Statement of Purpose

Shuofan Zhang (Economics Department)

It is my desire to pursue a PhD in Economics at University California, San Diego as part of my long-term professional goal of becoming an academic researcher. I have a strong interest in applied econometrics and macroeconomics.

This passion stemmed from the first econometric course of my master's program that was taught by Professor Farshid Vahid. I was fascinated by the idea of tackling real-world problems based on economic theories and statistical methods, thereby I cannot stop reading the textbook (Wooldridge 2016). It was then that I discovered three errors in the equations. Surprised by these errors, Professor Vahid wrote to Professor Jeffrey Wooldridge, which resulted in for me his appreciation, public praise in class, and extra course credit points. Encouraged by this experience, I decided to pursue my interest in econometrics and transferred my specialization from actuarial studies to applied econometrics.

In the year that followed, I received the *Student Excellence Award* for achieving the highest mark in seven out of eight courses. In April 2018, I was chosen as one of the four representatives of Monash Business School to participate in the *Econometric Game*[1], where the best PhD and master students from thirty prestigious universities including Harvard University and University of Cambridge study a same question and deliver academic papers for competition. This year's research topic was around the detrimental effects of an individual's unemployment on that individual's happiness as well as on a group's wellbeing. After reviewing the relevant literature for the empirical support on selecting explanatory variables, our team constructed an ordered probit model with the raw responses to the survey item that elicits respondents' overall life satisfaction as the dependent variable. Under the assumption of the homogeneous spillover effects amongst individuals in a group, we then estimated the multiplier between the effects on an individual and a group. The potential simultaneity bias was discussed. This great experience of conducting rigorous research provided me with important new insights into how unemployment affects well-being, and significantly stimulated my interest in empirical research in economics.

Measures of happiness, previously psychologists' tools, now are widely used by economists; likewise, statisticians' findings are infiltrating econometrics.

[1]The Econometric Game is hosted by the University of Amsterdam. See http://econometricgame.nl/

As an attempt to introduce a statistical technique to econometrics, my master's thesis employed deep learning to facilitate the hypothesis test design, which was supervised by Professor Dianne Cook. Deriving hypothesis tests and the asymptotic distributions constitute a considerable part of the econometric literature. However, the derivation is often complex and the resulting test may be powerless. For example, the Hausman tests and the unit root tests all suffer from low power. In my thesis, I trained a binary deep learning classifier to detect linear relationships from no structure in a scatter plot as an alternative to the conventional t-test. Tested on a large unseen dataset, the power $(1 - \beta)$ of the classifier was always close to the t-test holding the type I error (α) constant. Given that the t-test is the known uniformly most powerful test under such experimental settings (Neyman and Pearson 1933), this finding suggests the deep learning model has the potential to approach the unknown best test in more complicated situations. The study was then extended to test the null of homoscedasticity against heteroscedasticity using the retrained classifier. A small dataset of human evaluations was collected via an online questionnaire (Majumder, Hofmann, and Cook 2013) and a specific form of the White test (White 1980) was applied to provide a reference level of the test accuracy. The classifier achieved much higher accuracy than both the White test and humans. Using the data plot as the test statistic instead of a single quantity, this approach could avoid complex derivations while exploiting the useful information in the plots (Anscombe 1973). Although the training design could be challenging, these results empower future research. For instance, by replacing the scatter plot with time plots one can use this approach to test the unit root in time series. I am currently writing a first-author paper on these results to submit to the *Statistical Analysis and Data Mining*. This study was also presented by Professor Cook as the 50^{th} *Belz* Lecture for the *Statistical Society of Victoria*[2].

After completing the master degree, I was excited to accept an offer from Professor Heather Anderson and Professor Farshid Vahid as a research assistant to work on a paper that studying the high-dimensional predictive regression with the LASSO[3] (Koo et al. 2016). In this project, I reviewed the literature to investigate the compatibility (or restricted eigenvalue) condition and its implication of choosing the tuning parameter λ for the ℓ_1 norm to achieve the prediction optimality, while taking into account the potential consequences of inconsistency for variable selection by the LASSO (Bühlmann and Van De Geer 2011). I have been self studying Real Analysis by reading the book *Principles of Mathematical Analysis*(Rudin 1976) to better understand the relevant concepts. Comparison between the LASSO, the adaptive LASSO and

[2]The Victorian Branch of the Statistical Society of Australia Inc. See https://www.statsoc.org.au/branches/victoria/

[3]LASSO is the abbreviation for Least Absolute Shrinkage and Selection Operator.

the group LASSO is under considering. In addition, the out-of-sample mean squared errors of forecasting GDP growth and inflation rate using the LASSO on 146 macroeconomics variables are compared against other approaches including an autoregressive model and a principal component analysis. The potential co-integrating relationships in the selected variables are being studied. Observing the series chosen by this data-driven methodology and seeking possible economic theory-based explanations are intriguing and encourage me to learn more. I want to continue to work on macroeconomic problems and apply cutting-edge techniques to challenging areas such as high-dimensional data and forecasting economic or financial series.

Meanwhile, I am co-authoring a paper with the Learning and Teaching team at Monash University that measuring student levels of perceptions of live-streaming, a new technology implemented in the lectures. Our study adapted the CRiSP[4] questionnaire (Richardson et al. 2015) which was validated by a combination of factor analyses. Results revealed three reliable scales: acceptance, usability, and confidence. Following the validation results, I have investigated the correlations between student perceptions of live-streaming and their study attitudes using the estimated factor scores. The treatment effect of live-streaming on students academic performance is also being studied by the fixed effects model with panel data accounting for the unobserved heterogeneity.

Though I am open to a variety of research in Economics, I find the work of Professor Elliott, Timmermann, and Hamilton are especially interesting to me. I am confident that my talent in Mathematics combined with my various research experiences in Economics will allow me to contribute positively to the PhD program at University California, San Diego.

References

Anscombe, FJ. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27 (1): 17–21.

Bühlmann, Peter, and Sara Van De Geer. 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media.

Koo, Bonsoo, Heather M Anderson, Myung Hwan Seo, and Wenying Yao. 2016. "High-Dimensional Predictive Regression in the Presence of Cointegration."

Majumder, Mahbubul, Heike Hofmann, and Dianne Cook. 2013. "Validation of Visual Statistical Inference, Applied to Linear Models." *Journal of the American Statistical Association*

[4]CRiSP is the name of classroom response system perceptions questionnaire.

108 (503). Taylor & Francis Group: 942–56.

Neyman, J., and E. S. Pearson. 1933. "IX. on the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 231 (694-706). The Royal Society: 289–337. doi:10.1098/rsta.1933.0009.

Richardson, Alice M, Peter K Dunn, Christine McDonald, and Florin Oprescu. 2015. "CRiSP: An Instrument for Assessing Student Perceptions of Classroom Response Systems." *Journal of Science Education and Technology* 24 (4). Springer: 432–47.

Rudin, Walter. 1976. "Principles of Mathematical Analysis (International Series in Pure & Applied Mathematics)." McGraw-Hill Publishing Co.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica: Journal of the Econometric Society*. JSTOR, 817–38.

Wooldridge, Jeffrey M. 2016. *Introductory Econometrics: A Modern Approach, 6th Edition*. Cengage.