Mini Project 1
Motherland ties among Chinese statisticians

LIU Haoyu and RONG Yi

March 13, 2017

1 Introduction

Motivated by a recent project on social ties and favoritism in Chinese science by National Bureau of Economic Research, we would like to see whether there exist similar motherland ties among Chinese statisticians. Different from the project mentioned above, the problem we are going to solve in this mini project is if there are preferences on nationality when Chinese statisticians choose their collaborators.

Specifically, we look at the coauthorship behaviors of five Chinese COPSS award laureates (Jeff Wu, 1987, Jianqing Fan, 2000, Tony Cai, 2008, Tze Leung Lai, 1983, and Wing Hung Wong, 1993) in the journal *Annals of Statistics*. The COPSS award is recognized as one of the most prestigious awards in statistics, and the five winners we select in this mini project are among most respectable Chinese statisticians. The journal *Annals of Statistics* publishes research papers of the highest quality reflecting the many aspects of contemporary statistics, and there are around 1000 papers included in total from 2003 to 2012.

The dataset we work on is a subset of the data on coauthorship and citation network for statisticians by Pengsheng Ji and

Jiashun Jin.[1] We apply several biased estimator techniques that introduced in class, including Linearized Bregman Algorithm for Generalized Linear Models (LB)[2] and Least Absolute Shrinkage and Selection Operator (LASSO).

2 Results

First, we show the solution paths by LB and LASSO, for Jeff Wu (Figure 1 and Figure 2), Jianqing Fan (Figure 3 and Figure 4), and Tony Cai (Figure 5 and Figure 6), respectively. Other plots are omitted here because of limited spaces. As the dataset covers papers between 2003 and 2012, relatively newer COPSS award winners (Jianqing Fan, 2000, for example) have denser solution paths compared to relatively older COPSS award winners (Jeff Wu, 1987, for example). There are both positive and negative coefficients on the solution paths. Positive coefficients suggest above average collaboration probabilities, while negative coefficients suggest below average collaboration probabilities. Compared with the results in J. Xiong, F. Ruan, and Y. Yao (2016), although COPSS award winners do not usually cooperate with other COPSS award winners, they do cooperate with people rather than work alone.

In order to find the possible preference on nationality and the coauthorship behavior, we look at the positive coefficients from now on, which indicates a high probability of coauthorship. Ta-

[1] P. Ji and J. Jin. Coauthorship and citation networks for statisticians. arXiv:1410.2840, 2014.

[2] J. Xiong, F. Ruan, and Y. Yao. A Tutorial on Libra: R package for the Linearized Bregman Algorithm in High Dimensional Statistics. arXiv:1604.05910, 2016.

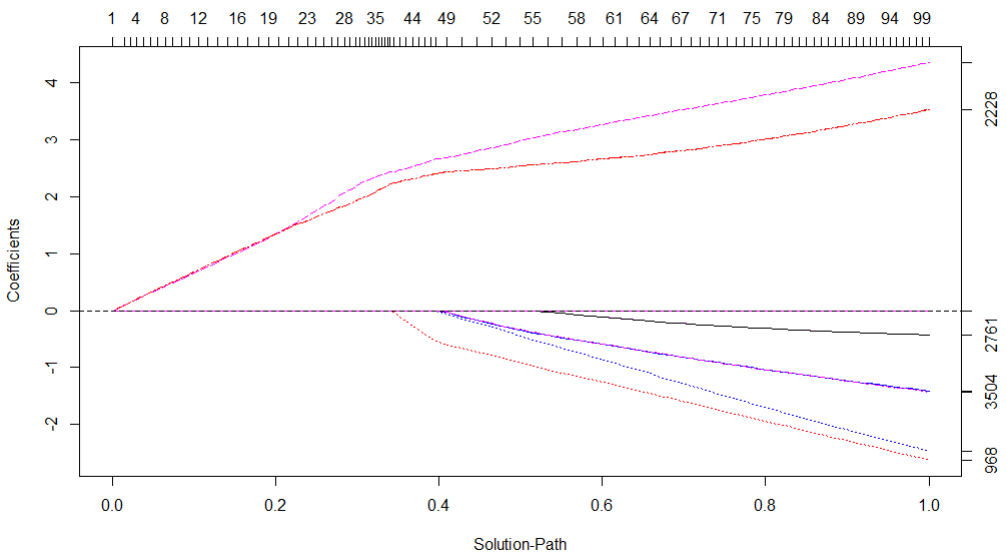

Figure 1: Solution Path by LB for Jeff Wu

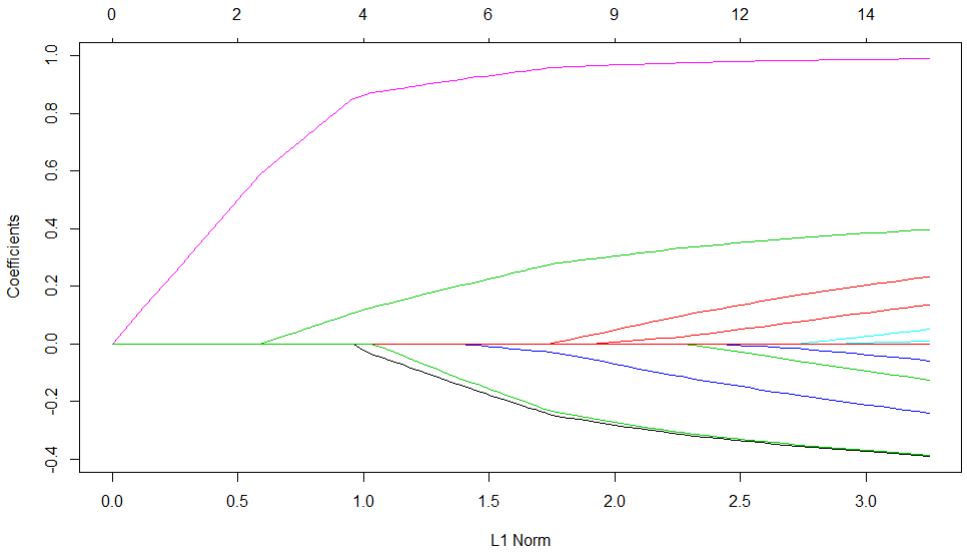

Figure 2: Solution Path by LASSO for Jeff Wu

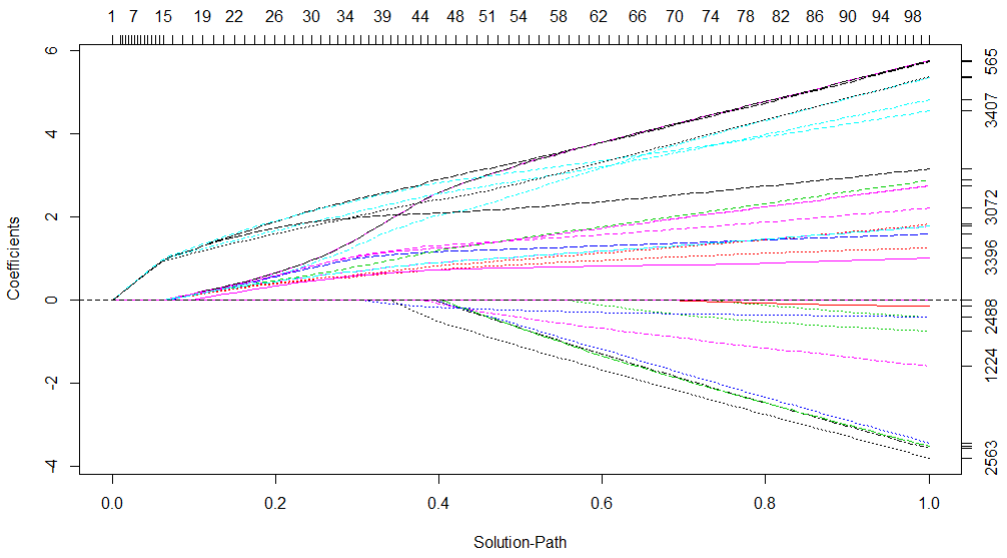

Figure 3: Solution Path by LB for Jianqing Fan

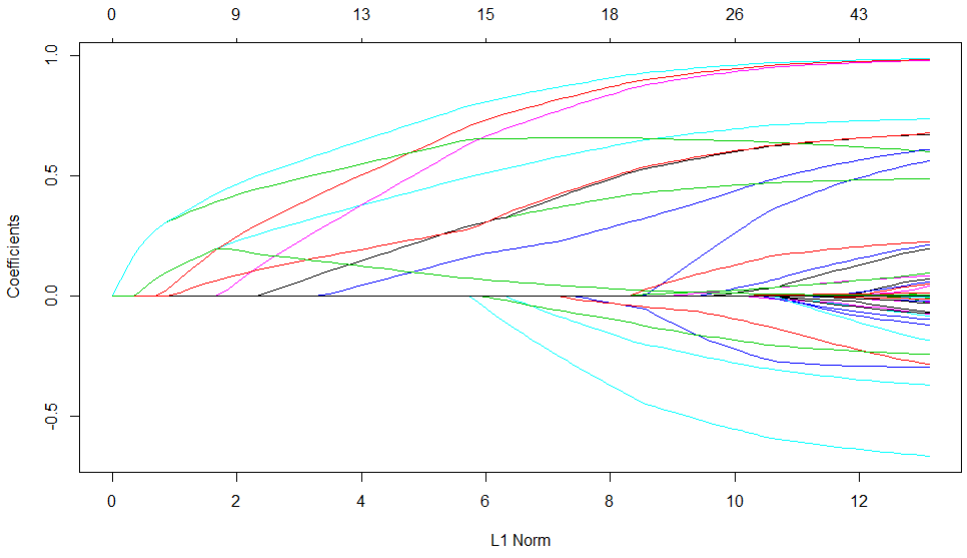

Figure 4: Solution Path by LASSO for Jianqing Fan

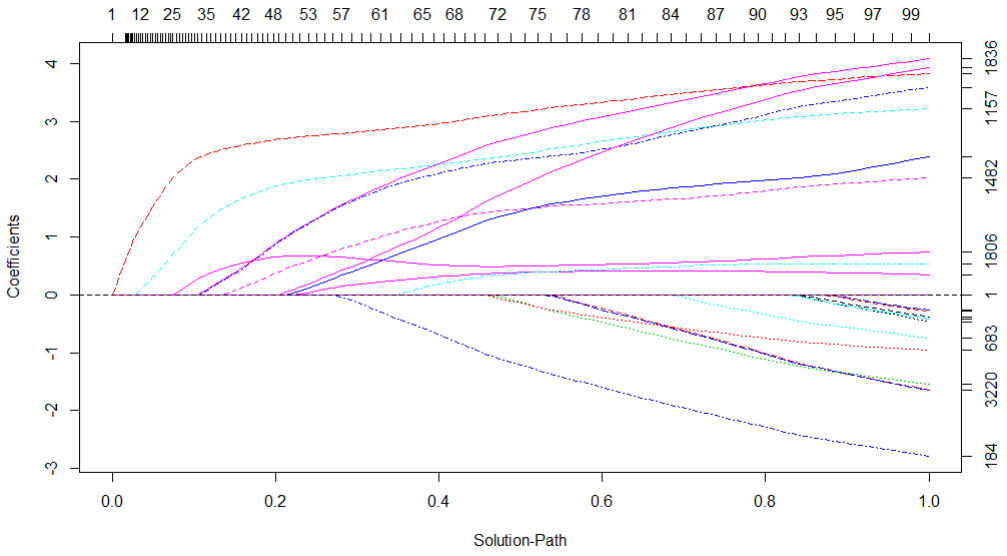

Figure 5: Solution Path by LB for Tony Cai

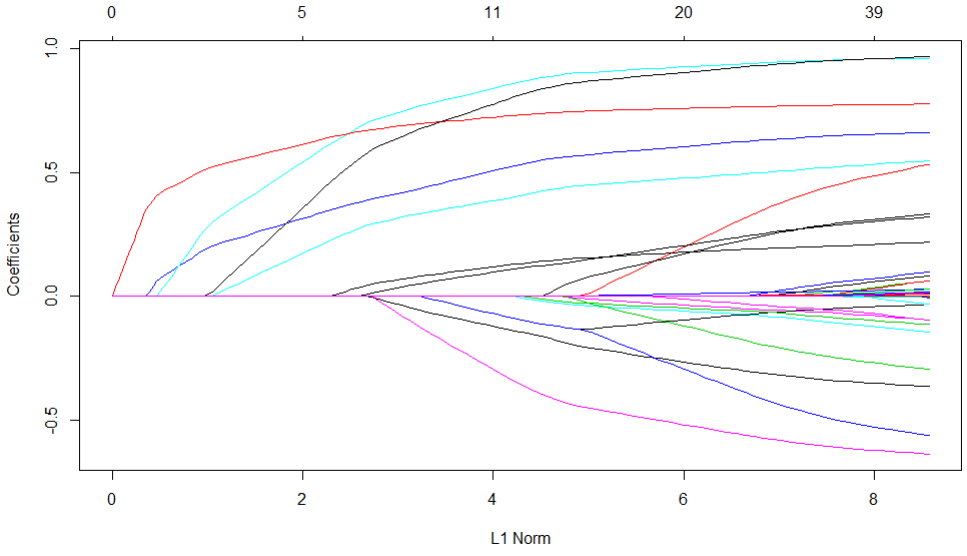

Figure 6: Solution Path by LASSO for Tony Cai

COPSS award winner	Possible coauthors	Percentage of Chinese
Jeff Wu	M. Ai and P. Qian, etc.	100%
Jianqing Fan	Y. Zhou, Y. Fan, Y. Feng, Y. Sun, W. Zhang, T. Yu, and R. Song, etc.	86%
Tony Cai	T. Jiang, J. Liu, and M. Zhou, etc.	50%
Tze Leung Lai	S. Chen, etc.	50%
Wing Hung Wong	A. Ding, L. Wang, and Q. Chen, etc.	100%

Table 1: Summary of positive coefficients

ble 1 summarizes the positive coefficients. To verify that our predictions are valid, we check the coauthors suggested by the results in reality. For example, Jianqing Fan does have two annals of statistics publications in the journal *Annals of Statistics* with Yong Zhou:
Cai, J., Fan, J, Zhou, H. and Zhou, Y. (2007) *Marginal hazard models with varying coefficients for multivariate failure time data.* The Annals of Statistics, 35, 324-354.
and
Fan, J., Lin, H. and Zhou, Y. (2006) *Local partial likelihood estimation for life time data.* The Annals of Statistics, 34, 290-325.
This means our predictions have certain reference value.

As we can see from Table 1, even though these Chinese COPSS award winners have extremely rich overseas working experience and do have lived in American for decades, almost all of them has a strong preference on Chinese nationality when choosing their collaborators. Their coauthorship behaviors have particular pattern. This phenomenon may be caused by language similarity, living and working habits, or just simply motherland ties. The reasons may be confirmed by future research.

To summarize, we apply several biased estimator tools in the

class. By analyzing the dataset on coauthorship and citation networks for statisticians, we find Chinese statisticians are more likely to work with another Chinese when they cooperate.

Reference

[1] P. Ji and J. Jin. Coauthorship and citation networks for statisticians. arXiv:1410.2840, 2014.

[2] J. Xiong, F. Ruan, and Y. Yao. A Tutorial on Libra: R package for the Linearized Bregman Algorithm in High Dimensional Statistics. arXiv:1604.05910, 2016.

Appendix: Codes

```
library(Libra)
library(glmnet)
data<-read.table(file.choose())
data<-t(data)
author<-c(392,1476,3057,3169,3315)
for (i in 1:5){
  print(author[i])
  y<-as.vector(2*as.matrix(data[1:960,author[i]])-1)
  x<-as.matrix(2*as.matrix(data[1:960,-author[i]])-1)
  path.lb<-lb(x,y,kappa=1,family="binomial",normalize=FALSE)
  plot(path.lb,xtype="norm",omit.zeros=FALSE)
  path.lars<-glmnet(x,y)
  plot(path.lars)
  lasso<-cv.glmnet(x,y)
  coefficient<-coef(lasso,s=lasso$lambda.min)
  print(which(coefficient>0))
}
```