ISOM 673 -4101: Social Network Analytics

Empirical Assignment 5

<u>Diversification and Social Status in Venture Capital</u>

Student:

Feifan Gu

Under supervision of

Dr. Demetrius Lewis

Emory University

Goizueta Business School

CONTENTS

Chapter 1

Introduction

In this exercise, I analyzed how the co-investment networks of venture capital firms influence their strategies, in terms of the types of startup companies that they invest in. I analyzed the relationship between investors’ status and their diversification, the relationship between investors' status, diversification and the chance of successful investment, and the diversification strategy of these high-status investors.

I considered investors of the Venture Capital Type and deals that have occurred from 1990 onward in this whole report.

Chapter 2

Investor Firm Status and Investment Diversification

In this part, I studied the relationship between firms' status and their diversification of industry choices when investing. I used regression and plot to illustrate how and why investors with low and high statuses share "similar" less diversified strategy while middle-status investor's strategy is much more diversified.

Here status is represented as the eigenvector centrality of a 'leadership relationship matrix' between investors, and I used two ways to measure investors' diversification, a traditional one and a new one.

2.1 (Question 1A) relationship between status and traditional diversification

Here, I created a new variable called 'hhi2', which measures the cumulative concentration of each venture capital firm's portfolio through each year it has made an investment with the formula:

$$\text{Herfindahl} = \sum_{i}^{n} (deal\ size_i)^2$$

So the bigger the index, the bigger the concentration, the smaller the diversification.

Also, I computed and included 3 new control variables 'Whetherfirst', 'WhetherIT' and 'Whetherearly' in the regression to isolate the effect of status from the rationale of minimizing risk. With firm's age and year controlled and firm-fixed effects incorporated, the regression shows as below:

```
Oneway (individual) effect Within Model

Call:
plm(formula = hhi2 ~ lagstatus + I(lagstatus^2) + lagwhetherfirst +
    lagwhetherearly + lagwhetherIT + age + as.Date(as.character(Deal_Year),
    "%Y"), data = regression_uni1, effect = "individual",
    model = "within", index = "Investor_Id")

Unbalanced Panel: n = 5952, T = 1-27, N = 32306

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-9276.722  -765.789   -97.046   549.618  8177.800

Coefficients:
                                        Estimate Std. Error  t-value Pr(>|t|)
lagstatus                              -6785.540    295.854 -22.9354  < 2e-16 ***
I(lagstatus^2)                          8829.783    396.803  22.2523  < 2e-16 ***
lagwhetherfirstTRUE                     1064.335     37.300  28.5343  < 2e-16 ***
lagwhetherearlyTRUE                     -391.759     30.813 -12.7141  < 2e-16 ***
lagwhetherITTRUE                         -82.805     35.319  -2.3445  0.01906 *
age                                    25246.386  11692.762   2.1591  0.03085 *
as.Date(as.character(Deal_Year), "%Y")   -69.692     32.013  -2.1770  0.02949 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    9.2884e+10
Residual Sum of Squares: 5.9656e+10
R-Squared:      0.35774
Adj. R-Squared: 0.2125
F-statistic: 2096.49 on 7 and 26347 DF, p-value: < 2.22e-16
```

In this regression, the variables that we focused on, lagstatus and lagstatus^2 are all significant. Considering their coefficient and that diversification is opposite to concentration, it can be concluded that diversification is positively correlated with lag of status, and negatively correlated with the squared lag of status, using the traditional definition of diversification. Therefore, combining the two effects status having on diversification, it is not pure and may force diversification move in different direction when lagstatus is in different range. In specific range, linear term may weigh more than quadratic term while in another range, the opposite may happen. The relationship can be shown more clearly on a 2-D graph. Considering the coefficient of linear term is positive and the coefficient of quadratic term is negative, it should be a parabola opening downward. I will show the plot when using a new definition of diversification in Question1C, which is consistent with the inference here.

2.2 (Question 1B) relationship between status and new diversification

Here, we use a new diversification taking how different industry categories might be related into account. We not only consider different categories, but also measure how different they are using a network between industries. The new measure of diversification is called ‘niche width’:

$$\text{niche width} = 1 - \frac{1}{1 + \frac{\sum_{it} distances}{total\ number\ of\ industries - 1}}$$

Here distance refers to the Jaccard distance between each pair of industry categories for each year. Since sum of distances is in the denominator and there is a negative sign before the fraction, niche width should positively measure the diversification.Then I reran the regression using glm and used the ‘Mundlak’ approach to incorporate fixed effects:

```
Call:
glm(formula = nichewidth ~ lagstatus + I(lagstatus^2) + lagwhetherfirst +
    lagwhetherearly + lagwhetherIT + as.Date(as.character(Deal_Year),
    "%Y") + age + mundlak, family = quasibinomial(link = "logit"),
    data = regression_uni2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9667  -0.1257   0.0347   0.2000   0.8800

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                             -8.274e-01  4.449e-02 -18.599   <2e-16 ***
lagstatus                                7.519e+00  1.076e-01  69.876   <2e-16 ***
I(lagstatus^2)                          -7.663e+00  1.478e-01 -51.843   <2e-16 ***
lagwhetherfirstTRUE                     -2.499e-01  1.133e-02 -22.060   <2e-16 ***
lagwhetherearlyTRUE                      1.988e-01  8.986e-03  22.123   <2e-16 ***
lagwhetherITTRUE                         4.457e-03  8.176e-03   0.545    0.586
as.Date(as.character(Deal_Year), "%Y")   4.708e-05  2.717e-06  17.327   <2e-16 ***
age                                      8.287e-02  1.525e-03  54.339   <2e-16 ***
mundlak                                  8.557e-02  8.580e-03   9.973   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 0.09533599)

    Null deviance: 6627.5  on 32305  degrees of freedom
Residual deviance: 3452.2  on 32297  degrees of freedom
  (6281 observations deleted due to missingness)
AIC: NA

Number of Fisher Scoring iterations: 5
```

As we can see, lagged status and lagged status squared are both significant. The coefficient of lagged status is positive and the coefficient of lagged status squared is negative, and they share the same order of magnitude (e+00), even have similar absolute value. This indicates a similar relationship between status and diversification as the conclusion in Question1A. If we drew a 2D-graph between them, it should be a parabola opening downward, which means when lagged status weighs more than lagged status squared, diversification moves in the opposite direction as lagged status; when lagged status weighs less, diversification moves in the same direction as lagged status.

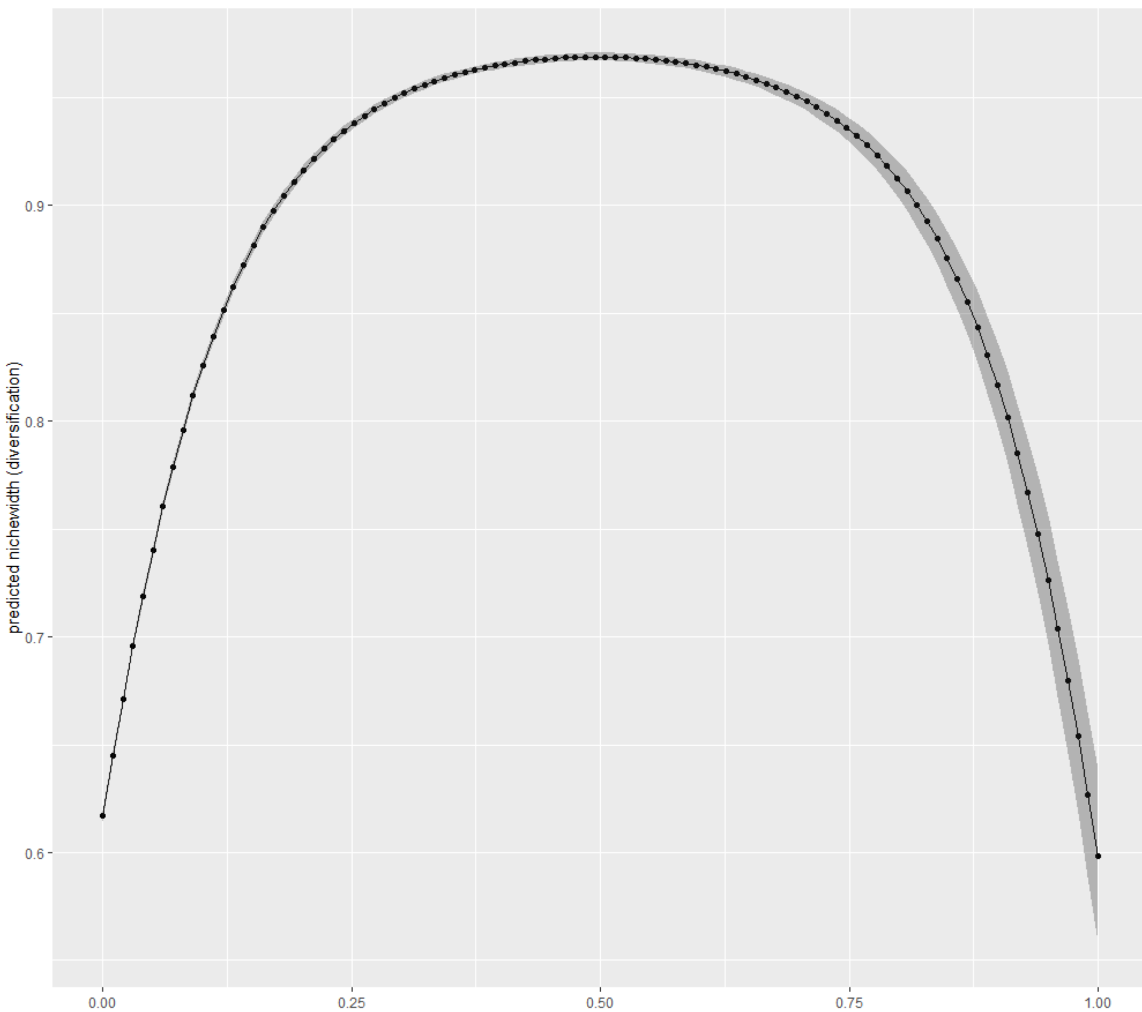
We will further justify such conclusion in the following part.

2.3 (Question 1C) Diversification strategies of investors with different status

Only using the lagged status and lagged status squared in the regression in 1B, I reran the regression:

```
Call:
glm(formula = nichewidth ~ lagstatus + I(lagstatus^2), family = quasibinomial(link = "logit"),
    data = regression_uni2)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.32026  -0.17493   0.04711   0.25060   0.91972

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.476805   0.005325   89.55   <2e-16 ***
lagstatus       11.908981   0.119726   99.47   <2e-16 ***
I(lagstatus^2) -11.987717   0.162242  -73.89   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.1333444)

    Null deviance: 6627.5  on 32305  degrees of freedom
Residual deviance: 4645.5  on 32303  degrees of freedom
  (6281 observations deleted due to missingness)
AIC: NA

Number of Fisher Scoring iterations: 5
```

The effect is basically the same. Then I used the model to predict 100 points between minimum and maximum of lagged status, which is 0 and 1, and plotted the fitted value and the 95% confidence intervals:



This curve is a parabola-like curve with its open end pointing down, which clearly shows the different diversification strategies of low, middle, and high-status venture capital firms. Low and high-status investors share similar less diversified strategy, investing in fewer industries or industries that are closely related to each other, like sharing common domain knowledge; middle-status investors tend to have the most diversified strategy, investing much more industries which even are not closely related to each other.

2.4 Conclusion

This phenomenon makes sense in real world. An investor with low status, which means having less chance to lead investments, is more likely to be a young and small-sized firm. It's domain knowledge and experience can only support it to invest in several quite close-related industries or even one industry; An investor with middle status, which tends to be in the fast developing stage, will more likely try different industries to explore which they can perform well in, possibly in a trial-and -error procedure; An investor with high status, possibly the leader of the investment industry, more likely has found the investment industry that it is best at, so it will focus more in this area which could yield more guaranteed return for the firm.

Chapter 3

Investor Firm Status, Diversification and Effects

In this chapter, I studied on what kind of investors are more effective at diversifying their portfolios in terms of their status and extend of diversification. Here we define successful investments as those having a deal type of "IPO", "Merge/Acquisition" or "Buyout/LBO" and use the niche width measure of diversification.

3.1 (Question2A) relationship between the interaction of lagged status and lagged diversification and number of successful investments

Here I should include the same control as in 1A and 1B, which indicates using pglm regression to combine plm and glm. The regression of lagged status and lagged diversification to number of successful investments ran like this:

```
Maximum Likelihood estimation
Newton-Raphson maximisation, 11 iterations
Return code 2: successive function values within tolerance limit
Log-Likelihood: -7266.844
11  free parameters
Estimates:
                                    Estimate Std. error t value  Pr(> t)    
(Intercept)                       -1.972e+00  7.863e-01  -2.508  0.01213 *  
lagstatus                         -5.390e+00  2.386e+00  -2.259  0.02389 *  
lagnichewidth                      2.323e+00  2.296e-01  10.121  < 2e-16 ***
lagwhetherfirstTRUE                3.990e-01  8.716e-02   4.577 4.71e-06 ***
lagwhetherearlyTRUE               -4.467e-01  4.262e-02 -10.482  < 2e-16 ***
lagwhetherITTRUE                  -3.102e-02  6.390e-02  -0.485  0.62736    
age                                2.269e-01  1.747e-02  12.987  < 2e-16 ***
as.Date(as.character(Deal_Year), "%Y") -2.230e-04  4.768e-05  -4.677 2.92e-06 ***
mundlak                           -2.157e-01  1.313e-01  -1.644  0.10026    
lagstatus:lagnichewidth            6.740e+00  2.493e+00   2.703  0.00687 ** 
sigma                              7.001e-02  4.464e-03  15.682  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this regression, lagged status, lagged niche width and their interaction are all significant. What we focus most here is the interaction term, whose coefficient is 6.740, positive, indicating that a higher level of both status and diversification together have a positive effect on the number of successful investments.

3.2 (Question2B) visualization of the relationship

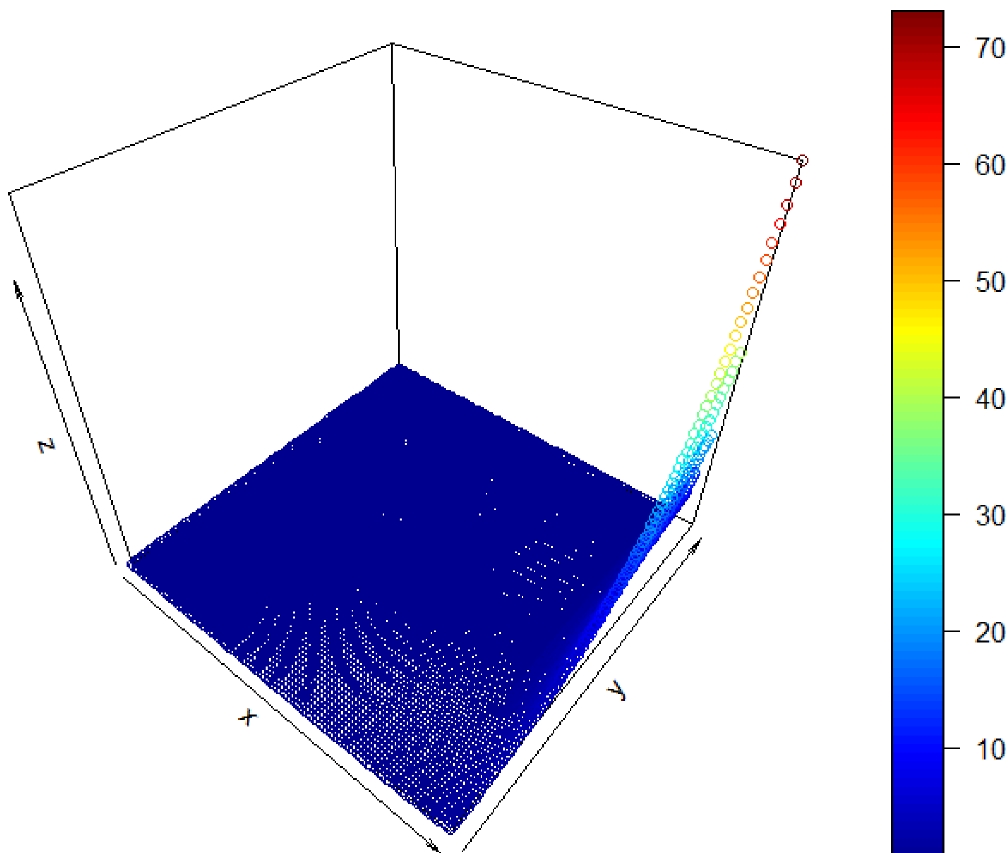
First, I reran the model with only lagged status and lagged diversification (niche width) involved using glm:

```
Call:
glm(formula = numofsuccess ~ lagstatus + lagnichewidth + lagstatus:lagnichewidth,
    family = "poisson", data = regression_uni3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8240  -0.6043  -0.3059  -0.1171  14.0832

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              -9.3507     0.1541  -60.67   <2e-16 ***
lagstatus               -44.9284     1.7892  -25.11   <2e-16 ***
lagnichewidth             9.5360     0.1787   53.37   <2e-16 ***
lagstatus:lagnichewidth  49.0334     1.8449   26.58   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 40552  on 32305  degrees of freedom
Residual deviance: 25167  on 32302  degrees of freedom
  (6281 observations deleted due to missingness)
AIC: 33069

Number of Fisher Scoring iterations: 8
```

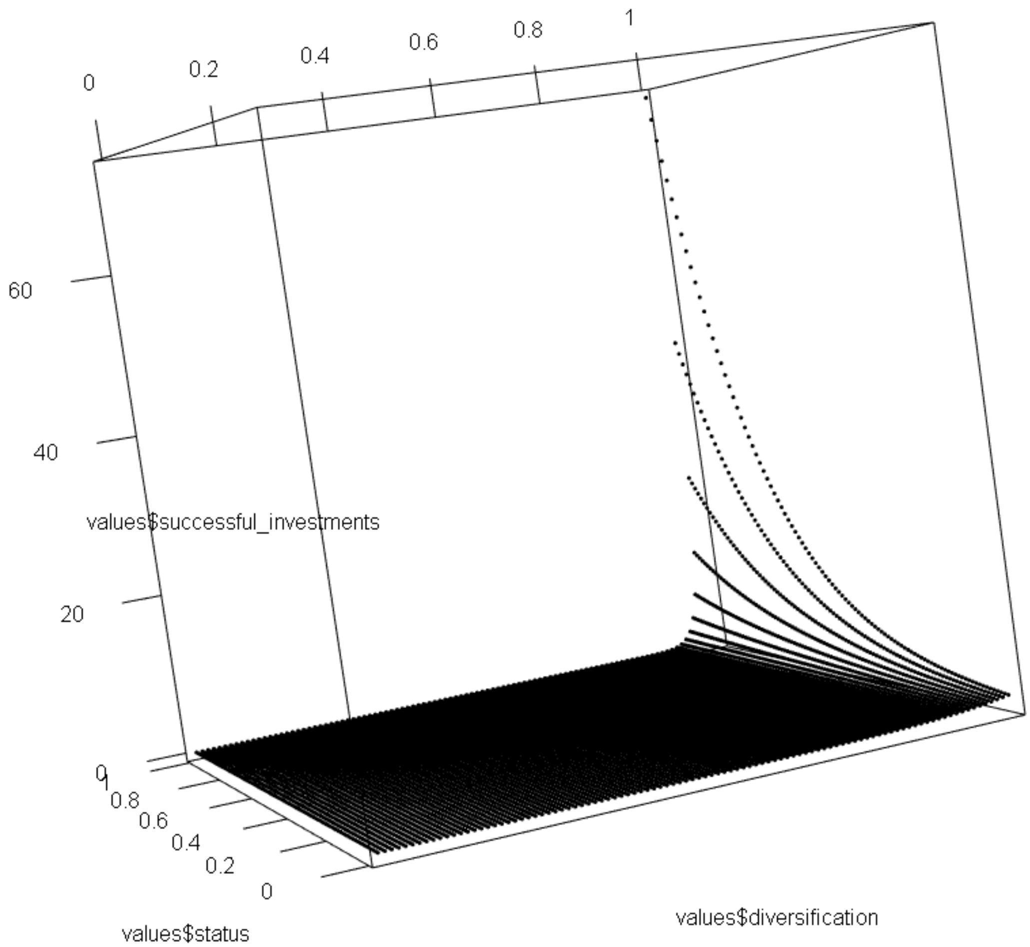
Still, they are all significant and the interaction term has a positive coefficient.

Then I predicted the number of successful investments on 10,000 points in the range of lagged status and lagged niche width and used three ways to plot the relationship between status, diversification and number of successful investments.

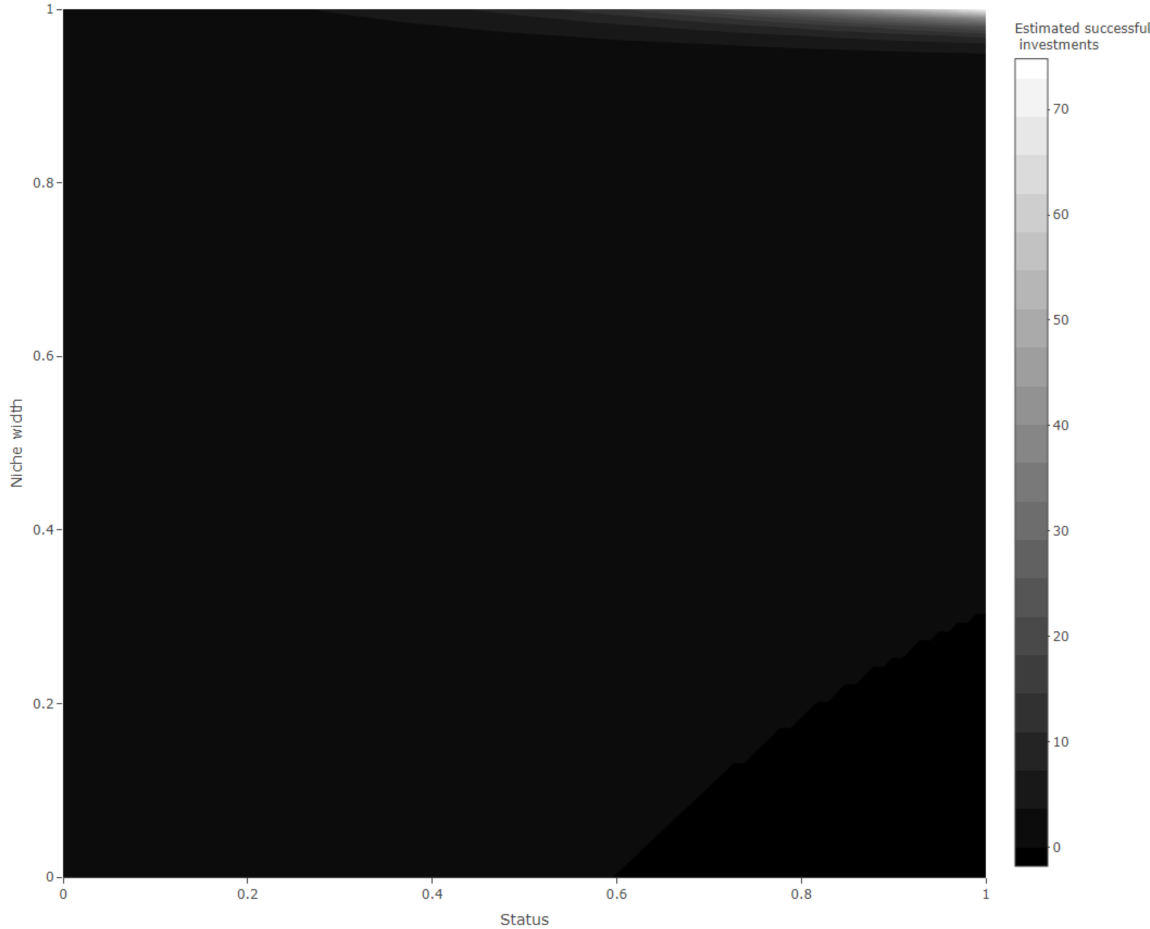
Regular 3d plot:



Interactive 3d plot:



Contour plot:



The patterns of the three plots are consistent. We can easily tell that those people with both high status and high diversification are most likely to be successful at diversifying their portfolios; Those who have low status and low diversification are least likely to succeed in diversifying.

3.3 Conclusion

As you can see, single status or single diversification variable influences slightly on the chance of successful investments. The interaction of two of them both being high could generate something magic, resulting in a suddenly extremely high chance of success. This could be explained like this: high-status investors could make the best of diversification, since they can be somehow good at dealing with a huge amount of diversified industries. How does the high status give them the magical power? We will figure their secrete out in next chapter.

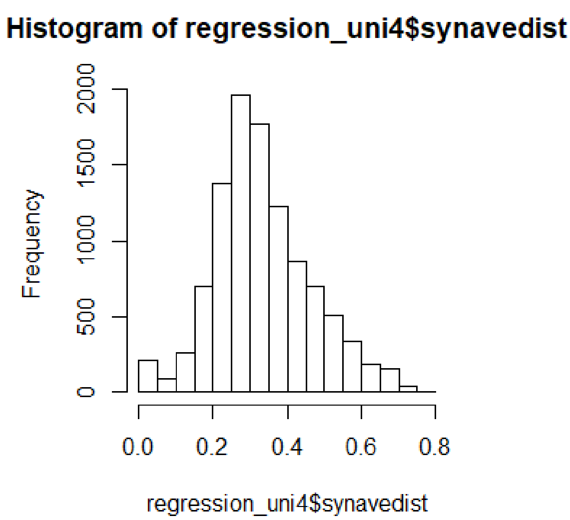
Chapter 4

Diversification Strategies of High-status Investors

Chapter 2 (Question1) suggests that low and high-status investors share the similar tendencies to diversity, while Chapter 3 (Question2) suggests that only high-status investors are better at diversifying. In this chapter, I studied on the secret of high-status investors, examining the strategies that make them better.

4.1 regression: how high-status investors coordinate other investors' expertise

Here I ran a pglm between lagged status, an investor's own average distance from the industry medoid and the investor's syndicate partners' average distance using family=Gaussian, since the distribution of y is more like a Gaussian distribution.



We hope to find that the coefficient of 'lagged status: investor's own average distance' is negative, so that we can conclude that high-status investors tend to coordinate syndicate partners who are good at the industries that they themselves are not that good at.

However, the regression is like this:

```
Maximum Likelihood estimation
Newton-Raphson maximisation, 6 iterations
Return code 2: successive function values within tolerance limit
Log-Likelihood: 7618.818
10  free parameters
Estimates:
                      Estimate Std. error t value Pr(> t)
(Intercept)          0.2531884  0.0043915  57.655  <2e-16 ***
lagstatus           -0.3202065  0.0184626 -17.344  <2e-16 ***
ownavedist           0.2359974  0.0081451  28.974  <2e-16 ***
lagwhetherfirstTRUE  0.0010840  0.0046646   0.232  0.8162
lagwhetherearlyTRUE -0.0084392  0.0028336  -2.978  0.0029 **
lagwhetherITTRUE     0.0011463  0.0027193   0.422  0.6734
age                  0.0002655  0.0002493   1.065  0.2869
lagstatus:ownavedist 1.0062481  0.0450238  22.349  <2e-16 ***
sd.mu                0.0397667  0.0026766  14.857  <2e-16 ***
sd.eps               0.1047646  0.0009343 112.133  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Though all the variables that I focused are significant, unfortunately, the coefficient of ‘lagstatus:ownavedist’ is positive, opposite to what the assignment requirements hint. Sorry that I do not have enough time to figure out which step might yield the wrong coefficient. I am pretty sure that I made most of it correct, calculating coordinates using MDS, generating medoid coordinates, calculating distances, etcs., or else it would not run successfully and printed out a result.

The coefficient should be negative, so that I can conclude that high-status investors tend to use their influence in the investor network to coordinate other investors’ expertise on deals that are further away from their own expertise.

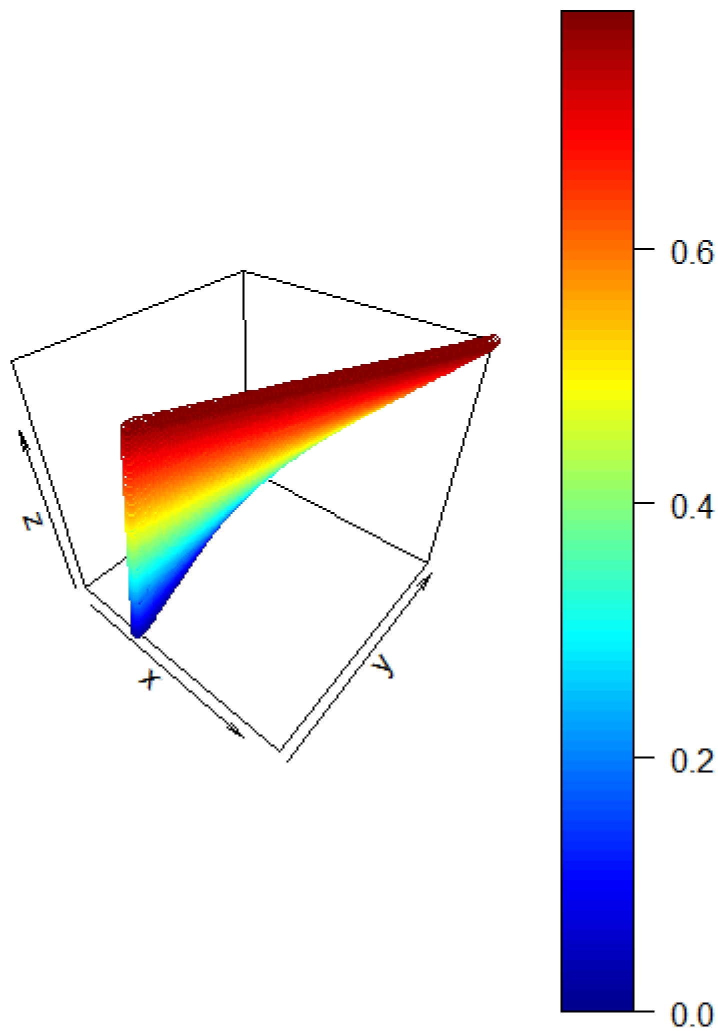
4.2 visualization of the relationship in 4.1

We ran a glm without those controls:

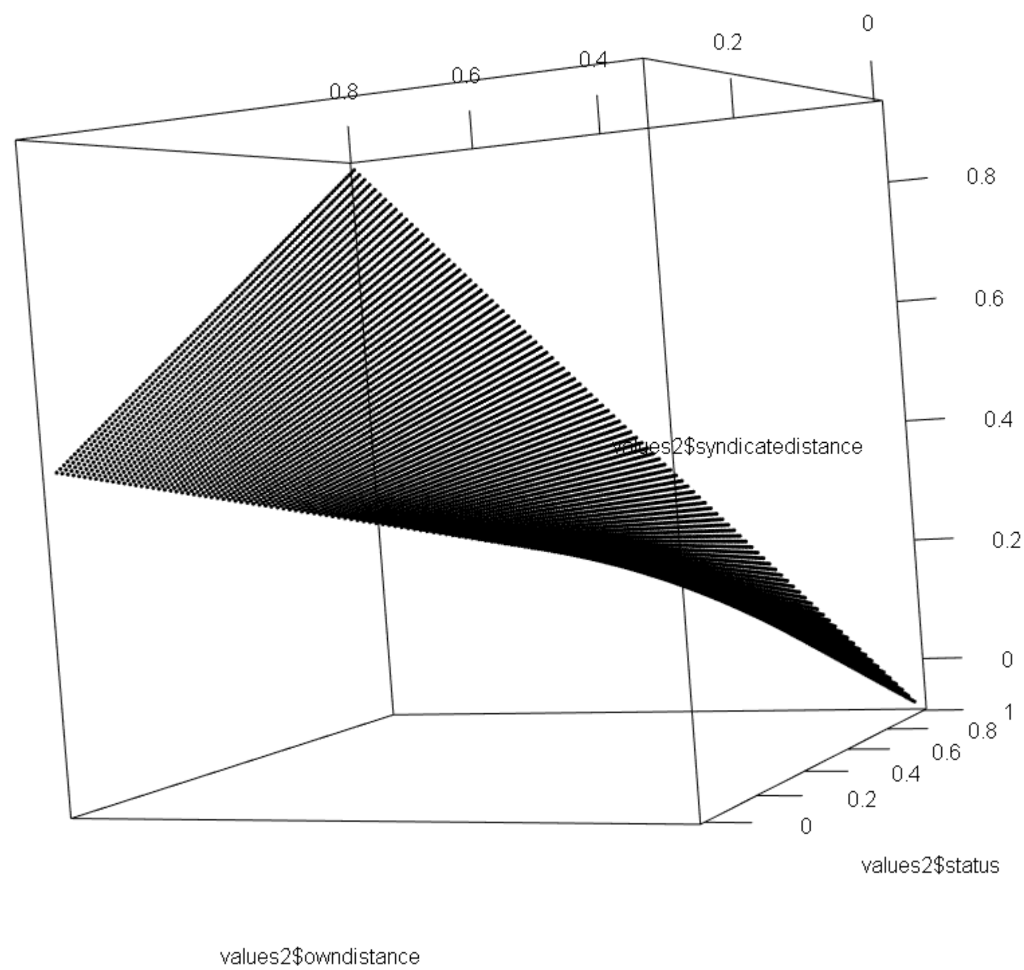
```
Call:
glm(formula = synavedist ~ lagstatus + ownavedist + lagstatus:ownavedist,
    family = gaussian, data = regression_uni4)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.43108  -0.06004  -0.00827   0.05459   0.47266

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           0.254066   0.003070   82.76   <2e-16 ***
lagstatus            -0.324274   0.017113  -18.95   <2e-16 ***
ownavedist            0.229543   0.008083   28.40   <2e-16 ***
lagstatus:ownavedist  1.026184   0.046281   22.17   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.01234563)

    Null deviance: 156.13  on 9711  degrees of freedom
Residual deviance: 119.85  on 9708  degrees of freedom
  (661 observations deleted due to missingness)
AIC: -15111

Number of Fisher Scoring iterations: 2
```

Then still three ways to plot using the prediction of the model above:

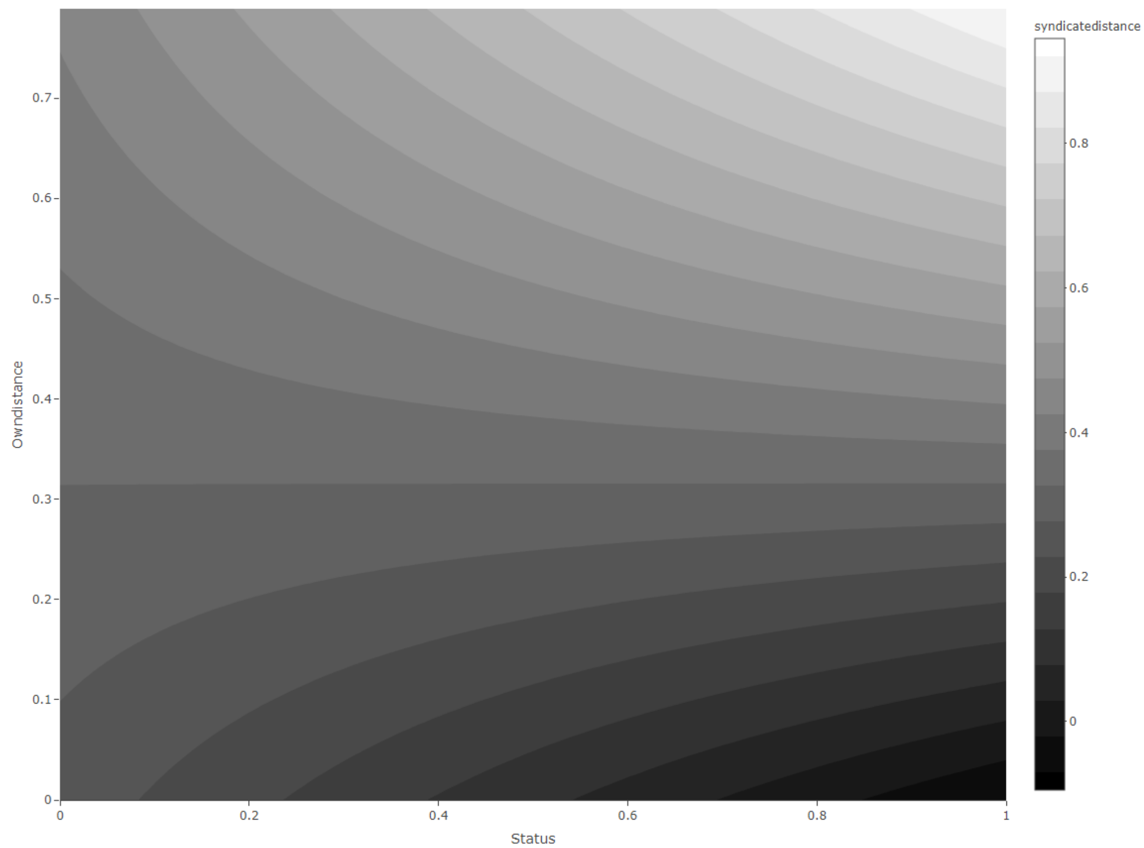
Regular 3D plot:



Interactive 3D plot:



Contour plot:



These plots are consistent with the regression result in Question3A, but of course not what the professor want. I assume that these plots should look in another way and could give such inference: when status is high, syndicate distance is negatively correlated with own distance. This can reveal the strategy of these high-status investors: when they themselves are familiar with the industry, they won't bother to find other experts in this area; when they are not that closely related to that industry, they will use their high status to coordinate experts in this sector to raise their chance of conducting a successful investment.

4.3 Conclusion

Sorry for may not showing the exact regression result and plots that the professor implied. Here, I only address what the professor expected about how high-status investors develop strategies to diversify more effectively.

To utilize other investors' talent, high-status investors would make full use of their status to cooperate with those investing experts in certain industries when they themselves are not familiar with, which is consistent with a saying, "learn from others' strong points to offset our weakness". This is such a successful strategy that make the high-status investors the only kind of investors that are better at diversifying.