20190228 Report

Fan Cheng

Feb 27, 2019

1 Workshop on Econometrics and Data Analytics

`Date`: Fri, Feb 23, 2019

1.1 Speaker: Denni Tommasi, Monash Business School

`Title`: LATE With Mismeasured or Misspecified Treatment: Women's Empowerment and Family Health

`Note`: Only listened to part of it due to the arrival of someone from eSolution to deliver my Mac. Met Denni later on Tuesday in Clayton for the milestone report. He was really active in listening and asking questions. More talk later.

1.2 Speaker: Chaohua Dong, Zhongnan University of Economics and Law

`Title`: High Dimensional Semiparametric Moment Restriction Models

`Note`: Jiti's collaborator from China. Didn't quite get the idea of the report.

1.3 Speaker: Jenny Williams, University of Melbourne

`Title`: Can Electronic Monitoring Redue Reoffending?

`Note`: Jenny was really funny. She was more like telling a story rather than reporting.

2 Milestone Reviews

2.1 Monday 25/2/2019 Caulfield

Speaker: Sium Bodha Hannadige (Confirmation)

Sium's thesis title is "Forecasting with Factor Augmented Regression Models: A Bootstrap Approach"

Note: She is working on factor augmented regression models. A bootstrap approach is used for the model estimation. She also listed the timetable for the future research.

2.2 Tuesday 26/2/2019 Clayton

2.2.1 Speaker: Chelle Wang (Progress Review)

Chelle's thesis title is " Adequacy and Sustainability of the Australian Retirement System using Modelling and Optimisation"

`Report title`: Optimising Australian Retirement Outcomes Under The New Pension Loan Scheme and Means Testing Owner Occupied Equity.

`Note`: A long title, maybe shorter. She wanted to study the influence of the new pension loan scheme so she built a complex dynamic model to look into the pension. Not finished yet. A simpler model, like OLS, could be added to emphasize the necessity of the dynamic model. The slides were made of long paragraghs of the report, making it difficult to follow her words. She was just reading her slides and there were no details and no time to think about them.

2.2.2 Speaker: Yuejun Zhao (Progress Review)

Yuejun's thesis title is "An Econometric Analysis of Mental Health Service Utilisation in Australia"

`Report title`: Widowhood and Cognitice Decline among US Elders: Mind Out if Time, or Time Not Out of Mind?

`Note`: Great Report! She focued on the direct impact of late-life widowhood and the gender difference. HRS dataset is used. Firstly, measure cognitive capacity by word recall tests and short TICS questionaire. Then, use the efficient GMM estimator and DPDM-FE model to look into the fixed effects (time-invariant) and synamic component (time-varying).

She was very elegant when reporting. Her slides were simple and there were some figures of the brain to illustrate the brain mechanism. I could get her idea even when I didn't know much about this area. She also listed her progress and the questions of the panel last year with her solutions. That was a good way to show her progress.

3 Enrolment

Finalised my enrolment on Feb 26 in Clayton. Campus card would be available today.

4 Coursework Timetable

ETX6500_CA_S1_ON-CAMPUS Statistical Inference

This class is held in CA_H/H4.87 at 15:00 every Thursday from 7/3-18/4, 2/5-30/5.

`TODO`: get a textbook.

5 Book Reading

I have been reading Chap 13 of the book `Modern Multivariate Statistical Techniques`. Here are the notes for the first half of this chapter.

5.1 Introduction

`Key words`: proximity, entity, Multidimensionl Scaling (MDS)

`Proximity`: degree of "closeness"

`MDS` is primarily a data visualization method for identifying "clusters" of points, where points in a particular cluster are viewed as being "closer" to the other points in that cluster than to points in other clusters.

The MDS methods in this chapter are classical scaling and distance scaling (divided according to whether the distances are f metric or nonmetric type, metric MDS and nonmetric MDS).

The standard treatment of classical scaling yields an eigendecomposition problem. The distance scaling methods use iterative procedures to arrive at a solution.

Example: Airline Distances

5.2 Two golden Oldies

The primary goal of MDS is to rearrange the entities in some optimal manner so that distances between different entities in the resulting spatial configuration correspond closely to the given proximities.

Classic examples: pairwise comparision of color stimuli and of Morse-code signals.

5.3 Proximity Matrices

Proximity can be a continuos measure of how physically close one entity is to another or it could be a subjective judgment recorded on an ordinal scale, but where the scale is sufficiently well-calibrated

as to be considered continuous.

Let δ_{ij} represent the dissimilarity of the ith entity to the jth entity. We arrane the m dissimilaritied, $\delta_i j$, into an $(m \times m)$ square matrix,

$$\Delta = (\delta_{ij}),$$

called a proximity matrix. The proximity matrix is usually displayed as a lower-triangular array of nonnegative entries, with the understanding that the diagonal entries are all zeroes and that the matrix is synnetric.For all $i, j = 1, 2, \ldots, n$,

$$\delta_{ij} \geq 0,\ \delta_{ii} = 0,\ \delta_{ij} = \delta_{ji}.$$

In order for a dissimilarity measure to be regarded as a metric distance, we also require that δ_{ij} satisfy the triangle inequality,

$$\delta_{ij} \leq \delta_{ik} + \delta_{kj},\ for\ all\ k.$$

5.4 Comparing Protein Sequences

The comparison problem is complicated by the fact that each sequence is actually a "word" composed of a string of letters selected from a 20-letter alphabet. The trick here is to align the two sequences so that the letters can be "matched".

Sequence alignment methods are generally divided into global and local methods. Global alignment tries to align all the letters in the two entire sequences, whereas local alignment assumes that the two sequences are highly similar only over short segments of letters. A sequence alignment is declared to be "optimal" if it maximizes an alignment score.

Example: Two Hemoglobin Chains.

SIM algorithm: local similarity program using dynamic programming techniques.

5.5 String Matching

A popular numerical measure of the similarity between two strings is edit distance (also called Levenshtein distance). The usual definition of edit distance is the fewest number of editing operations (insertions, deletions, substitutions) which would be needed to transform one string into the other.

Pattern matching: find a given pattern within a body of text.

Example: Employee Carees at Lloyds Bank.

The following Sections are about Classical Scaling and Distance Scaling. I have not finished them. By the end of this week, I will finish them and then focus on the citation of the important terms mentioned in this chapter.

5.6 R packages

There are also some R packages for Multidimensional Scaling.

Classical scaling can be carried out in R by using the command cmdscale [in library(mva)]. Sammon mapping can be computed using R command sammon [in library(MASS)]. A Fortran program, bmds, written by Ph and Raftery to compute Bayesian MDS is available at the STATLIB website. R contains the command isoreg [now in package stats, moed from package modreg] to compute isotonic regression. Kruskal's method of nonmetric distance scaling using the stress function and isotonic regression by using the command isoMDS [in library(MASS)].

6 TODO

I will finish reading the book first and try to run the commands mentioned above on the dataset on the book website.If possible, I plan to test the computaion efficiency on the example datasets. Then I will go on reading the important citations in this chapter.

7 Feedback