Statement of Purpose

It is my desire to pursue a Ph.D. in Economics as part of my long-term professional goal of becoming a professor. I have a strong interest in applied economics and econometrics.

After graduating from college, I spent three years working for MSH China[1] to meet my student loan obligations. Having achieved this, I began the next phase of my education – a Master degree in Actuarial Studies at Monash University[2]. My plan was to get a degree to facilitate my future career, but it was during this time that I discovered my passion and talent for academic research. The first trigger happened in the Introductory Econometrics course taught by Professor Farshid Vahid. I found three mistakes in the formulas of the OLS estimator while doing extracurricular reading on the textbook (Wooldridge 2016). I brought this to Farshid and even he was surprised by these typos. He wrote an email to Wooldridge, this resulted in a thank you from Wooldridge, a public praise and some bonus points from Farshid. Highly encouraged by these rewards, the more I studied Econometrics, the more I enjoyed it. I decided to transfer to Master of Applied Economics and Econometrics. In the year that followed, I achieved the best performance in seven out of eight courses. In April 2018, I was chosen as one of the four representatives to participate the Econometrics Game[3]. In two days' time, we wrote a paper to find "the multiplier between the detrimental effect of unemployment on the wellbeing of the individual who is unemployed and the total effect of that person's unemployment on the wellbeing of the group as a whole" (P.Frijters 2018). 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, etc. I was impressed by the breadth and depth of the top paper given the tight schedule, while feeling deeply proud of ourselves. Our skills and knowledge open the door to a unique perspective of this world. Our work may boost an individual's happiness and promote a country's development.

I learned a fact, by communicating with the PhD students from other top universities, that knowing some cutting-edge techniques, such as machine learning or data visualization, is very useful for doing research in Economics. Coming back from the game, I delved into my master thesis which relates to computer vision. This project was supervised by Professor Di Cook.

[1]MSH China is the Asia pacific headquarter of MSH INTERNATIONA, a world leader in the design and management of international healthcare solutions. 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. http://econometricgame.nl/

The central question was: can we train the computer to read residual plots? Residuals plots contain more information than single statistics (Anscombe 1973). It has been proved that the hypothesis testing of human observations of visualized plots is valid (Majumder, Hofmann, and Cook 2013). In the hope that previously manual process can be aided and supplemented by computers, I built a binary classifier using a deep learning algorithm which detects linear relationships from no structure in the scatter plots. 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 (J.Neyman and E.S.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 heteroscedasticity from homoscedasticity using the binary classifier. To provide a reference level of the test accuracy, a small experiment of human observations was conducted via online questionnaire and a specific form of white test (White 1980) was employed. In this experiment, the classifier achieved much higher accuracy than the white test and humans. Our research innovated the use of computer vision on hypothesis testing and provided evidence of its validity when the structure in the plots is very specific. Now we are working towards the publishment of these results.

After completed the master degree, I was jubilant to accept the offer from Professor Heather Anderson as a research assistant to work on a paper about the high-dimensional predictive regression with the Lasso estimator (Koo et al. 2016). In this project, I review 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, whereas detect the potential consequences in consistency for variable selection of the Lasso (Bühlmann and Van De Geer 2011). The comparison between the Lasso, the adaptive Lasso and the group Lasso is also considered. In addition, the Lasso's performance of forecasting GDP and inflation on 146 economics variables is tested against other approaches including an Autoregressive model and a principal component analysis etc. The potential co-integrating relationships in the selected variables are being studied. I find it is interesting to stare at the series chosen by this data-driven methodology and think about the possible economically theory-based explanation. Meanwhile, I am working for the learning and teaching team in Monash University to analyze the treatment effect of the education innovations used in the lecture by IV estimation. In addition, we adapted factor analysis to construct the students' study approaches scores (deep or surface) from a survey's data.

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.

J.Neyman, 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.

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.

P.Frijters. 2018. "The Central Question of the Econometric Game." http://econometricgame.nl/econometric-game-2018/.

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.