

Statement of Purpose

Qianfan Wu

Personal Website: <https://sites.google.com/view/qfwu>

I want to pursue a Ph.D. in IROM, specialized in **Statistics** or **Decision Science**. My career aspiration is to become a professor or a team leader in an industry research group. I have research experience related to genomic data analysis, survival analysis, statistical learning, and have published my research twice [1,2].

My motivation for using statistical methods to analyze complex datasets started from the [Experiential Learning Project](#) I participated in at Northern Illinois University. As an active research student consultant, I analyzed large unstructured demographic datasets and built several predictive models to efficiently locate low-income households for Northern Illinois Food Bank. This experience sparked my interest in designing statistical models and efficiently using them in different domains.

I gained research experience in statistical learning by working as a research assistant at Brigham and Women's Hospital / Harvard Medical School. Supervised by Professor [Weiliang Qiu](#), I conducted extensive literature research on machine/deep learning feature extraction, and tested auto-encoders, a neural network-based method, on several genomic datasets. In addition, due to some recent paper's inappropriate use of performance measures, I conducted multivariate simulation analysis to compare the robustness of popular performance measures. I also summarized the pros/cons of the current research and discussed possible improvements. At the end of my internship, I presented my research at the statistics seminar at Brigham Women's hospital and published my literature research paper [1] on *Journal of Biometrics and Biostatistics*, as the first author.

I enhanced my research and programming skills through my internship as a quantitative researcher at [Algo Depth](#). Through my research, I proposed several novel ideas in using survival analysis to predict the execution time of limit orders in the stock market. Also, I worked closely with other team members to implement Cox/APT survival models on the trading system. Moreover, based on independent research, I developed an [online course](#) on applying survival analysis in finance. With well-designed videos and materials, the course largely facilitated the promotion of my research and brought in plenty of insightful discussion by viewers. Furthermore, I designed a flexible information retrieval application using Python and C++ that largely accelerated the speed of processing financial data from open sources. This project cut the data acquisition budget and improved strategy implementation efficiency.

My full-time working experience at [M-CAM International](#) has further reinforced my motivation to pursue a Ph.D. in quantitative disciplines. Supervised by [Dr. David Martin](#) and [Dr. Carol Corrado](#), I worked on a research project on using statistical methods to measure the innovation ability of companies. I performed hypothesis testing using patent citation data and tested the method previously proposed by the team. In addition, I developed a new financial data feature construction method, which is designed to discover optimized time series window size and eliminate redundant features. The new method improved the prediction accuracy of a company's future innovation-driven performance

and led to the development of [The Conference Board Innovation \$\alpha^{\circ}\$ index](#). The paper I prepared on this topic has been accepted by *the American Economic Association 2020 Annual Conference* [2].

I am applying for the UT-Austin IROM Department because of its strength and interdisciplinary nature in data science research. The course offered by the program will complement the skillsets I've already built. Though I am open to a wide variety of research within IROM, I am particularly interested in constructing predictive models using feature selection techniques using real-world data. I especially find the work of Prof. **Carlos M. Carvalho** interesting and wish to contribute his research group on solving high-dimensional problems in finance. The research work of Prof. **Mingyuan Zhou**, **Sinead Williamson**, and **Rui Gao** in statistical learning and Bayesian predictive models are also appealing to me. I am confident that my quantitative skills and research experience will positively contribute to the IROM department.

References

- [1]. **Wu, Q.**, Boueiz, A., Bozkurt, A., Masoomi, A., Wang, A., DeMeo, D.L., Weiss, S.T. and Qiu, W., 2018. Deep Learning Methods for Predicting Disease Status Using Genomic Data. *Journal of biometrics & biostatistics*, 9(5). (Published).
- [2]. Corrado, C., Martin, D. and **Wu, Q.**, 2019. Innovation α° : What Does an Intangibles-intensive Stock Price Index Tell Us about Business Profitability and Productivity? *American Economic Association Annual Conference 2020* (Accepted; Estimated publication 1/20/2020).