

# 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<sup>1</sup> 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, Australia. 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<sup>2</sup>. When I checked the formulas I derived for the estimated coefficient of OLS with the textbook (Wooldridge 2016), I found our equations differed. I was confident about my answers. Therefore, 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, and some bonus points from Farshid. With this experience, my confidence and interests both increased exponentially. The more I studied Econometrics, the more I enjoyed it. I decided to transfer to Master of Applied Economics and Econometrics after the first semester.

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 held in Amsterdam University. 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 also feeling deeply proud of being part of this brilliant group - the future economists.

By communicating with the PhD students from other top universities, I also learned a fact that knowing some cutting-edge techniques, such as machine learning or data visualization,

---

<sup>1</sup>MSH China is the Asia pacific headquarter of MSH INTERNATIONAL. Founded in 1974, MSH INTERNATIONAL is a world leader in the design and management of international healthcare solutions. <http://www.mshchina.com/>

<sup>2</sup>Farshid Vahid is a professor in the Department of Econometrics and Business Statistics at Monash University.

is very useful for doing research in Economics. Coming back from the game, I delved into my own research project which is related to computer vision. This project was supervised by Professor Di Cook<sup>3</sup>. The central question was: can we train the computer to read residual plots? Residuals plots contain more information than single statistics (Anscombe 1973), thus are a primary means to diagnose statistical models. Majumder, Hofmann, and Cook (2013), has proven hypothesis testing of human observations of visualized plots to be a valid methodology. In the hope that previously manual process can be aided and supplemented by computers, a binary classifier was built using a deep learning algorithm which detects linear relationships from no structure in the scatter plots. Having trained a satisfactory model using simulated data, we tested it against the same dataset used by Majumder et al. An interesting discovery was that our deep learning model behaved similarly to the conventional t-test. Theoretically, t-test is the uniformly most powerful test under such experimental settings (J.Neyman and E.S.Pearson 1933). The fact that the deep learning model approaches t-test gives hope that it may be able to find 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. This gives more evidence of the validity of hypothesis testing of computer vision of data plots. Now we are working towards the publishment of these results.

After completed the master degree, I accepted the offer from Professor Heather Anderson<sup>4</sup> as a research assistant to work on an interesting project about the high-dimensional predictive regression with Lasso<sup>5</sup> estimator (Koo et al. 2016). In this project, I help to investigate the consistency and efficiency of Lasso in variable selection by reviewing the relevant literature (Bühlmann and Van De Geer 2011). We are using this methodology to forecast GDP on 146 series to test its performance against other approaches including Autoregressive model etc. 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).

---

<sup>3</sup>Dianne Helen Cook is an Australian statistician, the editor of the Journal of Computational and Graphical Statistics, and an expert on the visualization of high-dimensional data.

<sup>4</sup>Heather Anderson is the Maureen Brunt Professor of Economics and Econometrics, in the Department of Econometrics and Business Statistics at Monash University.

<sup>5</sup>Least Absolute Shrinkage and Selection Operator

## References

- Anscombe, F.J. 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.