Research Statement

Ta-Cheng Huang

National University of Singapore

November 2020

Below I summarize my research projects, grouped by the fields.

1 Microeconometrics

1.1 Unobserved Heterogeneous Treatment Effects

In microeconometrics, unobserved heterogeneity in treatment effects has been the centerpiece in literature for decades. Here unobserved heterogeneity refers that the treatment effects vary among otherwise observationally identical people. For example, when evaluating the impact of a job training program on earnings, we may find that treatment effects are different between two identical job seekers, who have the same covariate values that econometricians can observe. Therefore, the difference in treatment effects is attributed to some unobserved characteristics, for example, working attitude.

One main implication of unobserved heterogeneity in treatment effects is that the summary treatment effects that aggregate over different parts of the population will differ. When treatments self-selected, which is very common in economics, Imbens and Angrist (1994) show that we can only identify the compliers' treatment effects. Therefore, even if the treatment is effective for the compliers, it might not be so for others with the same covariate values.

From the perspective of policy/program evaluation, having a test for unobserved heterogeneity in treatment effects is of great importance. We can often estimate the costs when we conduct a cost-benefit analysis for a treatment or a policy. However, it is challenging to assess the benefits because we do not know whether the treatment effect estimates are valid for a larger subpopulation. In the working paper with Yu-Chin Hsu and Haiqing Xu titled "Testing for unobserved heterogeneous treatment effects with observational data," we propose a nonparametric test for unobserved heterogeneity in treatment effects when the treatment is self-selected. This test can answer the following

policy-relevant research question: "Whether the measured treatment effects are not only valid for the compliers but also for others with the same covariate values?"

1.2 Regression Kink Design

Regression Kink design is useful to estimate causal effects when a policy variable has kinks or discontinuities in the first derivative. For example, the causal effects of unemployment insurance benefit on unemployment duration are of great interest from a policy perspective. In the United States, a compensated unemployed individual receives a weekly benefit determined as a fraction of his highest-earning quarter in the base period up to a fixed maximum amount. Here, the unemployment benefit is a function of past earnings, and there is a kink at the maximum amount of the benefit. Card *et al.* (2015) show that we can identify the average treatment effect on treated, which can be roughly interpreted as the treatment effect for individuals whose benefit is close to the maximum amount.

My work in progress, "Nonparametric tests for treatment effect heterogeneity under Regression Kink design," is the first to propose tests for treatment effect heterogeneity under the Regression Kink design. The proposed tests study the following three policy-relevant questions: (i) whether a policy treatment is beneficial for at least some subpopulations defined by covariate values; (ii) does the treatment have any impact on at least some subpopulations; and (iii) are the treatment effects heterogeneous across subpopulations.

In my another work in progress, "Monotonicity test for treatment effects on treated under Regression Kink Design," I propose nonparametric monotonicity tests for treatment effects on treated under Regression Kink design. The tests allow us to test whether the treatment effects are monotonically correlated with conditioning covariates of interest. All these tests are applied to study the effect of unemployment insurance benefits on unemployment durations in Louisiana between 1981 and 1982 by using data from the Continuous Wage and Benefit History Project.

2 Econometric Theory

2.1 Varying Coefficient Models

Varying coefficient models, also known as functional coefficient models, are appropriate for many economics applications, and in particular, when additive separability of covariates is unsuitable for the problem at hand. For example, in studies of marginal returns to education, some empirical studies find that the returns to schooling would rise for an individual with more advanced schooling, or the marginal returns to education are also plausibly increasing in work experience.

To nonparametrically estimate these coefficient functions, it is known that the selection of smoothing parameters, which is also called bandwidth, is of crucial importance. One difficulty in selecting smoothing parameters optimally in a varying coefficient regression model is that, there is more than one function to be estimated and coefficient functions have different degree of smoothness. In the published paper with Xirong Chen and Qi Li titled "An alternative bandwidth selection method for estimating functional coefficient models," we propose an alternative approach which assigns each coefficient function a different smoothing parameter and we choose the smoothing parameters simultaneously. Therefore, all coefficient functions can be estimated by selecting a smoothing parameter optimally.

2.2 Model Specification Testing

Model specification test is one of the most important topics in econometrics since the cost of a misspecified model is considerate in terms of biased estimation and misleading inferences (Hausman, 1978). Much work has been motivated on providing consistent tests of parametric and semiparametric null hypotheses against general non-parametric alternatives, especially in regression (Fan and Linton, 2003).

In the published paper with Hongjun Li and Zheng Li titled "A modified bootstrap of kernel-based specification test with heavy-tailed data," we provide a new resampling strategy to improve the finite sample performance of a nonparametric kernel-based specification test in the presence of heavy-tailed error terms. Based on the test statistic of Li and Wang (1998), we propose to generate the bootstrapped samples using a modified wild bootstrap. This new method matches all moments of the error terms if the error has a symmetric distribution and matches the first and all even moments when error distribution is asymmetric around zero. This new resampling method has better finite sample performance than the traditional one when the distribution of the error terms is symmetric and heavy-tailed.

3 Applied Econometrics

3.1 Finance

Estimating implied volatilities is an important intermediate step to extract the state price densities (SPDs), the moments of which provide information about market sentiment. For example, skewness gives us a general idea about whether investors believe prices are going up or down for the stock associated with the option. Another example is the tail of the SPDs. It tells us how likely people believe extreme events to happen. The challenges

in obtaining good predictions of implied volatility come from the fact that the available usable is scarce on daily basis. For example, on average we have only 8 call options prices for S&P 500 Index on a given day. Common ways to solve this issue rely on either simple parametric polynomial estimation or nonparametric kernel estimation after taking longer periods. While it would be preferable to use nonparametric estimation, there is much value in extracting the densities daily, which poses a problem since kernel estimators have high volume data requirements.

In the published paper with Pablo Crespo titled "Implied volatility estimation via ℓ_1 trend filtering," we propose ℓ_1 trend filtering which is introduced by Tibshirani *et al.* (2014) to estimate implied volatility. ℓ_1 trend filtering is a variation on Hodrick-Prescott filtering which substitutes a sum of absolute values (i.e., an ℓ_1 norm) for the sum of squares used in Hodrick-Prescott filtering to penalize variations in the discrete derivatives. When taking our estimation procedure to real daily data against the benchmark model used in practice, ℓ_1 trend filtering consistently generates better predictions. This paper fills a gap in the literature by providing a simple methodology which can access feasible, flexible form empirical extraction of state price densities in conditions in which datasets are relatively small.

3.2 Synthetic Control

Synthetic control is an important tool in the set of methodologies for estimating treatment effects. It is, however, dependent on the assumption of trend stationarity. In the working paper with Pablo Crespo titled "Estimating Causal Effects for Cointegrated Nonstationary Series: An Augmented Synthetic Control Approach," We propose an alternative approach based on modern techniques for automatic forecasting when candidate controls are cointegrated to relax the assumption.

In the presence of non-stationary cointegrated series, Monte Carlo simulations show that this method is more robust than synthetic control and can identify treatment effects in a variety of forms. An empirical application reexamines the work of Abadie and Gardeazabal (2003), demonstrating the method's ability to replicate their results.

4 Field Experiments

4.1 Health Economics

Most of the RCTs on weight loss are goal-based, i.e., participants receive rewards if they reach a goal. Goal-based interventions require participants to invest upfront for a later, uncertain reward. However, risk-averse people may be unwilling to invest heavily for an

uncertain reward, especially if the goal is challenging. Another problem with goal-based interventions could be the lack of a structure for participants to follow. This is important because people joining the program with little to no knowledge on how to approach fat loss, more so since they are already overweight, may have difficulty identifying and executing the means to achieving the goal.

To deal with the challenges, we designed a subsidy on health-improving activities. Our subsidy provides some structure, helping these people by signalling what they should be spending and working on to reach the goal. Moreover, instead of having participants invest upfront for cash at the end if they reach their goal, we intend to provide the cash upfront in the form of a conditional subsidy to help them reach this goal. We have conducted a pilot study with 100 participants and got some promising results. We plan to run a larger scale field experiment with 600 participants to establish the statistical significance and study the mechanism in early 2021.

4.2 Labor Economics

Helping low-performing employees has become more critical due to strong job protection policies. In the work in progress with Hua Chen and Noah Lim titled "Helping Weaker Sales Agents via Temporary Transfers: A Field Experiment Study," we propose a novel tool – temporary transfers – for the retailing business, which has multiple stores with different traffic levels.

We designed two types of transfers: upward and downward transfers. Upward transfers move low-performing salespeople to a store with heavier traffic, while downward transfers shift them to a lighter traffic store. We hypothesized that upwardly transferred salespeople can improve the efficiency of converting a walk-in to a customer. Meanwhile, transferred salespeople can earn more commissions, which will serve as a new reference income when they return to the previous store. On the other hand, the downwardly transferred salespeople will have longer interaction time to practice sales skills, and lower expected income will motivate them. We plan to run a field experiment with 128 salespeople in early 2021.

5 Future Plans

In the future, I intend to continue my research in the area of statistical inferences in causal effect models with a special focus on endogenous treatments. As for the empirical side, I plan to investigate the impacts of policy interventions, especially in the presence of self-selected treatment choices in healthcare, labor economics, and other similar applications.

Last, I am also open to investigating various applied economic questions for which strong econometric and data skills, in conjunction with rigorous analyses and careful narratives, are valued.

References

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, **93** (1), 113–132.

Card, D., Lee, D. S., Pei, Z. and Weber, A. (2015). Inference on causal effects in a generalized regression kink design. *Econometrica*, **83** (6), 2453–2483.

Fan, Y. and Linton, O. (2003). Some higher-order theory for a consistent non-parametric model specification test. *Journal of Statistical Planning and Inference*, **109** (1-2), 125–154.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, **46** (6), 1251–1271.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** (2), 467–475.

Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. **87** (1), 145–165.

Tibshirani, R. J. *et al.* (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, **42** (1), 285–323.