

基于纵向数据的神经网络与线性模型的参数估计法

李旖宁 202006131039T

摘要

本文基于房地产板块 5 年的股票收盘价以及财务指标构成的纵向数据，建立传统线性模型以及神经网络模型，并将其相结合，并对股票数据进行参数估计，通过 R 软件的编程和计算，相比于传统线性模型，神经网络-最小二乘混合模型的预测效果更好。

关键词：神经网络；纵向数据；线性模型参数估计

1. 神经网络-最小二乘混合模型

神经网络模型是对自然的神经网络的模仿，它可以有效地解决复杂的有大量互相相关变量的回归和分类问题。神经网络的原理是把上层节点的值加权平均送到下层节点，最终到输出层节点，然后根据误差大小反馈回前面的层，再加权平均，如此反复的训练，直到误差在允许范围之内，具有一个隐藏层的一般神经网络的加权过程如下：

$$y_j = f^* \left(\sum_k w_{kj} Z_k + w_{0j} \right) = f^* \left\{ \sum_k w_{kj} \left[f \left(\sum_i w_{ik} Z_i + w_{0k} \right) \right] \right\}$$

其中 w_{ik} 是自变量 x_i 隐藏在第 k 个节点的权重， w_{kj} 是隐藏层第 k 个节点对于第 j 个因变量的权重， z_k 是隐藏层第 k 个节点的值， f^* 为激活函数。

而神经网络-最小二乘模型就是神经网络模型和最小二乘法混合而成的模型，先用神经网络模型对数据进行处理，再利用最小二乘法对参数进行估计。

2. 数据来源与采集时间

本文数据来源于西南证券金点子财富管理终端、中财网，采集时间为 2021 年 5 月 15 日，选取 18 支房地产股票数据，考虑 2016 年至 2020 年各个公司财务数据对年末股票收盘价的影响。共选取 5 个反映股票投资价值的财务指标进行分析，指标分别为每股收益、每股净资产、净利润、净资产、净资产收益率。

3.R 语言分析结果

首先我们建立神经网络模型，以股票收盘价作为因变量，五个财务指标为因变量建立神经网络模型，设置隐藏层为 1，我们可以画出神经网络图如图 1 所示。该神经网络模型的误差是 584.45，经历了 2712 次加权。

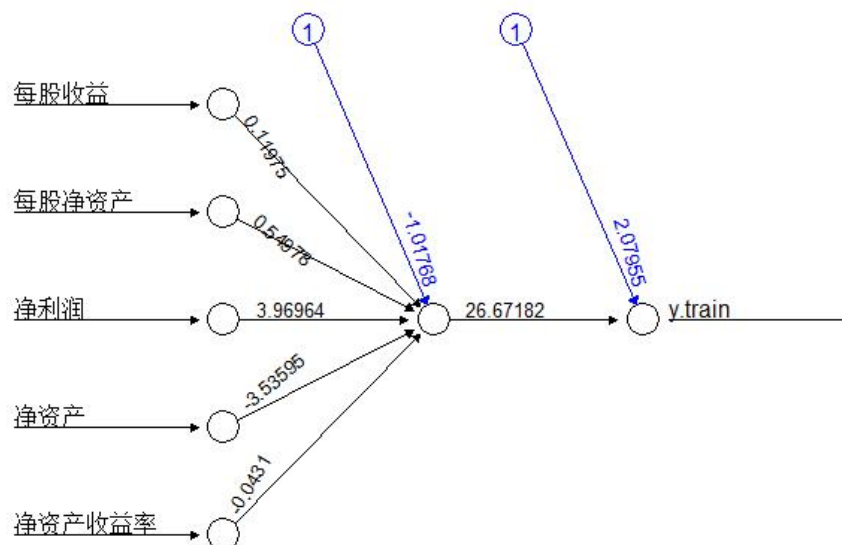


图 1 2016-2020 股票年末数据神经网络图

接着我们建立纵向数据的线性模型，通过 `crPlots` 函数判断自变量和因变量是否为线性关系，结果见图 2。

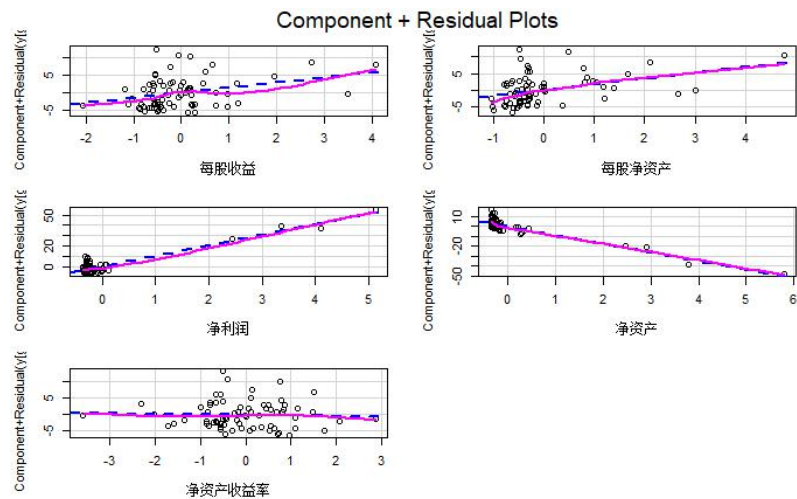


图 2 2016-2020 股票年末数据自变量与残差图

根据图 2 可以大致判断每股收益、每股净资产和净资产收益率与股票价格并非线性关系，由于股票价格是正测量数据，而且每股收益、每股净资产和净资产收益率与股票价格大致为指数关系，因此我们考虑连接函数为 $h(\mu) = \ln(\mu)$ ，分别建立在独立工作矩阵、等相关结构、AR(1) 下的边际模型。

接着我们建立纵向数据的其他线性模型，方法分别为最小二乘法、加权最小二乘法、约束极大似然估计、神经网络-最小二乘法，不同方法下的模型参数估计结果见表 1。

表 1 不同方法下的模型参数估计

变量选择方法	常数项	每股收益	每股净资产	净利润	净资产	净资产收益率
最小二乘法	9.2671689	1.5024149	1.7619076	10.299746	-8.6060025	-0.1489724
加权最小二乘法	9.2462961	0.9081362	0.4442924	8.180685	-5.177898	0.3897914
约束极大似然估计	9.2315738	0.8144424	0.1209712	7.8207795	-4.5407321	0.4805298
神经网络	2.05427789	0.12091761	0.54674219	3.94735072	-3.51529935	-0.04311215
神经网络-最小二乘法	2.080	0.197	0.5498	3.970	-3.536	-0.0431
Gee 估计-独立工作矩阵	0.2032	1.776	1.468	1.542	-3.244	-0.354
Gee 估计-等相关矩阵	-1.712	-0.0322	4.2805	3.381	-5.770	2.1385
Gee 估计-AR(1)	1.1042	2.7399	-0.0692	0.587	-1.867	-1.6487

表 1 显示，几种求解方法得到的参数估计结果有所差别，最小二乘法、加权最小二乘法、约束极大似然估计三种方法得到的参数估计得到的参数估计结果近似，神经网络系列方法和 Gee 系列方法得到的参数估计值整体比前三种方法更小了。Gee 估计-等相关矩阵的每股收益参数估计为负数，Gee 估计-AR(1) 的每股净资产参数估计为负数，加权最小二乘法、约束极大似然估计估计的净资产收益率为正数。模型效果评价见表 2。

表 2 不同参数估计法的股票数据模型效果

变量选择方法	内预测	外预测
最小二乘法	3.255	2.727
加权最小二乘法	3.306	2.931
约束极大似然估计	3.332	2.993
神经网络	7.051	5.241
神经网络-最小二乘法	2.001485*e-15	2.282*e-15
Gee 估计-独立工作矩阵	7.938	14.495
Gee 估计-等相关矩阵	20.149	22.130
Gee 估计-AR(1)	6.355	11.684

表 2 中预测效果最好的为神经网络-最小二乘法，误差接近 0。

4.结论与展望

神经网络-最小二乘法在股票的预测效果中有较为突出的表现，希望以后可以发掘研究神经网络与其他方法的结合。