Statement of Purpose

Shuofan Zhang, Economics Department

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

It was the first econometric course of my master's program taught by Professor Farshid Vahid that piqued my interest. I was fascinated by the idea of tackling real-world problems based on economic theories and statistical methods. I could not stop reading the textbook written by Wooldridge (2016). In my reading, I discovered three errors in the equations. Surprised by these errors, Professor Vahid wrote to Professor Jeffrey Wooldridge, which resulted in 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 *Monash Business School Student Excellence Award* for achieving the highest mark in seven of my eight courses. In April 2018, I was chosen as one of the four representatives of Monash University to participate in the *Econometric Game*[1], where teams of postgraduate students examine a research topic and deliver academic papers for competition. Thirty prestigious universities were represented, including Harvard University and the University of Cambridge. The research topic considered 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 capturing 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 and implications were 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 tools used by psychologists, are widely adopted by economists nowadays; likewise, statistical advances are infiltrating econometrics.

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. The derivation of hypothesis tests and their asymptotic distributions constitute a considerable part of the

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

econometric literature. However, the derivation is often complex and the resulting test may lack power. For example, the commonly used unit root tests all suffer from low power in distinguishing the unit root null from stationary alternatives. In my thesis, I trained a binary deep learning classifier to test the null of no structure against linear patterns 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 known as the uniformly most powerful test under such experimental settings according to the Neyman–Pearson lemma, this finding suggests that 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 re-trained classifier. A small dataset of human evaluations was collected via an online questionnaire and a specific form of the White test was applied to provide a reference level of the test accuracy. The classifier achieved much higher accuracy than both the White test and the human evaluations. 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 (see Anscombe (1973) for examples). Although the training design could be challenging, these results encourage future research. For instance, by replacing the scatter plot with time plots one can use this approach to perform a unit root test. I am currently writing a first-author paper on these results to submit to *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 my master's degree, I was excited to accept an offer from Professor Heather Anderson and Professor Farshid Vahid as a research assistant to work on extensions to a paper studying high-dimensional predictive regression with the Least Absolute Shrinkage and Selection Operator (LASSO) (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 the group LASSO is under consideration. 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 provoked my intellectual curiosity. I would like to continue to work on macroeconomic problems, applying cutting-edge techniques to challenging areas such as high-dimensional data and economic or

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

financial series forecasting.

Meanwhile, I am co-authoring a paper with the Learning and Teaching team at Monash University which measures student levels of perceptions of live-streaming, a new technology implemented in the lectures. Our study adapted the CRiSP[3] questionnaire which was validated by a combination of factor analyses. Our results revealed three reliable scales: acceptance, usability, and confidence. Following the validation results, I investigated the correlations between the three scales and the self-reported study attitudes using the estimated factor scores. The causal effect of live-streaming on academic performance is also being studied using a fixed effects model with panel data accounting for unobserved heterogeneity.

Although I am open to a variety of topics in economics, there are several professors at UC San Diego whose projects are especially appealing to me: Professors Elliott, Timmermann and Hamilton. After reading some of their papers, I believe their work is closely aligned with my skills and interests and that UC San Diego will be a great environment for me to thrive.

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." https://ssrn.com/abstract=2851677.

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

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

[3] CRiSP is the name of classroom response system perceptions questionnaire. (Richardson et al. 2015)