

**Kaizheng Wang**  
**Department of Industrial Engineering & Operations Research**  
**Columbia University**



*Model selection amidst distribution shifts*

**Wednesday, February 14th, 2024**  
**11:50 AM**

**96 Frelinghuysen Road, CoRE Building, Room 431**

**Zoom Meeting: Meeting ID: 969 0606 4706**  
**Password: 745339**

<https://rutgers.zoom.us/j/96906064706?pwd=ZklvbExpRVBJQ3c5dUhhYTFuR2ZrZz09>

**Light refreshments will be served in Hill 452, 11:15am**

**Abstract:** In this talk, I will present two vignettes on model selection amidst distribution shifts, where estimating the test error becomes challenging. The first vignette delves into kernel ridge regression under covariate shift, aimed at minimizing mean squared error over a target distribution using unlabeled data from that and labeled data with a different covariate distribution. Our analysis shows the adaptivity of hold-out validation with pseudo-labels, highlighting the benefit of training an undersmoothed imputation model to fill the missing target labels. The second vignette focuses on model selection in changing environments by synthesizing data across both present and historical epochs. We construct an adaptive window to estimate the generalization error while accounting for unknown temporal distribution shifts. Based on that, we develop a method for pairwise model comparisons. It is further integrated into a single-elimination tournament framework that enables near-optimal model selection from a pool of candidates.

**Bio:** Kaizheng Wang is an assistant professor of Industrial Engineering and Operations Research, and a member of the Data Science Institute at Columbia University. He works at the intersection of statistics, machine learning, and optimization. He received the SIAM Activity Group on Imaging Science Best Paper Prize in 2024, and the Second Place Award in the 2023 INFORMS Data Mining Challenge. He obtained his Ph.D. from Princeton University in 2020 and B.S. from Peking University in 2015.

