Statement of Purpose

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 a professor. I have a strong interest in applied econometrics.

After graduating from college, I spent three years working for MSH China[1] to meet my student loan obligations before beginning a Master's degree in Actuarial Studies at Monash University[2]. My goal was to facilitate a future career as an actuary. However, I developed a keen interest in econometrics and academic research. This passion stemmed from three mistakes I discovered in equations for the OLS estimator in the textbook (Wooldridge 2016) while doing extracurricular reading for the Introductory Econometrics course taught by Professor Farshid Vahid. Surprised by these typographical errors, Professor Vahid wrote to Wooldridge, which resulted in a thank you from him, public praise, and extra credit points from Farshid. Encouraged by these positive experiences, I decided to pursue my interest and transferred to a Master's degree in Applied Economics and Econometrics.

In the year that followed, I achieved the highest mark in seven out of eight courses. In April 2018, I was chosen as one of the four representatives to participate the Econometric Game[3]. We were required to write a paper around the topic of the economics of happiness over the course of two days. I can still feel the passion every time I remember how we worked together under great pressure discussing ways of measuring "wellbeing", constructing the ordered probit model, and screening for the relevant variables. I felt deeply proud of ourselves for being able to tackle a meaningful and interesting economics problem, while also being stimulated greatly to learn more.

Inspired by the diverse expertise of the PhD students from other top universities, I wanted to equip myself with modern techniques such as machine learning and data visualization to do cutting-edge research in Economics. Returning from the Econometric Game, I began work on my masters thesis on computer vision with rigor. This project was supervised by Professor Di Cook. The central research question was: can we train the computer to read residual plots? Residuals plots contain more information than single statistics (Anscombe 1973). It has been shown that hypothesis test based on human observations of visualized plots is valid

[1]MSH China is the Asia-Pacific Headquarter of MSH INTERNATIONA, a world leader in the design and management of international healthcare solutions. See http://www.mshchina.com/

[2]Monash University is a public research university based in Melbourne, Australia.

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

(Majumder, Hofmann, and Cook 2013). In the hope that human evaluation can be aided and supplemented by computers, I built a binary classifier using a deep learning algorithm to detect linear relationships from no structure in a scatter plot. Having trained a satisfactory model using simulated data, I tested it against a large unseen test set, as well as the same dataset used by Majumder et al. An interesting discovery was that the power $(1 - \beta)$ of the deep learning model was always close to the conventional 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 gives hope that the deep learning model may be able 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 binary classifier. A small experiment of human observations was conducted via online questionnaire and a specific form of the White test (White 1980) was employed to provide a reference level of the test accuracy. In this experiment, the classifier achieved much higher accuracy than both the White test and humans. Our research extended the use of computer vision on hypothesis testing and provided evidence of its validity when the structure in the plots is very specific. A draft paper is currently being written to be submitted to the journal: Statistical Analysis and Data Mining.

After completing the master degree, I was excited to accept an offer from Professor Heather Anderson as a research assistant to work on a paper that studying the high-dimensional predictive regression with the Lasso estimator (Koo et al. 2016). In this project, I reviewed 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 in consistency for variable selection of the Lasso (Bühlmann and Van De Geer 2011). Comparison between the Lasso, the adaptive Lasso and the group Lasso is under considering. In addition, the out-of-sample mean squared error of forecasting GDP growth and inflation using the Lasso on 146 economics variables is tested against other approaches including an autoregressive model and a principal component analysis. The potential co-integrating relationships in the selected variables are being studied. I find that it is intriguing to observe the series chosen by this data-driven methodology and seek possible economic theory-based explanations.

Meanwhile, I am working for the learning and teaching team at Monash University to analyze the treatment effect of the education innovations such as live streaming used in the lectures, using IV estimation. We adapted explanatory factor analysis to establish the validity of the online survey conducted for the users experience.

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* 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.

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.