Research Statement

Zheng Wang

My research interest lies within the fields of applied econometrics and applied microeconomics. I develop and use appropriate causal inference tools to understand issues related to education, gender, discrimination, and crimes. In my current research, I have a strong focus on networks and peer effects, where I aim to both improve current econometric methods and use novel data to answer interesting empirical questions.

Job Market Paper and Future Extensions

My job market paper, "**The linking effect: causal identification and estimation of the effect of peer relationship**", develops a **causal identification strategy** in the potential outcome framework to tackle the network endogeneity problem in the empirical study of peer effect. I achieve this by closely examining the meaning of "peer effect" in Rubin's potential outcome framework, which allows me to connect the network peer effect literature with the literature of multiple causal inference. By combining results from multiple causal inference and statistical network analysis, I prove that only three assumptions are necessary for causal identification. These assumptions are conditional independence of network links, no single-link confounder, and positivity, which I argue are easy to satisfy for many types of network data and are also implicitly assumed in existing econometric papers. Compared to existing approaches, my identification strategy does not rely on any specific network formation or outcome models, significantly improving its applicability. The main identification idea is that propensity scores of the unobserved confounders can be non-parametrically identified and estimated from the distribution of network links, which naturally points to the use of **propensity score-based estimators**. Unlike traditional propensity score estimation procedures where the probability of treatment is regressed on a set of observed pre-treatment variables, here, the propensity scores of the unobserved confounders are estimated using only the observed network links. These estimators are flexible and easy to implement with existing statistical packages.

The proposed identification and estimation procedure could be used to study various endogenously formed social and economic networks, such as friendship networks, buyer-supplier networks, banking networks, etc. As an **empirical application** of the methodology, I estimate the causal effect of high school friendships on students' bachelor's degree attainment. While previous literature finds that being exposed to more high-achieving boys in high school makes girls less likely to obtain a bachelor's degree, I show that this is not true when these high-achieving boys are considered friends by the girls. In fact, one additional high-achieving male friend increases the probability that a female student graduates from college by 3 p.p. Further analysis suggests that this positive influence is not a result of increased academic ability but rather a significant confidence boost. These results suggest that fostering friendships between

female students and male high-achievers could help offset the negative impact on female students from the presence of male high-achievers.

As with any new methodology, there are many ways to **extend my job market paper**. Concretely, I plan to work on two extensions that could benefit the empirical study of link effects. First, **a variation of the Oster test** could be developed to test whether unaccounted confounding is likely to cause unstable estimated peer effect parameters. This is important because, although all unobserved confounding can be identified and inferred, this is an infinite population property. With a finite sample, there could be some leftover confounding. An adapted Oster test could inform empirical researchers on whether the leftover confounding will likely cause the estimated effect to disappear or flip signs. Second, I plan to adapt **heterogeneous treatment effects techniques to identify the most important heterogeneities that determine peer influence**. In the friendship example, this means developing a machine learning procedure to identify which characteristics of friendships are the most influential to the outcome of interest. This work is motivated by the fact that when studying peer influence, the ultimate goal is to understand the mechanisms underlying the estimated effects and use that to design policies that maximise welfare. As with using machine learning to study traditional treatment effect heterogeneity, the procedure could solve the p-hacking problem. Finally, I am also interested in **collaborating with researchers in the fields of trade and finance** to apply my empirical strategy to the study of **trade networks and banking networks.**

Other Ongoing Projects

Econometrics

In my project, joint with Yann Bramoullé and Pierre-Philippe Combes, "**The perils of pairwise peer effect**", we show that using pairwise regressions to study peer effect can lead to spurious estimation results. The use of pairwise regressions as a way to avoid the infamous reflection problem has become increasingly common, noticeably after the influential paper by Bayer, Ross, and Topa (2008)[1]. We show that this approach is troubled with both identification and inference issues. Indeed, our theoretical result, which is also verified by simulations, suggests that pairwise regressions can give a null estimate when the true data-generating model is a linear-in-means model with a non-zero endogenous peer effect parameter. Moreover, we show that clustered standard errors routinely used in pairwise regressions cannot properly account for the dependence structure in the error terms, thereby significantly underestimating the uncertainty in the parameter estimates.

[1] Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. "Place of work and place of residence: Informal hiring networks and labor market outcomes." Journal of Political Economy 116, no. 6 (2008): 1150-1196.

Applied Microeconomics

My paper "**Can women be more competitive than men? Evidence from Duolingo**" examines the differential effects of competition difficulty on men and women when the competition does not involve face-to-face confrontation, and effort is the only determinant of the final ranking. I utilise a natural experiment from the weekly Duolingo Leaderboard competition, where language learners are randomly allocated to groups of 30 people to compete on the number of language lessons completed during a week[2]. These lessons are standardised and reward around the same points upon completion. While people within the same competition group could track each other's performance, they have no means of communicating with one another. With web-scraped data from over 25,000 Duolingo users, I find that when faced with more challenging competitors with a track record of higher performance, women tend to double down and put more effort into the competition. At the same time, men don't seem to respond the same way. These findings are in accordance with recent literature on grit which suggests that females do not shy away from competitions that emphasise efforts.

In my paper with Dalila Figueiredo, "**The Effect of Familiarity on Test Performance and its Implications on Fairness**", we design and conduct an experiment to examine whether students' familiarity with test topics used in the IELTS[3] exam affects their test performance and whether familiarity disparities across groups lead to unfair examinations. In our experiment, we randomly assign reading materials to participants before they are asked to complete an IELTS writing task. This allows us to evaluate whether topic familiarity causally affects test performance. Through a survey, we obtain the distribution of topic familiarity and **examine whether the test is unfair towards any group in three categories: gender, culture, and socioeconomic status**. We have recently conducted a pilot with Italian university students. The pilot results suggest that our reading materials successfully increase students' familiarity with their assigned topic and that their writing performance, as rated by a certified IELTS examiner, is improved.

Finally, in my project with Christian Dustmann, Rasmus Landersø, and Mikkel Mertz, "**Victimization and Crime Overlap**", we use Danish administrative data on the universe of victims and study whether and how different aspects of victimisation causally affect the future criminal activities of the victims. This paper explores the reasons behind an interesting observation that there is a significant overlap between victims and criminals. We will use a **deconfounding procedure similar to the one used in my job market paper** to account for the endogeneity issue that might be causing part of this overlapping pattern.

[2] Duolingo is a free online foreign-language learning platform with over 300 million users. For more information on Duolingo and the leaderboard competition, you can visit their website at https://www.duolingo.com/.
[3] IELTS, or International English Language Testing System, is a standardised English test many universities worldwide use for admission purposes.