

上市公司间关联网络的因子研发：利用 Usman Ali 与 David Hirshleifer 提出的动量溢出效应进行关联公司的收益预测

1. 引言

有效市场假说（Efficient Market Hypothesis）将一个价格完全反映所有可得信息的证券市场成为有效市场。然而，有效市场是一个绝对极端的概念，当投资者与分析师对于相关信息关注不足或者判断失误时，股票价格与其实际市值就会出现时效性的偏差。在 2019 年一篇关于动量溢出效应的文献中，Usman Ali 与 David Hirshleifer 正是利用了关联公司投资者与分析师造成的时效性偏差进行了跨公司收益的预测。通过对其关联股票的动量因子的计算与分析，Usman Ali 与 David Hirshleifer 将动量溢出效应总结为分析师的共同覆盖的成果。同时，由于一对关联公司被共同覆盖的分析师数量并像两个公司是否所处同一个行业是一个 0-1 的变量，共同覆盖的分析师数量相比所处同一个行业是一个 0-1 的变量可以更加精确地作为关联公司权重的衡量工具。

2. 研究过程

已有数据是由每一个公司名（EstPermID）及其对应的分析师（AnalystID）、分析师所在券商（BrokerID）以及分析师观测该公司的起止时间（StartDate、EndDate）构成。因此计算公司间的相关因子应当首先选取某一合适的时间段，在所选时间段下构建由公司与其分析师分布情况形成的矩阵，通过矩阵运算得到公司间相关因子。其中，本文选用 2014 年 10 月-2019 年 9 月的相关数据进行研究。

2.1 数据预处理

在取得数据进而进行数据分析数据前，所得数据中存在两个问题阻碍数据分析。首先是每一条数据所对应分析师观测的起止时间都精确到秒，因而很难选取到相同的时间进行分析；其次，结束时间（EndDate）从 285429 行开始显示空值。因此，先对数据中的时间列进行预处理是有必要的。

针对上述问题，首先我们将起止时间进行切片，只选出年月数据，呈现出“xxxx年-xx月”的形式。对于显示为空值的结束时间（EndDate），将值替换为样本结束时间——“2019年9月1日00:00:00”。将处理后的数据以 pkl 格式进行储存，以便后续选取使用。

2.2 数据分析

1.2 数据分析

所得数据中最关键的数据列是公司名（EstPermID）与分析师（AnalystID）。现将这两列数据单列出来，按照其固有顺序编成两个字典（cindex_to_id、aindex_to_id），以便后续对应查找，其格式为{ID1: 1; ID2:2; ...}。

其次，构建一个函数（task1），时期能够以单月为阶段，选取出在每个阶段内被关注的公司及其相关数据。再对选取出的公司进行分析，统计出在这个时间段关注过该公司的全部分析师，进而返回生成一个字典，其格式为{公司名:分析师名}。

接着，构建一个函数（task2），将 task1 函数返回的字典变成矩阵，横列为分析师名，纵列为公司名。对每一位分析师进行判断：在指定阶段内是否关注过该公司，若是，显示 1，若否，显示 0。最终形成一个由 0,1 构成的矩阵。再对该矩阵进行运算，求得其与自身转置矩阵的乘积。乘积所得结果矩阵中每个元素即为两公司在指定阶段所对应共同分析师的人数，以此作为关联因子。（例：第 i 行，第 j 列的元素即为第 i 个公司与第 j 个公司的关联因子，顺序与字典 cindex_to_id 中的顺序一致）。

最后，用 for 循环构建出从 2014 年 10 月到 2019 年 9 月，共 60 个月的时间段，以之为参数，运用之前定义的 task1 函数与 task2 函数，得出每个时间段下公司之间的关联因子。

3.结果分析

通过上述过程，得到公司与对应分析师的邻接表以及公司之间的关联因子矩阵。由于数据量过于庞大，现选取 2017 年 5 月的邻接表、前 100 个公司的关联因子矩阵展示如下。

2017_5	
30064771087	[30042006, 30160418, 30272528, 30316772, 30318604, 30280640, 30270392, 30306106, 30145475, 30307783, 30276948, 30045742, 30311739, 30295288, 30042327, 30053231, 30120488, 30155474, 30161056, 30171872, 30174275, 30186427, 30205460, 30209731, 30277738, 30280949, 30273980, 30319757, 30316897]
30064771095	[30300620]
30064771099	[30304524, 30195983, 30260412, 30203782, 30305725, 30293480, 30299861, 30293480, 30295472, 30311805]
30064771105	[30201200, 30137695, 30299862, 30293791, 30295471]
30064771108	[30199830, 30304120, 30183070, 30299093, 30321197, 30246256, 30310685]
30064771113	[30314608, 30313575]
30064873521	[30314608, 30313575]
30064771114	[30191900, 30314199]
30064771115	[30167754, 30296032, 30196365, 30279409, 30258806, 30258233, 30187920, 30207439, 30196365, 30309451, 30204605, 30274090, 30281335, 30316253, 30161034, 30293299, 30293327, 30165316, 30280155, 30303298, 30258542, 30319207, 30275958, 30308844, 30305825, 30263520, 30175415, 30319176, 30303349, 30273950, 30160699, 30278376, 30321603, 30315286, 30326981, 30321215, 30294662, 30319207, 30319161, 30251539, 30275926, 30304548, 30316306, 30201612, 30212158, 30321229, 30321230, 30321231, 30321232]
30064771117	[30280113, 30313030, 30277892, 30200292, 30314996, 30280461, 30299714]
30064771120	[30259154, 30187920, 30294625, 30302384, 30315816, 30272847, 30314997, 30315816, 30320129, 30324642, 30200082]
30064771122	[30299866, 30303034, 30274300, 30306821, 30298074, 30299862, 30322322, 30200675, 30204064, 30189350, 30199265]
30064771123	[30312912, 30306909, 30310313, 30301369]
30064771138	[30267490, 30300791, 30189745, 30294213, 30264763, 30166814, 30273914, 30303124, 30200232, 30315925, 30258867, 30326682, 30190325, 30323652, 30201115, 30337683, 30337683, 30199463, 30196813, 30303794, 30321477, 30172144]
30064771139	[30190594, 30236503, 30310496, 30187826]
30064771141	[30293300, 30272929, 30237131]
30064771142	[30295068, 30303763, 30303764, 30251713, 30208966, 30314996, 30191900, 30236343, 30196813]
30064771147	[30152360]
30064771148	[30305479, 30260033, 30260457, 30278376, 30300788, 30258235, 30172916, 30206197, 30303812, 30305593, 30294454, 30270285, 30175929, 30222334, 30301294, 30252960, 30295840, 30295675, 30295860, 30209822, 30258670, 30308948, 30302951, 30170068, 30301099, 30192927, 30311856, 30184867, 30187800, 30298599, 30312523, 30276028, 30260389, 30161032, 30305593, 30312620, 30317300, 30309507, 30296776, 30251820, 30210925, 30196639, 30303297, 30196639, 30302369, 30197791, 30303297, 30195932, 30281354, 30315762, 30312911, 30323205, 30258267, 30142813, 30319154, 30187023, 30303724, 30316069, 30205201, 30306585, 30320299, 30328326, 30330343, 30330028, 30321452, 30189127, 30321257, 30161032, 30280298, 30321258, 30323205, 30281612, 30319154, 30281376, 30315775, 30171585, 30204607, 30191071, 30189350, 30307142, 30314307, 30311837, 30305773, 30264133, 30304597, 30318279, 30315095]
30064771150	[30304999, 30304998, 30303837, 30258513, 30172908, 30304998, 30303837, 30304999, 30267804, 30196421, 30200070, 30237116, 30278670, 30299936, 30276458, 30293310, 30321258, 30299936, 30278670, 30272074, 30200070, 30264236, 30319913, 30304998, 30303837, 30205402, 30319319, 30189900, 30316272, 30172700, 30174155, 30278257, 30281554, 30306897, 30316220, 30314828, 30197506, 30273191, 30319319, 30204607, 30330051, 30328273, 30271356, 30281356, 30319300, 30314828]
30064771151	[30189451, 30187920, 30196365, 30305683, 30208589, 30314283]
30064771153	[30296833, 30251721, 30264133, 30311706, 30296833, 30251721, 30309530, 30314312]
30064771154	[30174288, 30280175, 30192482, 30187920, 30204854, 30305039, 30281475, 30293318, 30253147, 30189486, 30318330, 30173140, 30167958, 30275589, 30307172, 30195954, 30304527, 30311611, 30293096, 30321821, 30179099, 30174288, 30160699, 30236403, 30280175]
30064771155	[30258681, 30298617, 30279876, 30298617, 30279876, 30298617, 30275967, 30309970, 30298949, 30277692, 30278257, 30258345, 30174155, 30208154, 30275712, 30253156, 30320449, 30312912, 30281356, 30174155, 30314701, 30273191, 30281589, 30200070, 30237116, 30276458, 30278670, 30293310, 30299936, 30297214, 30212403, 30277692]
30064771156	[30306099, 30298172, 30296782, 30275716, 30251705, 30281106, 30297269, 30298055, 30178413, 30300902, 30187920, 30313523, 30294454, 30189350, 30195685, 30205285, 30272335, 30174220, 30159592, 30306099, 30209527, 30317944, 30189859, 30301369, 30322838, 30300996, 30313401, 30314855, 30306099, 30189891, 30325067, 30313401, 30312249, 30312032, 30329394, 30300996, 30315060, 30196813, 30302572, 30306088]
30064771157	[30207718, 30293289, 30281626, 30187920, 30199927, 30277083, 30204605, 30274090, 30304480, 30313683, 30314196, 30312909, 30204857, 30305350, 30293143, 30312910, 30318934, 30270046, 30174468, 30264660, 30270357, 30316142, 30318210, 30318799, 30318925, 30322794, 30325291, 30312120, 30322796, 30279000, 30319176, 30321231, 30321232, 30321230, 30321229, 30323868, 30267837, 30334323, 30191176, 30189126, 30314515, 30304548, 30299397]

图 1：2017 年 5 月公司与对应分析师的邻接表

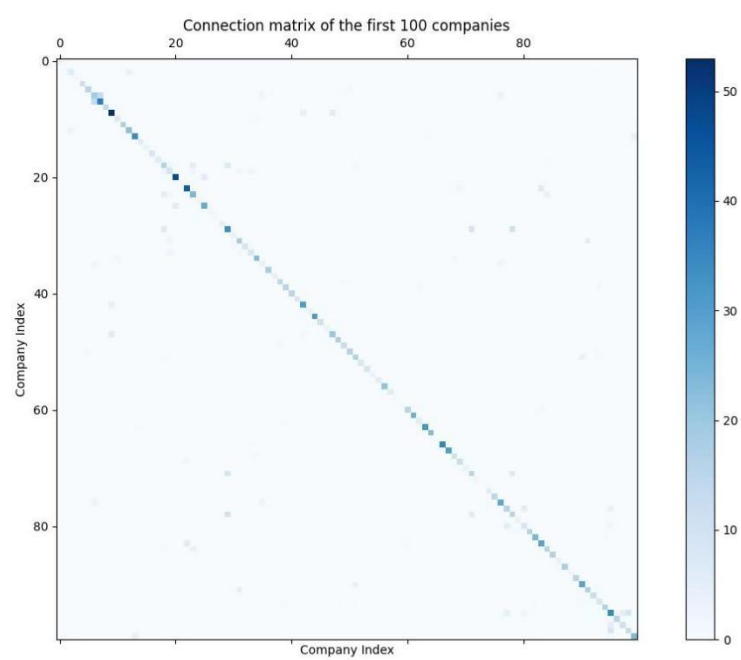


图 2: 前 100 个公司之间的关联因子矩阵

附录

1. 文件说明:

文件名	内容
connection_matrix(1).csv	前 100 个公司之间的关联因子矩阵