

# Likun Zhang

📍 Beijing, China    ✉ zhanglk6@ruc.edu.cn    ☎ +86 13711012799    in zhanglk6    🌐 zhanglk6

## Education

---


<b>Renmin University of China</b> (RUC), Beijing, China <i>PhD in Statistics, Institute of Statistics and Big Data</i>	<i>Aug 2021 – June 2026</i> <i>(expected)</i>
<b>University of California, Berkeley</b> (UCB), CA, United States <i>Statistics Exchange Program</i>	<i>Aug 2019 – Dec 2019</i>
<b>Sun Yat-sen University</b> (SYSU), Guangdong, China <i>B.S. in Statistics, School of Mathematics</i>	<i>Aug 2017 – June 2021</i>

- GPA: 4.0/4.0 (Ranking: 1/79)
- 2018 National Scholarship for academic excellence and whole person development (Ranking: 1/231)

## Publications

---

<b>Interaction Tests With Covariate-Adaptive Randomization</b> <i>Likun Zhang, Wei Ma</i>	Jan 2025
--	----------

- Published in Statistical Analysis and Data Mining: The ASA Data Science Journal; [doi.org/10.1002/sam.70003](https://doi.org/10.1002/sam.70003) 

## Research Projects

---


<b>Covariance-Driven Regression Trees: Reducing Overfitting in CART</b> <i>Likun Zhang, Wei Ma</i> <i>Under review</i>
--

- Propose a covariance-driven splitting criterion for regression trees (CovRT). This method is more robust to overfitting than the empirical risk minimization criterion used in CART, as it produces more balanced and stable splits and more effectively identify covariates with true signals
- Establish the consistency of CovRT and proves that its predictive accuracy is comparable to that of CART in high-dimensional settings
- Through simulations and empirical studies, we show that CovRT not only has smaller generalization error but also improves interpretability over CART. In particular, it reduces prediction risk by approximately 24% on the Boston Housing dataset using a much shallower tree

<b>Efficient Interaction Analysis in Randomized Controlled Trials</b> <i>Likun Zhang, Wei Ma</i> <i>Major revision at Biometrics</i>
--

- Introduce a model-free framework for interaction analysis in randomized controlled trials, advocating the adoption of a clearly defined target parameter for interaction and thereby avoiding assumptions about the functional form of the data-generating mechanism
- Study interaction analysis under covariate-adaptive randomization, including simple and stratified randomization, and minimization methods as special cases
- Establish that the usual method for interaction analysis can either exaggerate or understate uncertainty, leading to asymptotically conservative or anti-conservative results
- Compute the semiparametric efficiency bound and introduce a novel semiparametric efficient method, with nonparametric and machine learning techniques for adjusting baseline covariate information

---

<b>Interaction Tests With Covariate-Adaptive Randomization</b> <i>Likun Zhang, Wei Ma</i> <i>Published in Statistical Analysis and Data Mining: The ASA Data Science Journal</i>	<a href="https://doi.org/10.1002/sam.70003">doi.org/10.1002/sam.70003</a> 
--	---

- Investigate usual interaction tests and observe that they tend to be conservative
- Propose modifications to the usual tests to obtain corresponding valid tests with limiting rejection proba-

bilities equal to the nominal level

- Introduce a novel class of stratified-adjusted interaction tests that are simple, more powerful than the usual and modified tests, and broadly applicable to most covariate-adaptive randomization methods
- Encompass two types of interaction tests: one involving stratification covariates and the other involving additional covariates that are not used for randomization

## Experience

---

### **Age-Related Variation and Associated Factors of Intrinsic Capacity Across Adulthood**

*Beijing*

*Dec 2021 – June 2024*

*Beijing Institute of Geriatrics, Beijing Hospital*

- Characterized age-related variation in intrinsic capacity (IC) and identified factors associated with IC in a nationally representative Chinese population
- This cross-sectional study included 17,086 participants aged 25-89 years from seven major geographic regions in the PENG ZU cohort. IC was assessed across five domains (locomotion, cognition, vitality, psychology and sensory) using self-reported data and physical performance tests

### **Quantitative Investment Research With Machine Learning**

*Guangdong*

*Southern China Center For Statistical Science, SYSU*

*Feb 2019 – Dec 2019*

- Simulated stock price prediction strategies
- Optimized trading results with methods like the support vector machine (SVM), vortex indicator, TD sequence, Baum–Welch algorithm, Gaussian mixture distribution, and hidden Markov models

## Technologies

---

**Languages:** R, Python, C++, Java, C#, SQL