阅）

评阅人			
评分			
备注			

基于均线信息组合的趋势型量化交易模型

郑梓豪、吴亚林、谢森鹏

摘要 本团队的建模思路，是从对市场的认识出发，进而到模型的建立。即通过对指数运动内在规律的认识，通过数学表达和数学模型对这种规律进行**描述、解释和预测**。本文提出的基于均线信息组合的趋势型量化交易模型，反映了我们对指数运动规律和数学模型的几个认识：

首先，当短期趋势向上突破长期趋势，指数具有上行的能量；我们使用三条均线(MA3, MA5, MA7)的组合，对“长期趋势”进行刻画，当短期趋势(当前指数)突破“长期趋势”，便做多。然而，这种信息不是简单相加，而是通过建立其与下一周期的指数间的线性关系($Close^{(i+1)} = \theta_0 + \theta_1 MA3^{(i)} + \theta_2 MA5^{(i)} + \theta_3 MA7^{(i)}$)，通过机器学习算法(Linear Regression)得到的。通过这种简洁的表达，我们尝试对指数运动进行**描述**。此外，这种信息组合，较“5日线上穿20日线则做多”这类交易模型，有更好的表现力(更高的收益与更小的最大回撤)。

其次，模型的核心，即要求解一个刻画“距离”的最优化问题；我们随后研究了其基于范数的刻画方式与其统计含义，得知常用的“二范数”形式，即 $\|\Theta^T x - y\|_2^2$ ，其预测误差服从正态分布；而“一范数”形式，即 $\|\Theta^T x - y\|_1$ ，其预测误差服从拉普拉斯分布。而后的尾部更大，这也反映和**解释**了市场中的“长尾现象”。我们认为，采用“一范数”形式的表达方式，将更贴近于市场的真实情况。模型交易结果表明，利用这种表达方式，将有更优秀的模型表现(后者“**风险收益比**”高出约**2倍**)。此外，为避免“过拟合”，我们为模型添加了正则项。

最后，市场的模式在不断变化。我们通过添加“时间衰减项”，为模型赋予了描述该变化的能力。简单地说，我们并没有将每一条数据视作相同的个体，而是具有相对重要性的个体；而这个重要性是通过时间先后顺序体现的——越“新”的数据具有更高的重要性。在交易中定期更新模型，模型便具有描述指数运动规律变化的能力。模型测试结果表明，这种手段可以进一步提升模型表现。

最终模型在2011、2012年上的实测表现令人满意。其中2011年收益率达到1.96倍（由初始资本50万到145万元），最大回撤3.6%，风险收益比达54.9；2012年收益率达到2.05倍（由初始资本50万到151万元），最大回撤3.4%，风险收益比达59.2。

文章进一步讨论了模型的稳定性和通用性问题。在参数稳定性方面，我们关注最优参数的扰动分析；即在参数最优选择下，给参数施加扰动，观察模型表现的变化。其中，正则项系数 λ 及时间衰减率 α 的扰动对最终收益情况有一定影响，但不显著。这表明，该模型的最优设置处在“参数平原”上。周期通用性上，分别从“交易级别的周期通用性”及“输入变量的k线周期通用性”角度进行讨论。我们将数据调整为5分钟、10分钟级别，并在这些数据上进行测试。结果表明，该交易模型表现有所下降。但改变模型输入的k线周期则对模型影响不大。

关键词：趋势型 均线组合 机器学习算法

1 问题背景分析

1.1 为什么需要量化交易模型？

若要认识量化交易的意义，需要先了解“交易程序化”诞生的源动力。在期货市场中，市场的节奏非常快，对人心理素质的要求很高。普通投资者在机会面前容易“患得患失”：

“看对的行情，赚了一点点开头就跑了，错失了后面大段的收益；做错的头寸死不认输，结果亏损越来越大，一笔交易损失惨重；明明觉得机会来了，又心生畏惧，结果错过了最佳的入场点；明明是错误的判断，却又鬼使神差做了进去，事后后悔不已。”[1]

所以投资结果往往不理想；人们希望能够摆脱自身心态上的缺陷，希望做出的投资决策尽可能理性。这是“交易程序化”诞生的一个重要因素。

程序化交易具有一些人工交易无法比拟的优势：

- **速度快**。市场的机会稍纵即逝，使用程序化交易能迅速捕捉市场信号，提高交易频率。
- **纪律性强**。由于将交易规则和策略进行了程序化，交易系统严格按程序信号进行交易，避免了人为因素的干扰。

程序化交易的实质是其背后的量化交易模型。事实上，若要实现程序化，就不得不用计算机语言（数学语言）去描述我们的交易策略；若要实现盈利，就不得不用数学工具对市场进行**描述、解释和预测**。这样的整个过程，也是建立量化交易模型的过程。本文也将极力呈现这个过程。

1.2 需要什么样的量化交易模型？

虽然量化交易有巨大的优势，实际操作上也表现不俗。但优秀的期货投资者战绩并不输于这些机器。事实上，我们认为，基于对市场理解而建立的量化模型能充分结合理性直觉，具有更优秀的表现力。

我们在模型建立过程中也将秉持这个原则，力图模型的简洁、直观。

2 模型建立与策略制定

“5日线上穿20日线则做多”这种交易规则，是基于对指数运动趋势的认识而制定的策略：当“短期趋势”突破“中长期趋势”，表示此时上行的力量较显著，故可做多；反过来便可做空。在实际测试中，这种简单策略确实能够实现盈利，不过收益率较低，且最大回撤较大。

受到这类趋势型策略的启发，我们首先研究了如何对指数进行描述、解释和预测，确定了以机器学习中经典的“Linear Regression”算法为原型的改进算法；其次，针对

此模型对股价运动的刻画能力，我们为其制定了相应的交易策略。结果显示，相比于liyang2.m¹中的模型，我们的量化交易模型的收益率更高，且最大回撤更小。

2.1 Linear Regression模型研究

我们提出下面的基于均线信息组合的线性模型²：

$$\text{Close}^{(i+1)} = \theta_0 + \theta_1 \text{MA3}^{(i)} + \theta_2 \text{MA5}^{(i)} + \theta_3 \text{MA7}^{(i)} \quad (1)$$

其中 $\text{MA}k^{(i)}$ 为第 i 个时间周期的“ k 周”³移动平均线； θ_i 为待求参数。

机器学习中的Linear Regression算法思想，即给定一组 n 个特征及其 m 条样本，选择最优参数 θ ，使得其线性组合尽可能贴近我们的输出特征[2]。即：

$$\text{minimize } J(\theta) = \frac{1}{m} \sum_{i=1}^m \|\theta^T x^{(i)} - y^{(i)}\|_2^2 \quad (2)$$

其中 $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}$ 为各特征组成的向量， $x_i^{(i)}$ 为输入的各个特征； $y^{(i)}$ 为输出，或理解为我们希望近似⁴的目标。此算法的基本思路是希望利用这些特征，对历史数据反映出的模式进行学习，并将学习到的模式（表现在 θ 的取值上）用于具体生产实践中。

在这里，我们自然也希望对指数中的规律进行学习，并做准确的预测。并希望这些规律是具有稳定性、持久性的。本文选用三条均线作为特征，如 $x^{(i)} = \{\text{MA3}^{(i)}, \text{MA5}^{(i)}, \text{MA7}^{(i)}\}$ ，并以下一周期的价格作为预测值，如 $y^{(i)} = \text{Close}^{(i+1)}$ 。不过，在直接套用上述目标函数的形式之前，我们先做一些分析。

2.1.1 目标函数形式与其统计含义

Linear Regression算法的根本思想，在于“近似”，即“距离”的概念。这个概念，用数学语言便是“范数”；显然，(2)式使用“二范数”对距离进行刻画，当然我们也可以采用其他范数：如“一范数”或是“无穷范数”。不过，使用不同的范数形式，将意味着不同的统计含义，而这很可能影响我们的模型收益。

在这一节，我们研究Linear Regression的目标函数形式与其背后对应的统计现象之间的关系，以期更好地理解模型。

1. “二范数”的统计含义

¹即本次大赛组委会提供的样例

²这里取定了各个输入特征是为了方便读者理解我们的建模出发点；事实上这些特征的选取是需要思考的，我们将在3.3节进行介绍

³若数据为1分钟级别，则 $\text{MA5}^{(i)}$ 表示5分钟线在第 i 时刻的指数

⁴approximate or fitting

问题(2)实际上是在考虑下面的线性估计:

$$y^{(i)} = \theta^T x^{(i)} + v^{(i)}, \quad i = 1, 2, \dots, m \quad (3)$$

我们先考虑 $v^{(i)}$ 是高斯误差的情况, 即 $v^{(i)}$ 是独立同分布的、服从均值为0, 方差为 σ^2 正态分布的数据误差。那么 $y^{(i)}$ 便也是一个随机变量, 不过它的具体分布是由参数 θ 给出的:

$$p_{\theta}(y) = \prod_{i=1}^m p(y^{(i)} - \theta^T x^{(i)}) \quad (4)$$

所以对应的对数似然函数是:

$$l(\theta) = \log p_{\theta}(y) = \sum_{i=1}^m \log p(y^{(i)} - \theta^T x^{(i)}) \quad (5)$$

于是最大似然估计问题便是:

$$\text{maximize} \quad \sum_{i=1}^m \log p(y^{(i)} - \theta^T x^{(i)}) \quad (6)$$

而(4)式可利用误差 $v^{(i)}$ 的已知分布 $p(z) = (2\pi\sigma^2)^{-1/2}e^{-z^2/2\sigma^2}$ 得到:

$$l(\theta) = -(m/2)\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|\theta^T x - y\|_2^2 \quad (7)$$

所以问题(5)显然可以表达为:

$$\text{minimize} \quad \|\theta^T x - y\|_2^2 \quad (8)$$

可以看到, 我们的建模问题(2)和带有高斯误差的线性估计问题(8)是一致的。**也就是说, 当我们采用“二范数”进行最优化求解时, 得到的数据误差是服从正态分布的[3]。**

2. “一范数”的统计含义

对于(3)式, 如果误差 $v^{(i)}$ 是拉普拉斯误差: $p(v) = (1/2a)e^{-|v|/a}$ ($a > 0$), 用同样的思路, 可以知道对该最优化问题的求解等价于:

$$\text{minimize} \quad \|\theta^T x - y\|_1 \quad (9)$$

也就是说, 当我们采用“一范数”进行最优化求解时, 得到的数据误差是服从拉普拉斯分布的[3]。相比于正态分布, 拉普拉斯分布的尾部更加平坦。从分布的尾部大小来说, 后者更符合市场的规律。两种分布的情况请见图1。

到这里, 我们有理由相信, 采用“一范数”形式的模型会有更强的表现力; 有别于一般的Linear Regression模型采用“二范数”度量的目标函数形式; 本文的基础模型是一个用“一范数”度量的线性回归模型, 其目标函数是:

$$\text{minimize} \quad J(\theta) = \frac{1}{m} \sum_{i=1}^m \|\theta^T x^{(i)} - y^{(i)}\|_1 \quad (10)$$

而实际测试结果确实如此, 对模型表现力(回测)的讨论和探究请见文章3.3参数优化部分。

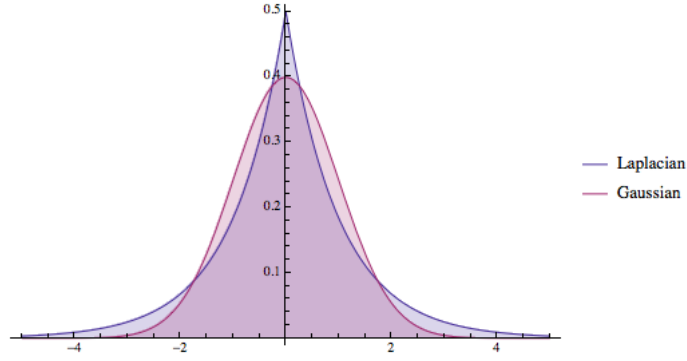


Figure 1: 正态分布与拉普拉斯分布的对比

2.1.2 模型泛化能力与正则项

由于模型对应的优化目标函数倾向于对历史数据尽可能地进行近似，如此会出现因“过拟合”而造成模型泛化能力不足的情况：

从模型上说，我们不希望少数的数据噪声对拟合效果造成太大影响（图2）。直观地理解，即我们不希望模型因指数某几次的“黑天鹅事件”影响了对整体指数运动模式的整体把握。若模型缺乏对整体规律的把握，表现在策略收益上，便是收益率等指标的剧烈波动。

所以我们为(10)配备正则项⁵：

$$\text{minimize } J(\theta) = \frac{1}{m} \sum_{i=1}^m |\theta^T x^{(i)} - y^{(i)}| + \lambda \|\theta\|_2^2 \quad (11)$$

其中参数 λ 具有调整 $J(\theta)$ 前后两项相对比重的作用。

实际测试结果亦显示了正则项对模型表现稳定性的重要性。这部分讨论和探究请见文章3.3参数优化部分。

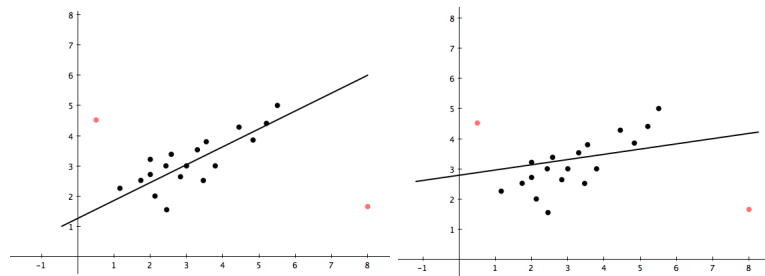


Figure 2: 正则项对拟合效果的影响。右图不具有正则化项，模型因此受到“噪点”影响而对整体数据模式把握欠佳

⁵事实上正则项的范数形式选择也有类似于上面的问题。对此项的讨论请见文献[3]

2.1.3 模型时变性与时间衰减项

对于时间序列问题，一个突出的情况是问题（或现象）的模式随着时间不断变化。同样地，市场的情况也在变化；如果我们的模型具备对“时间”的描述能力，则该模型便能更好地描述实际指数的运动规律。

一种可行的解决方案便是为模型配备时间衰减项[4]：

$$e^{\alpha(T-t)}$$

其中 T 为总的时间跨度， t 为当前时间（ t 越大表示离现在越近）， $\alpha(\alpha > 0)$ 为衰减速率。将此项加入(10)式易知，近期的样本有更大的训练权重（因为旧样本将遭到更大的惩罚）。图3展示了 $e^{0.3(10-t)}$ 的取值情况。

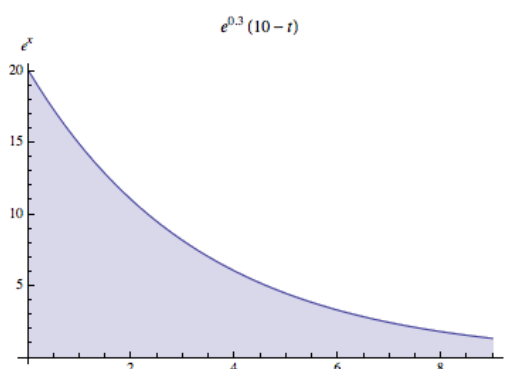


Figure 3: $e^{0.3(10-t)}$

最终，我们得到了一个在统计意义下更贴近市场规律的、具有一定泛化能力的，基于线性回归模型的趋势型模型⁶：

$$\text{minimize } J(\theta) = \frac{1}{m} \sum_{i=1}^m e^{\alpha(m-i)} |\theta^T x^{(i)} - y^{(i)}| + \lambda \|\theta\|_2^2 \quad (12)$$

2.1.4 小结：模型分析

至此，我们试着回答开始时提到的“模型如何对指数进行描述、解释和预测”的问题：

- **为什么选择均线信息组合作为特征？** 我们基于这样一个认识：（我们假设）人们盈利超过10%时，有一部分的人选择平仓离场；当盈利到达20%时又有一部分人平仓；即是说，人们面对不同的盈利规模，有不同强度的套利动机。而恰好均线信息在一定程度上反映了各投资者的成本情况（盈利规模）；且多个均线指标能够反映处在不同盈利水平的情况。

⁶这里的模型事实上是(1)，我们将它以一个优化问题的形式进行描述；其次，时变性的体现在这里便是“趋势”

- 为什么以下一个时间周期的指数作为模型输出？我们希望通过这个模型，了解指数的运动规律，故自然需要将下一个时间周期的指数作为输出。另外，采用这种输出的好处是，在实际进行交易时，在 t 时刻便对 $t + 1$ 时刻的情况有所预期，当 $t + 1$ 时刻的信息到来时，可以尽快采取相应的操作。

我们通过对市场的理解和认识，建立了简洁的量化交易模型。下面进行模型验证：
首先进一步对数据进行描述。矩阵 \mathbf{X} 中的元素为均线指标、 Θ 为待求参数：

$$\mathbf{X} = \begin{bmatrix} 1 & \text{MA3}^{(1)} & \text{MA5}^{(1)} & \text{MA7}^{(1)} \\ 1 & \text{MA3}^{(2)} & \text{MA5}^{(2)} & \text{MA7}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{MA3}^{(m-1)} & \text{MA5}^{(m-1)} & \text{MA7}^{(m-1)} \end{bmatrix} \Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_4 \end{bmatrix}$$

最后，我们依据模型，可将训练数据表示为：

$$\Theta^T \mathbf{X} = [\theta_0 \ \theta_1 \ \theta_2 \ \theta_4] \begin{bmatrix} 1 & 1 & \dots & 1 \\ \text{MA3}^{(1)} & \text{MA3}^{(2)} & \dots & \text{MA3}^{(m-1)} \\ \text{MA5}^{(1)} & \text{MA5}^{(2)} & \dots & \text{MA5}^{(m-1)} \\ \text{MA7}^{(1)} & \text{MA7}^{(2)} & \dots & \text{MA7}^{(m-1)} \end{bmatrix} = \begin{bmatrix} \text{Close}^{(2)} \\ \text{Close}^{(3)} \\ \vdots \\ \text{Close}^{(m)} \end{bmatrix} = \mathbf{y}$$

我们选取IF2011.mat中的前500条1分钟级数据⁷进行训练，利用CVX[5][6]可求解得到最优参数 Θ ⁸，并计算 $\Theta^T \mathbf{X}_{\text{test}}$ 得到 $\mathbf{y}_{\text{predict}}$ ，其中 \mathbf{X}_{test} 为501到1500天的数据， $\mathbf{y}_{\text{predict}}$ 为对应的预测值；我们将预测值与真实值 \mathbf{y}_{test} 进行对比（图4）：

可以看到该模型的预测具有以下特征：

- 总体走势上能够跟随指数的真实情况。且在指数连续上升阶段，预测值大部分在真实值下方；更重要的是，对于指数短暂的“调整”，预测值始终保持了较好的趋势延续性。这为策略制定提供了重要思路。
- 时延性。从指数极值点来看，该模型给出的极值点总是比真实值来得晚；这是模型的一大不足。在后面的分析中我们也可以看到，这是致使策略亏损的原因之一。

进一步通过图4的特点来看，我们能够看到模型中各个项的“身影”：

- 首先是时延性，这是因为模型使用均线组合导致的。所以当指数运动出现拐点时，模型未能及时给予反映。不过，模型具有这样特点，可以这样理解：当趋势转换得到确认时，模型才会出现拐点。这给我们建立交易策略立下基础。

⁷数据抽取方式请见程序data_transformer.m

⁸CVX是Stanford University开发的基于MATLAB语言的凸优化工具箱。利用它可高效方便地求解大部分最优化问题，其用法贴近自然语言；具体请见文献[5]

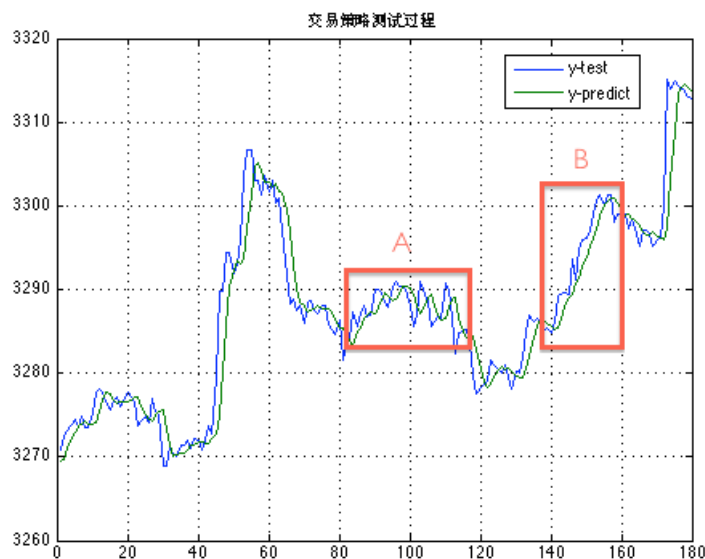


Figure 4: 模型预测效果（只选取了其中一段进行展示）。模型在区域A可能会发生亏损；但在连续上升阶段，例如B区，却对指数暂时的调整具有抗性，会继续保持其预测趋势。

- 其次是模型表现出的良好的“趋势”跟随能力。首先，在指数连续上升阶段，预测值大部分在真实值下方；反过来，当指数出现连续下跌，预测值大部分在真实值上方。这是策略可行性的重要保证。其次，模型对该模式的把握是通过时间衰减项做到的。
- 最后是模型对指数短暂“调整”的抵抗力。这项特性应当是正则项作用的结果；由于引入正则项，预测的指数变得较为平滑。这一个特性，使得我们可以把握住较大的机会。不过同时我们也注意到，当指数出现如图4中A区域的走势，那么可能会出现亏损。

2.2 策略制定

基于上面的模型分析，我们建立一个简单的交易策略：当指数上穿预测值（指数在预测值上方），做多；当指数下破预测值（指数在预测值下方），做空。这个策略在程序实现上是比较方便的。

3 模型测试

我们已经完成了模型建立和交易策略的制定。此前我们是用数学角度去分析模型，并认为具有形式(12)的方案是最佳的。这一部分我们对模型进行测试，以验证数学分析的结果。

在这一部分，我们用1分钟数据进行模型验证。其中，用前500组作为训练数据⁹，此后的所有数据均作为测试数据（即用训练得到的 Θ 得到 y_{predict} 并进行测试）

3.1 目标函数形式

在前面的模型分析中我们提到目标函数中的范数形式及其统计意义。即采用“二范数”形式的模型，其预测误差是一个正态分布，而“一范数”的是一个有更长尾部的拉普拉斯分布。不过，由于量化交易模型是“模型+策略”的结果，所以，在这样的组合下，结果可能会有所不同。下面进行实实验证¹⁰。

分别采用IF2011，IF2012的1分钟级数据¹¹，分别有约5万、6.6万条测试数据。测试结果如下表。其资金走势图及最大回撤情况请见图5、6

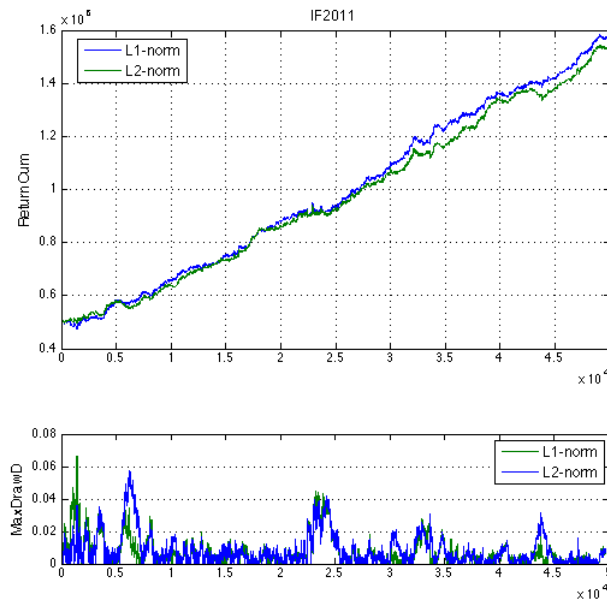


Figure 5: 交易模型2011年的资金走势及最大回撤图。在本年度，两种模型表现相当。

Table 1: 2011年模型表现

	收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
l_1 -norm	1.57×10^6	2.22	0.066	33.6	7.56	21.7
l_2 -norm	1.53×10^6	2.13	0.057	37.2	7.28	25.8

⁹ $\mathbf{X} = \{\text{MA3}^{(i)}, \text{MA5}^{(i)}, \text{MA7}^{(i)}\}$ $\mathbf{y} = \text{Close}^{(i+1)}$ 。其中的均线信息均为当前时间段的真实信息（即在这一分钟，可以通过模型得到下一分钟的预期；当下一分钟到来时，便可以给出买卖信号）

¹⁰依照前文的思路，此时目标函数是(8)和(9)的形式，即 $\|\theta^T \mathbf{X} - \mathbf{y}\|_1$ 和 $\|\theta^T \mathbf{X} - \mathbf{y}\|_2^2$ 。

¹¹转换过程可见 `data_transformer.m`

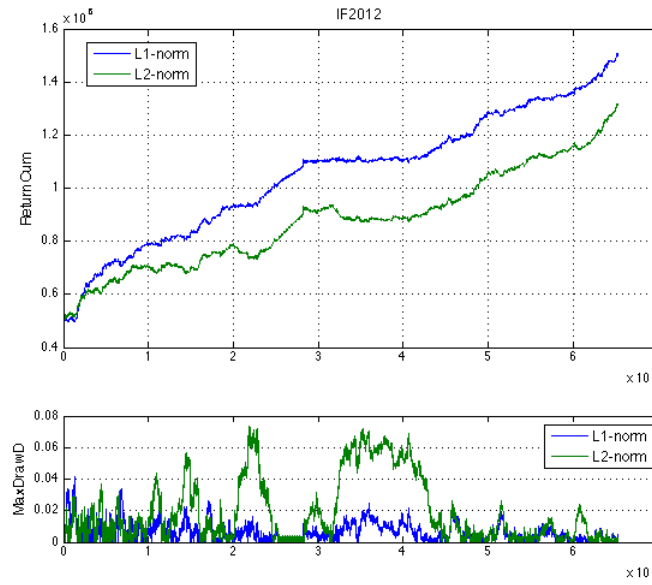


Figure 6: 交易模型2012年的资金走势及最大回撤图。 l_1 -norm模型具有较低的风险，也正因如此，其最终收益表现比后者要高

Table 2: 2012模型表现

	收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
l_1 -norm	1.50×10^6	2.05	0.041	49.7	8.0	26.8
l_2 -norm	1.32×10^6	1.67	0.073	22.8	6.18	18.3

易见 l_1 -norm的表现更好，尤其在2012年，其年化收益和风险收益比均要优于 l_2 -norm。所以从交易的角度说，选择 l_1 -norm是比较好的。

具体地看预测误差的分布图，可见结果与我们的理论分析是相吻合的（图7）：

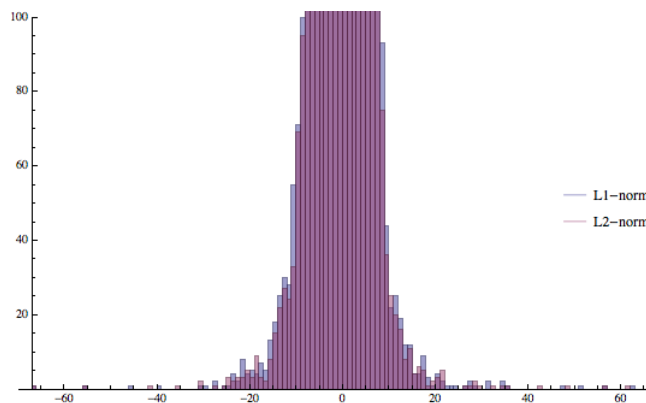


Figure 7: 预测误差的分布图

3.2 正则项

通过上面的对比，我们决定继续沿用“一范数”形式的目标函数。这一节，我们为模型配备正则项，以期望它通过改善拟合平滑度来提高收益。

取正则项系数 $\lambda = 0.1$ ，模型表现结果如下表：

Table 3: l_1 -norm with regularization term模型表现

	收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
2011	1.49×10^6	2.06	0.040	51.1	7.28	24.7
2012	1.47×10^6	2.00	0.038	52.9	7.93	24.1

Table 4: l_1 -norm without regularization term模型表现

	收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
2011	1.57×10^6	2.22	0.066	33.6	7.56	21.7
2012	1.50×10^6	2.05	0.041	49.7	8.0	26.8

从结果中看到，虽然带有正则项的模型在收益上稍逊一筹，但其在最大回撤以及风险收益比中表现较后者更为突出。并且，其各项性能指标在2011和2012年都比较接近，表现了一定的稳定性。进一步地，我们对比了2011年某时段内的交易（图7、8）：

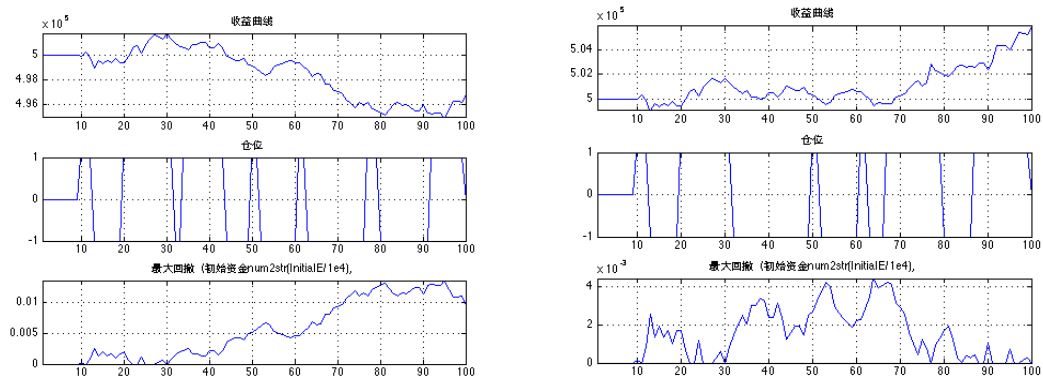


Figure 8: 交易模型在2011年某时段内的资金走势及最大回撤图。左图为不带正则项的模型表现，右图为带有正则项的模型表现。容易看到，当最大回撤不断上升时，不带正则项的模型交易频率明显更多；在指数震荡时，这往往意味着更多的亏损（平仓）

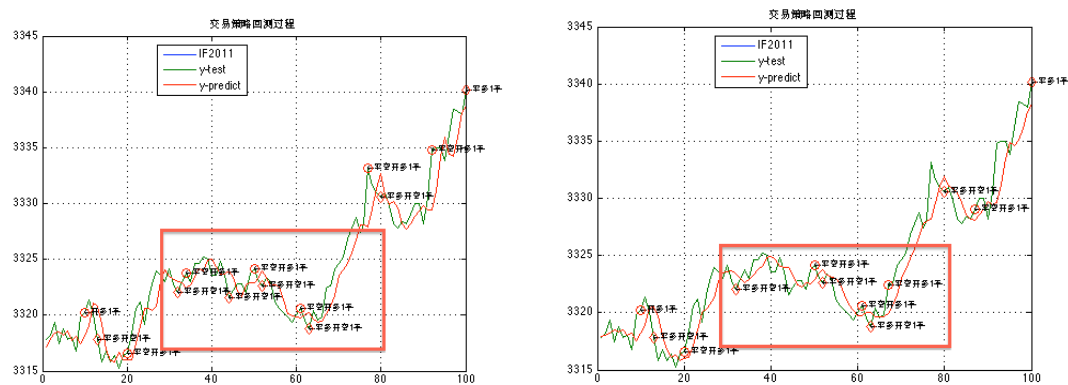


Figure 9: 交易模型在2011年某时段内的资金走势及最大回撤图。左图为不带正则项的模型表现，右图为带有正则项的模型表现。容易看到，不带正则项的策略有更多的止损交易。事实上这是过拟合导致的

3.3 时间衰减项

前面的测试结果，均是利用由一组训练集上得到的 Θ 进行计算的；当考虑时间衰减项的加入后，我们便需要不断训练模型（不断更新 Θ ）。我们以1000条数据作为一组，记作 \mathbf{x}_i 和 \mathbf{y}_i ，记由此组数据训练得到的为 Θ_i 。那么，由该组 Θ_i 通过 $\Theta_i^T \mathbf{x}_i = \mathbf{y}_{i+1}$ -predict，对比 \mathbf{y}_{i+1} -predict和 \mathbf{y}_{i+1} -test便可以进行交易模型测试。

我们取 $\alpha = 0.03$ 、 $\lambda = 0.1$ ，对比测试结果如下表：

Table 5: 2011年模型表现结果

	收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
带时间衰减项	1.45×10^6	1.96	0.036	54.9	6.76	26.4
不带衰减项	1.49×10^6	2.06	0.040	51.1	7.28	24.7

Table 6: 2012年模型表现结果

	收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
带时间衰减项	1.57×10^6	2.20	0.04	54.8	8.35	24.8
不带衰减项	1.47×10^6	2.00	0.038	52.9	7.93	24.1

总的来说，两者的表现都相当不错。虽然在2011年带有时间衰减项的模型最终收益稍不如后者；但其最大回撤较低，这直接导致其风险收益比更高。其原因可能是前者对指数的趋势有更佳的把握。

不过，这种提升并不明显，而且使计算量偏大（每个周期均需要进行模型训练、求解）。在模型稳定性分析中，我们将对衰减参数 α 进行分析。

4 模型稳定性分析

在确认了模型具有较好的表现能力后，我们进行模型稳定性分析。稳定性分析的意义在于，若能确认模型参数对其表现具有稳定性，则可以说明交易模型的“生命周期”较长，需要维护的成本较低，面对市场的剧烈变化仍能保持收益的稳定。

分析的方法采用控制变量法，通过改变某一参数取值，研究其与模型表现间的关系。

4.1 正则化系数 λ 及时间衰减率 α

为了描述正则化系数 λ 与时间衰减率 α 对收益的影响，我们设定 $\lambda = 0.1, 0.3, 0.5, 0.7, 0.9$ 以及 $\alpha = 0.002, 0.003, 0.004, 0.005$ ，对其两两组合进行验证；我们关注的是年收益率（An-Return）、最大回撤（MaxDrawD）和风险收益率（RB-Ratio）

通过对参数的分析，尤其是可视化其与模型表现之间的关系。我们可以认为，该模型的最优参数选择是在一个较平滑的曲面上。故该模型并未出现过度优化的情况。

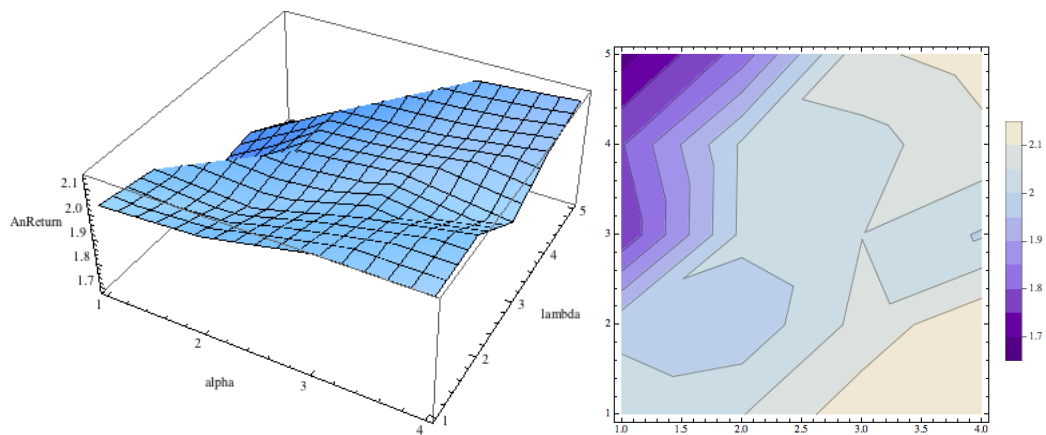


Figure 10: An-Return年收益率与 λ 和 α 的关系。在收益率最高的区域附近（右图右下角），且整体的值变化幅度也不大，所以该参数是稳定的

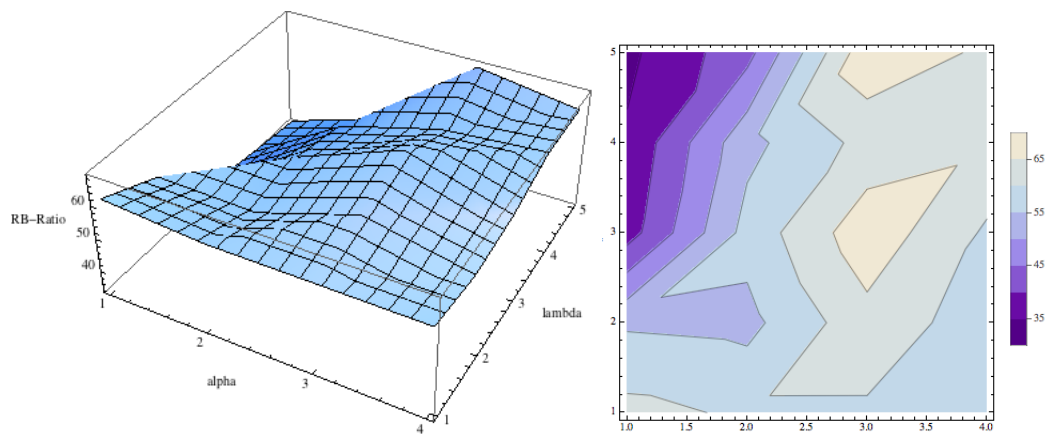


Figure 11: RB-Ratio风险收益率与 λ 和 α 的关系。在风险收益比最高的几个部分，曲面不很平滑。不过，对应着收益率较高的部分（右下角），该图像是比较平滑的，但总体来说，风险收益比的变化幅度较大。这很可能是与其计算方式有关（当最大回撤稍微变动，便影响该值）

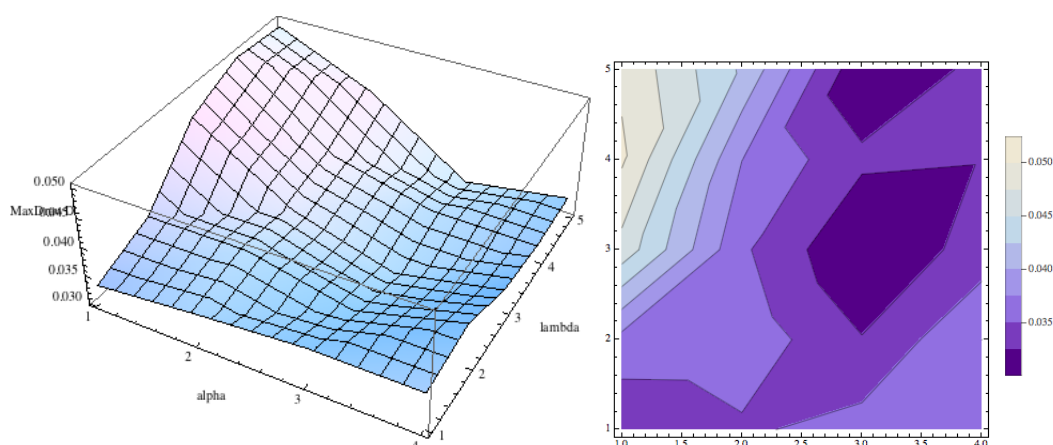


Figure 12: MaxDrawD最大回撤与 λ 和 α 的关系。同样的分析思路，可以看到模型最大回撤总体范围不大（0.03-0.05），同时在模型表现较好的区域，图像是比较平滑的。

4.2 周期通用性

关于模型周期通用性的分析，我们主要从两个方面入手。一是从模型本身，即输入变量的k线周期通用性；另一方面，是从交易的角度看，即交易级别的周期通用性。

4.2.1 输入变量的k线周期通用性

通过改变三条均线为 $MA5^{(i)}$, $MA10^{(i)}$, $MA20^{(i)}$ ，重新训练模型，并得到如下测试结果：

Table 7: 2011年新模型表现结果

收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
1.27×10^6	1.59	0.051	31.5	5.57	19.8

相比之下， $MA5^{(i)}$, $MA10^{(i)}$, $MA20^{(i)}$ 的均线组合对问题的描述能力不如 $MA3^{(i)}$, $MA5^{(i)}$, $MA7^{(i)}$ ，这有可能是因为参数调整不到位的缘故，因为此时的模型已经改变了。不过，由于模型表现并非相差很远，所以我们有理由相信，通过参数寻优，仍是能够得到稳定的最优表现。

4.2.2 交易级别的周期通用性

Table 8: 2011年新模型表现结果

收益	年化收益(倍)	最大回撤	风险收益比	夏普率	日均交易频率(次)
1.18×10^6	1.47	0.060	24.4	5.41	28.7

5 模型评价：缺点与不足

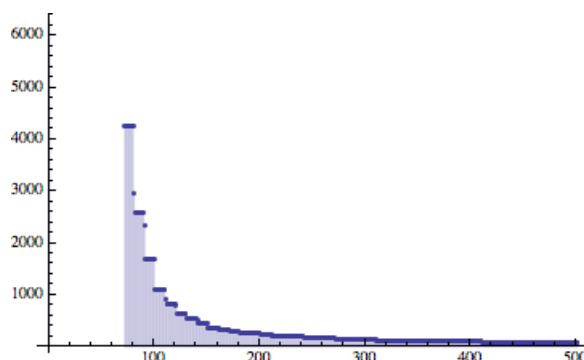
虽然该策略可以获得较高的收益，同时有较低的风险。但此模型仍有一些不足，需要得到认识。

- 线性假设的缺陷。关于均线组合与下一周期指数价格间的关系，我们使用简单、直观的线性表达，它虽然计算快速、简单，但这种假设的可信度仍然值得商榷。不过，从效果上看，这种趋势型模型的表现力和稳定性都是比较优秀的。

当然，我们可以仍然延续这种建模思路，将其中的关系用诸如神经网络等非线性算法进行表达。不过，我们利用线性模型的目的不在于“精确预测”，而是追求其中的“趋势”信息，这才是决定我们交易策略的关键。

- 交易信号可能出现闪烁，导致交易可能无法实现。由于我们通过“预测”得到“长期趋势”的一个预期，当下一时刻的指数出现时便可以做出判断，从而进行交易。但这个信号可能是不稳定的，即是说，虽然此刻满足买入（或卖出）的信号，但由于高频的特点，下一时刻的情况可能不再满足买入（或）卖出的条件。

所以，我们进行一些统计数据分析，以说明这种情况发生的可能性。我们导入股指IF2011.mat的tick级数据，统计tick之间指数变化的范围¹²：



其中指数在500ms内变化不大于6点（此即为我们预测的平均误差）的概率是99.9%，也就是说，即便我们错过了第一个买入时刻，在下一个时刻仍能够及时地发出交易信号。

- 时间衰减项带来计算复杂度的升高，同时模型表现提升不大。在未加入时间衰减项前，我们只需用一组数据进行训练，并在此后用该结果即可；但引入此项必须要周期性地更新。由上面的分析可以看到，这种添加方式并不一定能给模型性能带来显著提升，同时在“周期通用性”的检测上也表现不佳。所以，在实际使用时，可以考虑将该项去掉，或者对 α 做进一步参数寻优¹³。

¹²指数变化的统计步长为0.2，统计区间为[0, 10]，计算变化大于该值的数目

¹³由于计算资源有限，遗憾未能做更多验证

参考文献

- [1] 新浪博客, 为什么要进行程序化交易, http://blog.sina.com.cn/s/blog_54a980ed0100fw2k.html
- [2] Andrew Ng (Stanford), Machine Learning, <https://www.coursera.org/course/ml>
- [3] Stephen Boyd, Convex Optimization, Cambridge University Press. p351-352
- [4] Yin Zhu, Predicting User Activity Level in Social Networks, CIKM'13, Oct.27--Nov.1, 2013, San Francisco, CA, USA.
- [5] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. <http://cvxr.com/cvx>, September 2013.
- [6] Michael Grant and Stephen Boyd. Graph implementations for non-smooth convex programs, Recent Advances in Learning and Control (a tribute to M. Vidyasagar), V. Blondel, S. Boyd, and H. Kimura, editors, pages 95-110, Lecture Notes in Control and Information Sciences, Springer, 2008. http://stanford.edu/~boyd/graph_dcp.html.