

基于LASSO和神经网络的量化交易智能系统构建^{*}

——以沪深300股指期货为例

王宣承

摘要:本文立足于我国金融市场的现状提出了基于LASSO方法和神经网络模型的量化交易智能系统。该系统首先使用LASSO方法从众多技术指标中选出极少数最有效的指标作为输入变量,然后通过神经网络方法来搜索最优的交易规则,并以沪深300股指期货为例进行回测检验。结果显示:第一,与AIC和BIC回归模型相比,LASSO选出的变量少、预测高、且稳健性强;第二,经过神经网络的优化,交易系统的收益率和风险控制能力都得到了显著提高;第三,即使在考虑交易成本的前提下,该系统也可以获取超额收益。

关键词:LASSO;变量选择;神经网络;量化交易

JEL分类号:G21

一、引言

近几十年来,股票价格预测始终是金融研究领域的一个热点问题。总的来说,预测股价走势和买卖时点的方法包括基本面分析和技术分析。其中,基本面分析是通过分析宏观经济状况、行业发展趋势、公司同行业竞争水平和公司内部管理水平等诸多方面进行分析,以确定公司股票的内在价值;而技术分析则是通过股票的历史交易数据,如价格和交易量的变动等,来预测未来的价格走势。技术分析的假设是:历史会重复地发生,其分析工具包括金融统计模型和智能算法。

在技术分析早期发展历史中,统计模型曾发挥过主导性作用,如Box和Jenkins提出的自回归移动平均模型(ARMA)和Bollerslev(1986)提出的自回归条件异方差(GARCH)模型,均已成为刻画股票价格波动的经典方法;而用于评价波动风险的在险值(VaR)和条件在险值(CVaR)等指标,也被视为金融资产风险度量的行业标准。在股票市场上,各种技术指标更是数不胜数,其作用也各不相同,包括确定趋势(KDJ,

作者简介 王宣承:深圳市福田区发展研究中心研究员,上海财经大学统计与管理学院博士研究生;

***基金项目:**国家自然科学基金资助项目(71101083,71271128,71331006);上海市教育委员会科研创新项目(12ZZ072);上海财经大学创新团队支持计划。

MACD和Boll线),测量波动性(ROC)、强度(RSI)和市场人气(PSY和ADL)等。不同指标的组合使用可以构建特定的交易策略,为投资者带来超额收益。

上世纪80年代以后,智能算法凭借其强大的预测、学习和优化能力,开始迅速流行并得到广泛好评。智能算法可以发掘变量之间的非线性关系,并通过学习和训练,获得良好的动态预测能力。遗传算法(GP)、人工神经网络(ANNs)和支持向量机(SVM)等方法均被广泛用于在金融市场上进行投资和套利(Cao et.al, 2005; Chen et.al, 2010; El-Telbany, 2004)。然而,由于智能算法的优化过程主要基于随机与概率,对于运算过程存在模糊性和难以解释等缺陷。另外,在之前的研究中,技术指标通常被作为输入变量使用,他们被事先指定,数量极少且形式简单。由于没有对各种股票技术指标进行优化筛选,因此得出的结论很可能是有偏或非最优的。目前在统计学领域流行的一个变量选择模型——LASSO方法,可以对技术指标进行有效筛选,从而为交易规则的建立奠定坚实的基础。

本文将充分结合统计方法与人工智能技术的优势,构建一种量化交易智能系统。该系统首先通过LASSO方法从众多技术指标中选出极少数最有效的、解释力最强的指标,然后基于这些指标,使用人工神经网络方法来搜索最优的交易规则,并以沪深300股指期货为例进行实证检验。本文的结构如下:第二部分通过文献综述回顾基于技术指标的各种交易系统构建的特点与不足;第三部分介绍本研究中用到的各种变量选择模型,包括AIC、BIC和LASSO模型,以及用于预测优化的人工神经网络模型;第四部分使用本文提出的量化交易智能系统对沪深300股指期货进行预测和实证检验;第五部分是结论。

二、文献综述

Fama在1970年提出了著名的“有效市场假说”,即如果资产的价格完全反映了所有可以获得的信息,市场就是有效的;而如果当前的资产价格反映了所有过去历史的价格信息,市场就是弱有效的。按照“弱有效市场假说”,基于历史价格构建的技术指标分析是没有意义的,难以获得超额收益。

在早期的研究中,大部分结论支持了“市场有效性假说”。例如,Alexander(1961)检验了一系列交易规则,他建议投资者进行“追涨杀跌”的操作,当股票上涨一定比率时(如5%)便及时买入,当下跌一定比率时便卖出。尽管这种策略在道琼斯和标准普尔指数上获得了比“买入并持有”策略更高的超额收益,但在考虑交易成本因素后便不再有效。Fama和Blume(1966),以及Jensen和Benington(1970)在美国股票市场检验了滑动平均线和相对强度指标,也没有获得比“买入并持有”策略更高的超额收益。Jensen(1978)甚至声称,在经济学中没有其他理论能比“有效市场假说”得到更多支持了。

然而在最近20年,“有效市场假说”受到了越来越多的质疑和挑战。Brock等人(1992)对不同的移动平均线规则(MA)进行检验,在不考虑交易成本的情况下,可以捕捉到阶段性上升区间并据此获利。Bessembinder和Chan(1995)对6个亚洲国家的股票指数进行检验,结果验证了Brock结论的有效性,而且发现滞后于买卖信号一日交易将会降低获利的比率,而达到盈亏平衡点的交易成本为1.34~1.57%,远高于一般市场水平。Kung(2009)使用滑动平均线和交易范围突破(TRB)两个指标检验了台湾市场的有效性,结果发现在1975-1996年之间这两个策略是可盈利的,但在1997-2007年便不再有效。

智能算法的出现,为金融投资策略的构建提供了新的思路。神经网络、遗传算法等决策模型先后被用于股票市场,寻找或构建投资策略。White使用神经网络检验了IBM资产指数的价格变动中隐藏的变量间非线性关系;Kimoto(1990)使用模块化ANNs良好拟合了韩国股票市场指数TOPIX。Cao等人(2005)比较

了中国股票市场上FF三因素线性模型和神经网络的预测能力,证明后者的预测效果更好。Liao和Wang(2010)等人使用加入随机时间有效函数和布朗运动的神经网络模型,以交易价格和成交量为输入变量,分析了上证A股指数、B股指数以及4个国际股指的波动性,并对各指数进行了预测,拟合效果良好。戴念念和陈小伟使用小波神经网络拟合高阶CAMP模型,刘飞虎和罗晓光结合主成份分析和神经网络模型对财务风险进行预测。

Kaboudan(2000)对于遗传算法在股票市场的应用做出了开创性贡献,他证明了股票价格的可预测性,并据此构建了交易策略。然而,此后的研究更多地基于股票的变动趋势(而非价格)来寻找相关策略。Kaucic(2009)在大量宏观经济变量和公司金融指标中完成了变量选择和异常值诊断的工作,其构建的动态线性模型在MSCI欧洲股票指数上具有可观的预测效果。Chen等人(2008)将遗传网络规划(Genetic Network Programming)和21种股票技术指标(包括K线图和RSI、ROC等)相结合,成功找到可获利的交易规则并应用于日本股票市场,之后又将模型进行了深入优化和扩展(Chen et.al, 2010; Chen & Hirasawa, 2011)。

由此可见,之前绝大多数研究的关注点在于,检验某个特定的交易策略是否有解释和预测的能力,本文的研究角度与之略有不同。我们不对某一个具体的技术指标或交易策略进行检验,而在于通过变量选择和机器学习的方法,寻找一组特别重要的指标,并构建出一套可优化的交易策略。

三、模型和方法

(一)问题概述

为了在金融市场上获利,我们需要对市场未来的收益情况进行准确的预测。假设每日的平均价格表示为:

$$\bar{P}_t = \frac{C_t + H_t + L_t}{3} \quad (1)$$

其中 C_t, H_t, L_t 分别表示当日股票的收盘价、最高价和最低价。令 V_j 表示未来第 j 日的平均价格相对于当日收盘价的算术收益率,即:

$$V_j = \frac{\bar{P}_{t+j} - C_t}{C_t} \quad (2)$$

将未来 k 日的收益率加总,得到未来预期平均收益率 Y_t 。较高的 Y_t 值表示股票未来 k 日的平均价格会高于近日收盘价,因此具有投资价值,可买入或持有该资产;反之,较低的 Y_t 值表明股票近期有较大的下跌风险,应及时卖出或回避。

$$Y_t = \sum_{j=1}^k V_j = \sum_{j=1}^k \frac{\bar{P}_{t+j} - C_t}{C_t} \quad (3)$$

然后,我们将使用技术指标作为预测变量。技术指标是对股票价格各种属性或特征的数字化总结(Torgo, 2003)。为了便于执行和效果评估,还需要将模型的预测值 Y_t 转化为买卖信号。参照 Y_t 的上下4分位点选取买卖阈值,可以将 Y_t 分布到3个区间,分别对应买入、持有、卖出3种交易行为,即表示为:

$$signal = \begin{cases} sell & \text{if } Y_t < Q_1 \\ hold & \text{if } Q_1 \leq Y_t \leq Q_3 \\ buy & \text{if } Y_t > Q_3 \end{cases} \quad (4)$$

(二)变量选择模型

对于线性模型:

$$y_i = \mathbf{x}_i' \beta + \epsilon_i \quad i = 1, \dots, n, \quad (5)$$

其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ 为 p 维回归自变量, $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ 为回归系数, ϵ_i 为独立同分布且均值为零的随机误差项。在常用的最小二乘回归(OLS)模型中,为了使预测值和观测值的差距平方最小,需要求出在残差的平方和 $(\sum_{i=1}^n (y_i - \mathbf{x}_i' \beta)^2)$ 最小时回归系数 β 的估计值,即:

$$\hat{\beta}_{ols} = \operatorname{argmin}(\sum_{i=1}^n (y_i - \mathbf{x}_i' \beta)^2) \quad (6)$$

当自变量维数较高时,OLS回归通常使用最优子集法对显著变量进行筛选。最优子集法基于各种信息准则如AIC、BIC、 R^2 及Mallow's C_p 等,对自变量逐个进行删除或添加操作,最终获得简洁的模型。本文使用基于AIC和BIC准则的逐步回归法。

根据AIC准则,最优的模型应当与原模型拟合最好,这样的模型可以最小化:

$$AIC = -2\ln(g(\hat{(\theta)}_k | \mathbf{y})) + 2k \quad (7)$$

在正态线性模型中,上式又可写成:

$$AIC = n\ln(RSS_k) + 2k \quad (8)$$

其中 k 为自变量个数, RSS_k 为模型的残差平方和, $g(\hat{(\theta)}_k | \mathbf{y})$ 为模型的极大似然估计值。Akaike认为AIC达到最大值的变量子集为最优模型。BIC准则是基于Bayes方法构建的,它认为在正态线性模型中,最优模型应该是BIC达到最大,

$$BIC = n\ln(RSS_k) + k \times \ln(n) \quad (9)$$

Breiman(1995)曾批评到:最优子集法不仅计算量巨大,并且结果不稳健也不稀疏,会得到许多绝对值极小的参数,且本身无法克服多重共线性的问题。为此,Tibshirani(1996)提出了最小绝对收缩选择法(Least Absolute Shrinkage and Selection Operator,简称Lasso),Lasso本质上是一种带惩罚约束的最小二乘法,其参数估计值为:

$$\hat{\beta}_{lasso} = \arg \min(\sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda \sum_{j=1}^p |\beta_j|) \quad (10)$$

其中第二项为 l_1 惩罚项, $\lambda > 0$ 为协调参数,用于调节惩罚项的权重。Lasso方法通过将不显著系数缩减为0的方式,仅保留极少数显著的变量,并且极大地削弱了多重共线性的影响,非常适合对股票市场的重要技术指标进行筛选。

(三)预测模型

人工神经网络(ANNs)是一种人工智能方法,它具有强大的逼近非线性连续函数的学习能力,被成功应用于经济和金融预测领域(Pao, 2006; Cao et.al, 2005; 卢琦泽等, 2010)。相对于传统的预测方法,ANNs有两点优势:首先,它可以检测并提取出多个变量之间的非线性关系及交互作用;其次,它对输入变量的分布情况不做任何假定。Hornik等人(1989)证明,仅含一个隐层的前馈性神经网络,当在隐层配有logistic函

数时,可以渐进逼近任意非线性函数。同时 Homik 建议隐层节点数取值区间为 $n = [\sqrt{2n + m}, 2n + m]$, 其中 n 为输入层节点数, m 为输出层节点数。

人工神经网络由一系列相互连接的神经元构成,每个神经元执行两种计算,在输入端对各变量进行线性组合,在输出端给出非线性运算的结果,并连接到下一个神经元上。每个神经元的连接处都有一个权重,构建神经网络不仅需要构建网络的结构,还需要确定神经元之间的连接权重。反向传播神经网络(BP-ANNs)具有类似多层感知器的结构,训练样本从第一层(输入层)进入网络形成工作信号,再通过网络的正向传播到达最末层(输出层)形成误差信号,输入层和输出层之间会有一个或多个隐层。然后,误差信号从后向前反向传播,并在传播过程中逐层修改连接权重。BP算法(back-propagation algorithm)是一种常用的计算连接权重的优化算法,通过多次迭代拟合训练样本,以获得最小的预测标准误(史忠植, 2009)。具体步骤如下:

假设输入层信号表示为向量 $X = (x_1, x_2, \dots, x_m)'$, 隐层可表示为 $M = (m_1, m_2, \dots, m_h)'$, 输出层为 y 。通常在构建 ANNs 模型时,样本会分为两部分,一部分用来训练,另一部分用来检验。在训练阶段,隐层第 t 个节点的值是所有输入变量的加权平均值。

$$NET_t = \sum_{i=1}^m w_{ti}x_i + b_t \quad (11)$$

然后通过非线性 sigmoid 转换函数,每个隐层节点产生一个 0 到 1 之间输出值:

$$m_t = [1 + \exp(-NET_t)]^{-1} = f(NET_t) \quad (12)$$

各隐层节点的输出值和阈值 b_0 , 经过加权求和,传递到输出节点:

$$NET_0 = \sum_{i=1}^m v_t m_t + b_0 \quad (13)$$

在必要的情况下,输出节点可以再次进行 sigmoid 转换,得到输出层结果,由于本研究对期望收益率进行预测,取值范围超过 $(0, 1)$, 因此不必进行 sigmoid 转换,直接采用线性输出结果。

到目前为止,BP算法的第一阶段就结束了。第二阶段的内容在于调整各连接处的权重,这一过程可以通过最小化训练期的目标函数 SSE 来实现。

$$SSE = \sum_{j=1}^n (y_j - \hat{Y}_j)^2 \quad (14)$$

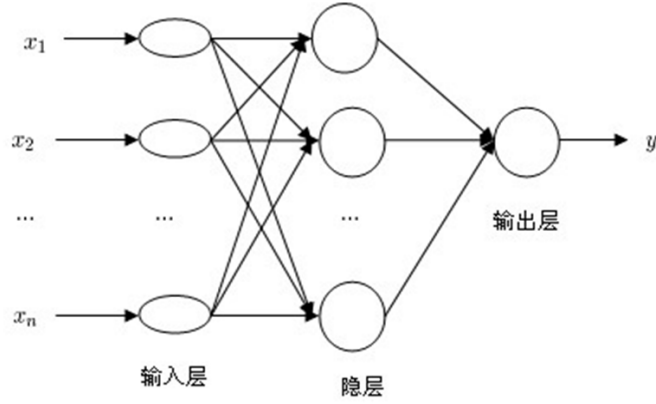


图1 神经网络结构图

其中 n 为训练期的样本个数。假设 Y 和 X 是单调的, 则输出变量的敏感性 S_i 可以表示为输出变量对输入变量的偏导数。

$$S_i = \frac{\partial \hat{Y}}{\partial X_i} = \sum_{t=1}^h \frac{\partial \hat{Y}}{\partial NET_0} \frac{\partial NET_0}{\partial m_t} \frac{\partial m_t}{\partial X_i} = \sum_{t=1}^h [f'(NET_0) v_t f'(NET_t) \omega_{ti}] \quad (15)$$

假设 $f'(NET_0)$ 和 $f'(NET_t)$ 为连续函数, 可以忽略, 则相对敏感性 \hat{S}_i 为:

$$\hat{S}_i = \sum_{t=1}^h v_t \omega_{ti} \quad (16)$$

具有较大的 \hat{S}_i 绝对值的输入变量, 对输出变量的影响更大。BP 算法的权重更新过程为:

$$\Delta \omega_{ij} = \eta \delta_j o_i \quad (17)$$

$$\begin{cases} \delta_j = f_j'(NET_j)(t_j - o_j) & \text{当 } j \text{ 为输出层节点} \\ \delta_j = f_j'(NET_j) \sum_k \delta_k \omega_{jk} & \text{当 } j \text{ 为隐层节点} \end{cases} \quad (18)$$

其中 $\Delta \omega_{ij}$ 为先前节点 i 到当前节点 j 之间权重的变化量, η 为学习因子 (恒量), δ_j 为节点 j 上实际输出与教学输入之间的误差, t_j 为教学输入, o_i 为先前节点 i 的输出。由于 BP 算法在每个训练期都要更新, 因此它也称为“在线学习算法”。

(四) 模型评价

在评价模型预测效果时, 最常见的评价指标包括平均绝对百分比误差 (RMSE)、平均绝对误差 (MAE) 和平均相对误差 (MAPE), 它们都是对某一变量预测值和真实值之间的差异进行测算的指标, 其数值越小表示预测越准确。他们可以用它们来检验未来预期平均收益率 Y_i 的预测有效性, 如下所示:

$$RMSE = \sqrt{\frac{\sum (Y_t - \hat{Y}_t)^2}{T}} \quad (19)$$

$$MAE = \sqrt{\frac{\sum |Y_t - \hat{Y}_t|}{T}} \quad (20)$$

$$MAPE = \frac{1}{T} \sum \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \quad (21)$$

然而,在金融市场预测中,仅对未来预期收益率进行有效预测可能是不够的,因为在交易时直接依据的是买卖点的决策。如果按照预期收益率转换得到的绝大多数决策信号是“持有”,那么对于投资者是没有吸引力的。因此,对于“买入”或“卖出”信号的准确预测是最关键的,他们直接影响着市场行为和潜在收益。

鉴于此,我们还将使用准确率(Precision)和检索率(Recall)对预测信号的有效性进行评估。准确率为预测正确的买入/卖出信号与所有买入/卖出信号预测值的比值。检索率为预测正确的买入/卖出信号与所有买入/卖出信号真实值的比值。通过混淆矩阵,可以很容易求出这两个比率,混淆矩阵为所有预测信号和真实信号的汇总矩阵。如下表所示:

表1 预测信号的混淆矩阵

		预测值			
		买入	持有	卖出	
真实值	买入	$n_{b,b}$	$n_{b,h}$	$n_{b,s}$	$N_{b,.}$
	持有	$n_{h,b}$	$n_{h,h}$	$n_{h,s}$	$N_{h,.}$
	卖出	$n_{s,b}$	$n_{s,h}$	$n_{s,s}$	$N_{s,.}$
		$N_{.,b}$	$N_{.,h}$	$N_{.,s}$	N

准确率和检索率可表示为:

$$Prec = \frac{n_{s,s} + n_{b,b}}{N_{.,s} + N_{.,b}} \quad (22)$$

$$Rec = \frac{n_{s,s} + n_{b,b}}{N_{s,.} + N_{b,.}} \quad (23)$$

除此之外,盈利水平是交易系统最有效和直接的测量指标,包括获利交易次数(比率)、平均收益率和相对于“买入并持有”策略的收益率等。

四、实证研究

(一)数据描述

本文的研究对象为沪深300股指期货(HS300)当月主力合约的日交易价格,具体包括每日的开盘、收盘、最高、最低价格和日成交量,数据来源是国泰安CSMAR系列研究数据库。数据的时间区间为2007年1月4日到2012年12月31日。图2为HS300股指期货在2012年内的价格走势,可以清楚的看到图中存在着多个拐点和趋势。

基于每日的开盘、收盘、最高、最低价格和日成交量、日成交额等数据,以2009-2011年的交易数据为样本内数据,可以计算出不同的股票指标,在本研究中共有49个指标与71个输出变量将被考察,如表2所示:

表2 各技术指标及其分类

指标类型	指标名称	种类数	输出指标个数
均线	EMA,SMA,WMA,EVWMA,ZLEMA,VWMA,MACD,TRIX	8	10
趋势	EMV,DEMA,ADX,AROON,CCI, TDI, VHF.close, VHF.hillow,priceDPO, ZigZag	10	17
震荡	RSI,ATR,chaikinVolatility,ROC, momentum, CMO,KST,MFI,WPR,volatility, SMI	10	16
成交量	chaikinAD,OBV,CLV,CMF,volumeDPO	5	5
动量	stoch,williamsAD	2	4
通道型	Bbands,DonchianChannel,Pbands	3	10
其他	SAR,过去10日收盘价之差	11	11
总计	——	49	71



图2 HS300股指期货的K线图及预期收益率 Y_t

(二)变量选择

为了消除量纲的影响,首先对自变量各技术指标及因变量未来预期收益率 Y_t 进行标准化处理,然后分别使用LASSO方法和基于AIC与BIC最小化准则的OLS模型,对71个自变量进行选择。

基于BIC和AIC准则,使用逐步变量选择法,分别得到16和42个重要技术指标,然而对于构建交易策略来讲,这个数量显然太多了。从统计学角度看,两种方法选出的变量中,最大方差膨胀因子分别是13765

和222373,这意味着非常严重的多重共线性,即相同的技术指标起着同样的解释作用。多重共线性对于参数估计影响极大,也不利于解释变量的作用效果。

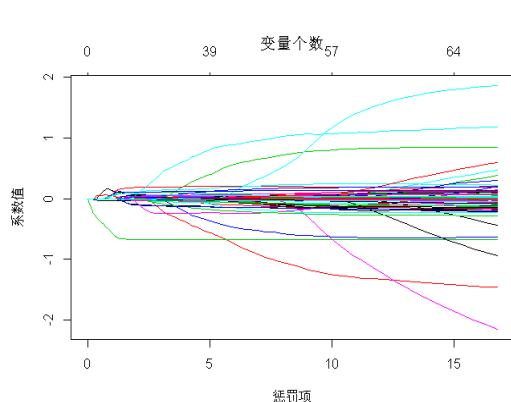


图3 LASSO解路径

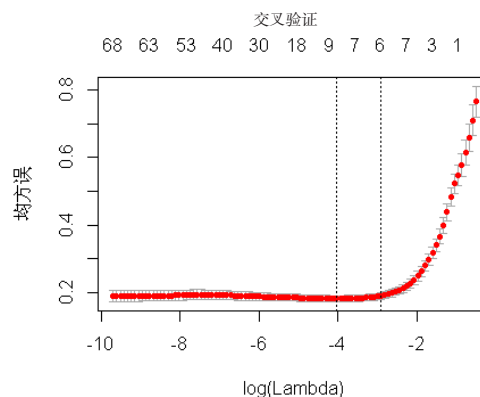


图4 交叉验证误差图

使用LASSO模型之后,由于加入了惩罚因子 l_1 ,不显著的变量被收缩为0,最终只选出6个技术指标。图3为各变量LASSO解系数的变化图,图4为交叉验证误差图,按照“1se准则”确定协调参数 $\lambda = 0.055$,6个变量足以将MSE的值降低到平缓的边界。其中4个变量的VIF值小于5,最大的VIF值为50.68,大大降低了模型的共线性。从预测效果上看,在样本内检验时,LASSO模型的RMSE和MAE均大于另两个模型,但MAPE最小,因此可以认为三者效果旗鼓相当;LASSO的准确率高而回忆率低,对于信号的检验更有效。使用2012年的数据进行样本外检验,其结果与样本内检验类似。图5为LASSO方法样本外检验的拟合效果图,可见预测值和真实值的波动基本上一致,拟合效果良好。

样本外检验

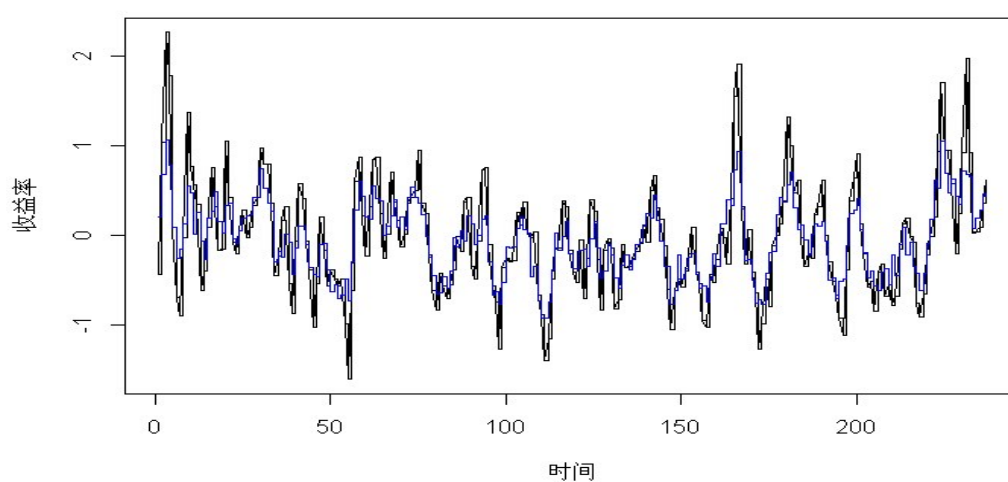


图5 LASSO方法的预测效果拟合图

表3 三种方法的变量选择结果

	AIC	BIC	LASSO
选 择 方 法	向前向后同时逐步筛选，以AIC为准则	向前向后同时逐步筛选，以BIC为准则	交叉验证次数10，以最小 λ 加上一个标准误差Se范围内的最大取值为准则。
共线性	最大VIF=13765， BIC=-1224.24	最大VIF=222373， AIC=-1224.24	最大VIF=50.68，有4个变量在5以下。
样本内预测			
RMSE	0.4004	0.3767	0.4358
MAE	0.305	0.2878	0.3333
MAPE	202.3812	281.8664	192.626
准确率	0.786	0.7908	0.8173
回忆率	0.7616	0.7702	0.6839
样本外预测			
RMSE	0.3208	0.3134	0.3354
MAE	0.2515	0.2462	0.2512
MAPE	242.4497	168.9552	178.4778
准确率	0.7564	0.7831	0.8518
回忆率	0.621	0.6565	0.4646
稳健性检验(滑动窗口)			
阶段1	ADX等共30个变量	MOM等18个	ROC,momentum,CLV,priceDPO,volumeDPO,Delt2,Delt3,Delt4。共8个变量
阶段2	ADX等共40个变量	MOM等12个	CLV,priceDPO,volumeDPO,Delt2,Delt3,Delt4。共6个变量
阶段3	ADX等共42个变量	ROC等共16个变量	ROC,momentum,CLV,priceDPO,Delt2,Delt3。共6个变量
涉 及 变 量 总 数	52	32	8

(三)稳健性检验

为了检验变量选择结果的稳健性，我们采用了滑动窗口(sliding window)的方式在不同的时间区间建模。整个过程分成3个阶段，在每个阶段中前3年为训练期，后续1年为检验期，验证完一个阶段后，时间窗口自动向前滑动1年。具体安排如下：

阶段1：训练期2007-2009，检验期2010；

阶段2：训练期2008-2010，检验期2011；

阶段3：训练期2009-2011，检验期2012。

结果如表3所示，AIC方法在每个阶段均选出30个以上的变量，3个阶段共选出52个变量；BIC在每个阶段选出的变量数在10-20个之间，3个阶段涉及的总变量数也达到32个，显然AIC和BIC方法在各阶段选出的变量存在着大量的轮换或不一致。LASSO方法在第一阶段选出了8个变量，后两个阶段只选出了6个

变量,且包含在之前的8个变量之中。这意味着,在为期6年的检验区间内,只需要使用8个变量即可完成对收益率的准确建模,因此LASSO方法的稳健性是极强的。

LASSO选出的8个变量或技术指标为:ROC, momentum, CLV, priceDPO, volumeDPO, Delt2, Delt3, Delt4。其中ROC全称为Rate of Change,即变动比率,用今天之价格除以其N天前之价格的比率。Momentum为动量指标,为当日收盘价与N日前收盘价之差,用以研究股价在波动过程中各种加速,减速和惯性的作用。CLV全称为Close Location Value,即收盘位置值,用来计算收盘价处于当天最高与最低价格变动范围中的位置,收盘位置值位于+1与-1之间。priceDPO,全称为price of De-Trended Price Oscillator,即去趋势价格振荡器。通过减去价格的滑动平均线,消除价格序列的趋势。volumeDPO,全称为volume of De-Trended Price Oscillator,即去趋势成交量振荡器。通过减去成交量的滑动平均线,消除成交量序列的趋势。Delt2, Delt3和Delt4分别是滞后2期、3期和4期的资产收盘价格。

(四)回测检验

鉴于OLS模型和LASSO模型均是线性模型,只能对变量之间的线性关系进行描述,无法挖掘其中的非线性关系。因此,选出若干重要指标之后,这些变量可以作为ANNs的输入变量,通过优化进一步提高系统的预测能力。ANNs对于变量的分布要求非常宽松,且善于拟合变量间的非线性关系。本次实验采用3层BP神经网络结构对未来收益率 Y_i 进行预测,参数设置如表4和表5所示;

表4 神经网络参数设置表

神经网络节点个数:
输入层:n,n为重要技术指标数量,由变量选择结果决定。
隐层:15,固定参数。
输出层:1,固定参数,即未来预期平均收益率 Y_i 。
隐层转换函数:非线性sigmoid函数
输出层转换函数:线性函数
学习进取率:0.01
最大迭代次数:1000
学习目标:0.0001

表5 交易系统参数设置表

交易策略类型:小资金(散户)策略、低频交易
初始资金:100万人民币
每次头寸仓位控制:100%
回测时间:2012年全年
止盈率:5% 止损率:5%
买入信号阈值:0.5 卖出信号阈值:-0.5
是否卖空:可卖空 交易成本:0

首先使用基于AIC与BIC的OLS模型和LASSO模型在训练样本上进行买卖信号的预测,并基于预测

的信号对2012年整年的日交易数据进行回测检验;然后将OLS模型和LASSO模型选择出来的16、42个和6个重要变量作为ANNs的输入变量,使用训练样本分别进行30次训练并选出表现最好的ANNs模型,从而构成了AIC-ANNs模型、BIC-ANNs模型和LASSO-ANNs模型。最后使用3种新模型对2012年的验证样本进行买卖信号预测和回测检验,在不考虑交易成本的情况下,结果如表6所示:

表6 三种变量选择模型在加入ANNs前后的回测检验结果

	AIC	BIC	LASSO	AIC+ANNs	BIC+ANNs	LASSO+ANNs
输入变量个数	42	16	6	42	16	6
交易次数	56	32	35	73	94	62
获利交易次数	43	27	29	50	76	53
获利交易比率%	76.79	84.38	82.86	83.33	80.85	89.47
最后获利金额	5311490	2326300	2587482	6491121	10935616	8595353
总收益%	531	233	259	649	1094	860
相对于“买入-持有” 策略的收益率%	524	223	249	642	1084	850
最大回撤表现	588640	351800	351802	940881	4332564	436143
夏普率	0.22	0.22	0.21	0.22	0.18	0.31
获利交易的策略 平均收益率%	5.14	5.08	5.09	5.23	5.13	5.21
损失交易的 平均亏损率%	-4.11	-4.99	-4.96	-2.85	-4.91	-4.24
所有交易的 平均收益率%	2.99	3.51	3.36	3.66	3.21	4.01
最大收益率%	6.43	6.43	6.43	6.43	6.22	6.43
最大损失率%	-5.24	-5.24	-5.24	-5.12	-5.45	-5.07

我们可以发现,在加入ANNs模型之前,AIC模型使用了多达42个技术指标,进行了56次交易,远高于BIC和LASSO模型,收益率达到531%,分别超过了BIC和LASSO各自收益率的2倍。但仅仅凭借这种表现就认为AIC是好的模型就过于片面了,过高的“最大回撤表现”和较低的交易成功率(76.79%)只能说明AIC是一个相对“激进”的投资策略。而且该策略需要同时优化42个指标的参数,在实践中显然过于复杂而缺乏可行性。BIC和LASSO在多个评价结果上相差不大,但是LASSO只用了BIC不到一半的变量个数。

在加入ANNs模型优化之后,3个模型的收益率都得到了显著提高,其中BIC-ANNs获得超过10倍的

收益率, LASSO-ANNs 以 8.6 倍的收益率居于次席, AIC-ANNs 收益率最少为 6.29 倍, 相对增幅也较小, 这可能与 AIC 选出的变量过多有关, 从而导致 ANNs 隐层节点相对较少且优化不充分。从交易次数上看 3 个模型也都有提高, 同样是 AIC 的增幅最少。BIC-ANNs 模型一年内总交易次数达到了 92, 接近 BIC 模型的 3 倍, 然而交易成功率却下降到 80.85%, 最大回撤表现达到惊人的 433256, 亏损交易的平均损失率接近止损点 (-4.91%), 还有高达 -5.45% 的最大损失率, 均说明该策略缺乏效率且隐含着巨大风险。LASSO-ANNs 模型的总交易次数只有 62 次, 是优化后的 3 个模型中最少的, 但其拥有接近 90% 的“获利交易比率”、最高的“夏普率”、“最大收益率”和“所有交易的平均收益率”, 同时拥有最小的“损失交易平均亏损率”、“最大损失率”和变量个数, 说明该模型不仅是有效的, 而且效率极高、风险较低。

由图 6 可以看到, 在测验期, 系统的获利是稳步提升的。在 HS300 指数从 2700 点跌至 2100 点时, 系统持仓基本都保持在负数, 即在做空; 而在末期 HS300 指数从 2100 点升至 2500 点时, 系统以做多为主。从整个趋势上看, LASSO-ANNs 模型准确把握了股指价格变化的趋势, 并获得了持续的正收益。



图6 LASSO-ANNs模型的获利状况

在交易成本方面, 本文主要考虑交易手续费和冲击成本。根据中国金融期货交易所的规定, 沪深 300 股指期货的手续费为合约规模价值的万分之零点二五, 而实际操作过程中, 期货公司收取的交易手续费一般不会超过万分之一。因此, 本文假设所有合约的交易手续费均为万分之一。冲击成本是指股指期货实盘交易时, 下单对市场成交价格造成的影响。理论上说, 冲击成本的大小和股指期货合约品种的流动性以及交易资金量大小有很大关系, 为了简化问题, 本文将冲击成本定为万分之二, 则总交易成本为万分之三。

考虑交易成本后, 各模型的回测检验结果如表 7 所示。可以看到, 各模型的相对表现与表 6 类似, 且各模型的最后获利金额均有所减少。其中 BIC+ANNs 和 LASSO+ANNs 两个模型减幅最大, 分别减少了

743%和364%。此时 LASSO+ANNs 模型不仅获利最多,且具有最高的获利交易比率、夏普率和所有交易的平均收益率,风险指标如最大回撤表现和最大损失率等也在可接受范围内,依然是表现最好的模型。

表 7 考虑交易成本后的各模型回测检验结果

	AIC	BIC	LASSO	AIC+ANNs	BIC+ANNs	LASSO+ANNs
输入变量个数	42	16	6	42	16	6
交易次数	54	50	32	62	52	61
获利交易次数	43	40	27	43	42	50
获利交易比率%	79.63	80.00	84.38	77.50	80.77	87.23
最后获利金额	4596200	2178348	2045258	4556559	3511270	4960141
总收益%	459	217	204	455	351	496
相对于“买入-持有” 策略的收益率%	456	214	201	452	348	493
最大回撤表现	574023	534098	352966	1354161	614186	587569
夏普率	0.21	0.22	0.20	0.20	0.18	0.26
获利交易的策略 平均收益率%	5.14	5.14	5.11	5.36	5.12	5.19
损失交易的 平均亏损率%	-4.86	-4.88	-4.99	-3.55	-4.40	-3.74
所有交易的 平均收益率%	3.10	3.14	3.53	3.07	3.29	3.89
最大收益率%	6.43	6.43	6.43	6.43	6.43	6.43
最大损失率%	-5.24	-5.24	-5.24	-5.02	-5.20	-5.07

在考虑交易成本后,LASSO-ANNs 模型的收益率随交易成本的增加而递减,当交易成本达到2.2%时,模型的收益率达到相对盈亏平衡点,即相对于“买入-持有”策略的盈亏平衡点(见图7)。该平衡点明显高于 Bessembinder 和 Chen(1995)的研究结论。这在一方面证实了 LASSO-ANNs 模型的有效性和盈利能力,另一个方面也反映出中国股票市场的非有效性,即投资者可以通过技术分析进行套利,获取超额收益。

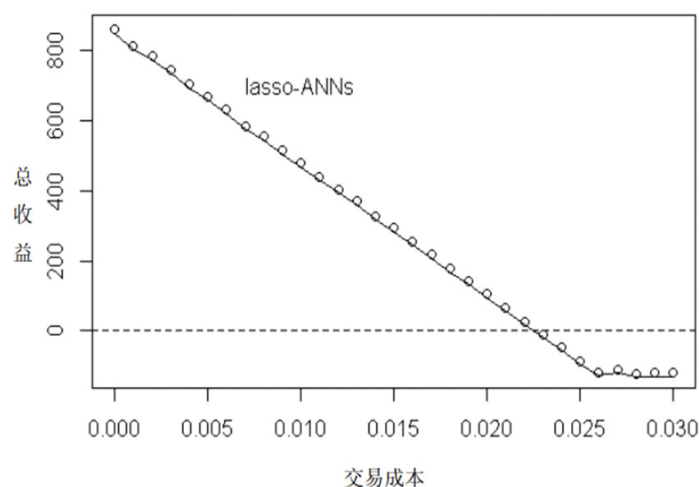


图7 不同交易成本下LASSO-ANNs模型的收益率

五、结论

本文使用基于LASSO和神经网络模型的量化交易智能系统对沪深300股指期货进行了回测检验,准确地预测了股指期货的价格变化趋势。在变量选择阶段,LASSO模型利用 l_1 惩罚项将不显著的系数收缩为零,从71个技术指标中挑选出6个最重要的指标,而模型的样本内和样本外的预测误差和准确率均不亚于由更多指标构建而成的AIC和BIC模型。在此基础上,由于神经网络模型具有较强的非线性拟合能力,能够从一定程度上弥补线性模型的不足,从而提升系统的预测能力,本文将LASSO选出的6个技术指标作为神经网络的输入变量,得到的LASSO-ANNs模型在收益和风险控制等方面均得到了极大的改进。通过3年滑动窗口方式的检验,LASSO变量选择模型具有较强的稳健性,即使考虑了交易成本因素,也能够获得比“买入并持有”策略更高的超额收益。

目前我国资本市场正处于金融创新的高峰期,随着“沪深300”股指期货的推出和融资融券试点的展开,我国的资本市场有了做空的品种和对冲系统风险的工具。由于投资者的交易理念、盈利模型发生了巨大的转变,“量化交易”的观念日渐深入人心,随之而来的是各种量化投资策略的引入以及金融投资产品在数量和规模上的爆发性增长。本文研究的LASSO-ANNs量化交易智能系统,一方面兼容了做空的机制,为构建股票或股指期货投资策略提供了一种新的思路。另一方面,基于Lasso的变量选择模型提供了最简单有效的技术指标变量,不仅在理论上避免了多重共线性、模型过度拟合等问题,在实践中也减少了时间和成本上的无谓消耗,使得交易策略的发掘和应用更加简便。最后经过神经网络模型的优化,交易系统的获利能力和风险控制得以保证。

在未来的研究中,变量选择的范围可以扩展到技术指标以外,借鉴基本面分析的思路,将宏观经济指标和公司财务指标纳入到股票价格预测的过程中。同时,LASSO-ANNs交易系统的参数也可以进一步优化,如买卖信号阈值的确定、止盈止损点的选取、以及在金融资产不能卖空的约束下系统的测验及改进等,这些工作的进展可以为人工智能技术在中国量化交易系统领域的推广和应用提供良好的适应性与广阔的发展空间。

参考文献

- [1]戴念念和陈小伟,2012,《基于小波神经网络的高阶 CAPM 实证研究》,《投资研究》,第12期,96-111。
- [2]刘飞虎和罗晓光,2013,《基于 PCA—RBF 神经网络的商业银行财务风险评价研究》,《投资研究》,第3期,88-97。
- [3]卢琦泽,叶德谦和南敏,2010,《基于遗传算法和神经网络的股票价格预测》,《电脑开发与应用》第23期,61-62。
- [4]史忠植,2009,《神经网络》,高等教育出版社,
- [5]Alexander, S. S, 1961, “Price movements in speculative markets: Trends or random walks”. *Industrial Management Review*, 2, pp.7-26.
- [6]Box, G. E., Jenkins, G. M., & Reinsel, G, 1977, “Time Series Analysis: Forecasting and Control”. *Technometrics*, 19(3), pp, 343-344.
- [7]Bollerslev, T, 1986, “Generalized autoregressive conditional heteroskedasticity”. *Journal of econometrics*, 31(3), pp.307-327.
- [8]Brock, W., J. Lakonishok, and B. LeBaron, 1992, “Simple technical trading rules and the stochastic properties of stock returns”, *Journal of Finance* 47, pp.1731-1764.
- [9]Bessembinder, H., & Chan, K, 1995, “The profitability of technical trading rules in the Asian stock markets”. *Pacific-Basin Finance Journal*, 3(2), pp.257-284.
- [10]Breiman, L, 1995, “Better subset regression using the nonnegative garrote”, *Technometrics*, 37(4), pp.373-384.
- [11]Cao, Q., Leggio, K. B., & Schniederjans, M. J, 2005, “A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market”. *Computers & Operations Research*, 32(10), pp.2499-2512.
- [12]Chen, Y., Mabu, S., & Hirasawa, K, 2010, “A model of portfolio optimization using time adapting genetic network programming”. *Computers & operations research*, 37(10), pp.1697-1707.
- [13]Cao, Q., Leggio, K. B., & Schniederjans, M. J, 2005, “A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market”. *Computers & Operations Research*, 32(10), pp.2499-2512.
- [14]Chen Y, Mabu S, Shimada K, et al, 2008, “Trading Rules on Stock Markets Using Genetic Network Programming with Sarsa Learning”. *Journal ref: Journal of Advanced Computational Intelligence and Intelligent Informatics*, 12(4): pp.383-392.
- [15]Chen Y, Hirasawa K, 2011, “A portfolio selection model using genetic relation algorithm and genetic network programming”. *IEEJ Transactions on Electrical and Electronic Engineering*, 6(5): pp.403-413.
- [16]El-Telbany, M. E, 2004, “The Egyptian stock market return prediction: A genetic programming approach”. *Electron Comput Eng*, 5, pp.161-164.
- [17]Fama, E. and M. Blume, 1966, “Filter rules and stock market trading”, *Journal of Business* 39, pp.226-241.
- [18]Hornik K, Stinchcombe M, White H, 1989, “Multilayer feedforward networks are universal approximations”, *Neural networks*, 2(5): 359-366.
- [19]Jensen, M. and G. Benington, 1970, “Random walks and technical theories: Some additional evidence”, *Journal of Finance* 25, pp.469-482.
- [20]Jensen, M. C, 1978, Some anomalous evidence regarding market efficiency. *Journal of financial economics*, 6(2), pp.95-101.
- [21]Kung J J. 2009, “Predictability of Technical Trading Rules: Evidence from the Taiwan Stock Market”. *Review of Applied Economics*, 5, pp.49-65
- [22]Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M, 1990, June, “Stock market prediction system with modular neural networks”. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, pp. 1-6.
- [23]Kaboudan, M. A., 2000, “Genetic programming prediction of stock prices”. *Computational Economics*, 16(3), pp.207-236.
- [24]Kaucic, M, 2009, “Predicting EU energy industry excess returns on EU market index via a constrained genetic algorithm”. *Computational Economics*, 34(2), pp.173-193.
- [25]Liao, Z., & Wang, J, 2010, “Forecasting model of global stock index by stochastic time effective neural network”. *Expert*

Systems with Applications, 37(1), pp.834-841.

- [26]Pao H T, 2006, "Comparing linear and nonlinear forecasts for Taiwan's electricity consumption", *Energy*, 31(12): 2129-2141.
- [27]Torgo L, 2003, *Data mining with R*. CRC Press.
- [28]Tibshirani R, 1996, "Regression shrinkage and selection via the LASSO". *Journal of the Royal Statistical Society. Series B (Methodological)*: 267-288.

Abstract: This paper proposes an intelligent system for quantitative trading based on LASSO method and artificial neural network model. The proposed system uses LASSO to select a few most effective variables from a large number of technical indicators. Then, the selected indicators will be treated as input variables of neural network model to search for the optimal trading rules, and CSI 300 Index will be used as an example for back-testing. The results show that: First, compared with AIC and BIC criteria of OLS regression models, LASSO method selected the least number of variables, while it shows comparable prediction accuracy and the strongest robustness; Second, since the trading system is optimized by neural network model, its profitability and risk control capabilities have been significantly improved; Finally, even considering the transaction costs, it can always outperform the "buy & hold" strategy and obtain excess returns.

Key Words: LASSO; Variable Selection; Neural Network; Quantitative Trading