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 and macroeconomics.

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 the equations 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 (for me) a thank you from him, public praise, and extra course credit points. Encouraged by this experience, I decided to pursue my interest and transferred to a Master's degree in Applied Economics and Econometrics.

In the year that followed, I received the Monash Business School Student Excellence Award for achieving the highest mark in seven out of eight courses. In April 2018, I participated the Econometric Game[3] as one of the four representatives of Monash Business School competing with the best PhD and master students from other prestigious universities. The research question for competition 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, we 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 the effects on an individual and a group. The potential simultaneity bias was discussed. This great experience of conducting rigorous research provided me important new insights into how unemployment affects well-being, and significantly stimulated my interest in empirical research in economics.

[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/

I have also participated in another interesting competition on Kaggle[4] forecasting housing price as part of the assessments for the statistics subject taught by Professor Dianne Cook, where I achieved the second minimum mean squared error by running a random forest regression. Studying more on the machine learning technique, I realized its similarity to econometrics and became curious about possible applications of this novel tool to economic studies.

In the context of economics, detecting violations of the Gauss-Markov assumptions in multiple regression models is crucial for robust interpretations of the empirical results. However, it is often difficult to derive powerful distribution test from complex mathematical theorems. Besides, Anscombe (1973) has shown that single statistics may contain less information than data plots. Being supported by Majumder, Hofmann, and Cook (2013), where the hypothesis test based on human evaluations of visualized plots has been shown as valid, I explored an innovative method of performing hypothesis tests that could avoid the arduous derivations while exploit the useful information in the residual plots. More specifically, I experimented 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 and compared the accuracy with the conventional t-test. Interestingly, 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 a 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 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 test, the classifier achieved much higher accuracy than both the White test and humans. These results empower future research on this new methodology. 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, Australia.

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 (Least Absolute Shrinkage and Selection Operator) (Koo et al. 2016). In this project, I reviewed literature to investigate the

[4]Kaggle is an online community of data scientists and machine learners, owned by Google, Inc. See https://www.kaggle.com/

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. Observing the series chosen by this data-driven methodology and seeking possible economic theory-based explanations are intriguing and encourage me to learn more.

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 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 research experiences in Economics will allow me 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 Statis-

tical 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.

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.

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.