Research Statement

Ying-Kai Huang
University of Pittsburgh

I am an applied economist who has broad interests in applied econometrics with state-of-the-art techniques in modern data science and empirical behavioral economics. One main focus of my research is developing applied econometric forecasting methods with neural networks. Another set of topics I study is how behavioral biases affect users' experiences and influence information aggregation processes on online platforms.

In my job market paper, "From Econometrics to Machine Learning: Application of Recurrent Neural Networks on Yield Curve Forecasting," I use recurrent neural networks to improve yield curve forecasting. In the paper, I construct a novel forecasting model using long short-term memory (LSTM) neural networks with autoencoder structures and show that the model improves forecasting root mean square errors compared to models in previous literature. The autoencoders in my model work as dimension reduction devices similar to nonlinear principal component analyses and transform higher dimensional information from the data into lower dimensional latent factors. Combining the autoencoders with LSTM, my model can be considered as a special type of state-space models. Moreover, I demonstrate the new model is actually a generalization of the previous seminal paper by Diebold and Li (2006). Without restrictions on yield curve functional forms, my model is more flexible than Diebold and Li's model. Moreover, the similarities between our models allow me to keep some desired theoretical properties in previous yield curve forecasting literature while having better out-of-sample forecasting performance.

In the second chapter of my dissertation, I focus on the problem of how loss aversion can influence consumers' evaluations of businesses. To answer this question, I use data from Yelp and utilize the rounding features of average ratings on the online platform. The average ratings on Yelp are rounded to the closest half stars to allow users to understand qualities of businesses more easily. For example, if the average rating for a business is 4.25, Yelp users will see a 4.5 star when they check the business on Yelp. And if the average rating is 4.24, users will instead see a 4.0 star. This rounding feature causes discontinuities for businesses with similar average ratings around the rounding thresholds. The discontinuities create potential surprises for consumers. When the star on Yelp for a business is higher than its average rating, consumers may feel disappointed when they go to the business if they check Yelp before their visits. On the other hand, if the star that consumers see is lower than the average rating, there is a positive surprise for them. Taking advantage of the rounding feature, I use a regression discontinuity design to check how these discontinuities influence people's evaluations of businesses with similar qualities. From my analysis, I show that loss aversion has a significant effect on people's experiences and can make the average ratings differ by around half a star for businesses with similar qualities. When evaluating the efficiency of the platform from a perspective of information aggregation, we can see a huge loss because of the behavioral bias caused by the rounding feature.

In my third chapter, "When Hope Hurts: How Special Occasions Lead to Attribution Bias," I work with two other students and we study how special occasions change people's expectations about their dining experiences and influence their evaluations. Similar to my second chapter, we utilize the data of Yelp and apply behavioral economic theories to form testable hypotheses. In the paper, we applied econometric methods such as propensity score matching and exact matching to create counterfactuals and estimate the effect of special occasions on Yelp users' ratings. In addition to traditional applied econometric techniques, we combine recently developed machine learning tools in natural language processing such as keyword parsing and structural topic models to determine whether consumers visited a restaurant on special occasions and analyze how special occasions change peoples' attentions when they evaluate a restaurant. By using those tools and analyzing text reviews from consumers, our analyses help us understand consumers' behaviors in a way that was not available in previous literature. For instance, we are able to conduct sentiment analysis on their reviews from the vocabularies they used and create a new control variables based on their emotions. We can also estimate the effects of some variables that could influence the focuses of consumers' reviews. Our results show that special occasions increased peoples' expectations and caused consumers' disappointments that led to lower ratings when they wrote their reviews. From our structural topic models, we also see that special occasions affected the evaluation standards of consumers and this result further supports our hypotheses about how special occasions change peoples' expectations.

In addition to the research in my dissertation, I also work with the experimental economics group at Pitt. In our ongoing project, we compare two most-common scoring rules, the binarized BDM and binarized QSR in economic experiments and evaluate their performance on eliciting beliefs from subjects with arbitrary risk preferences. Our results show that though both mechanisms lead subjects to provide accurate beliefs, they all create substantial noise in their responses. When we consider usage of the scoring rules under a fixed experimental budget, the responses in the binarized QSR are less noisy and this property makes QSR a preferred mechanism for incentivizing subjects if budgets are limited.

Going forward, I will continue my research in both financial forecasting with machine learning and empirical behavioral economics. For the yield curve forecasting research, I plan to connect my model more closely to the theoretical term-structure literature. To do so, I will connect the recently developed constraint optimization techniques for neural networks in financial math such as those developed in Kratsios and Hyndman (2017) to explore the possibilities of making my model satisfy the arbitrage-free property. In addition, I will study how my forecasting yield curves change before different kinds of market shocks in order to evaluate the validity of my model when facing financial crises. I would also like to discover more about theoretical links between recurrent neural networks and autoregressive time-series models and explore more potential applications of neural networks to applied econometrics. For the research in empirical behavioral economics, I will continue to use machine learning tools in text analytics to study text reviews on online platforms to shed further light on consumers' behaviors. My coauthors and I plan to extend the structural topic modeling techniques to test the main factors that shift consumers' focuses and improve the online information aggregation process.

References

Diebold, Francis X and Canlin Li, "Forecasting the term structure of government bond yields," *Journal of econometrics*, 2006, *130* (2), 337–364.

Kratsios, Anastasis and Cody Hyndman, "Deep Learning in a Generalized HJM-type Framework Through Arbitrage-Free Regularization," *arXiv preprint arXiv:1710.05114*, 2017.