

A GMM approach in coupling internal data and external summary information with heterogeneous data populations

王磊 南开大学

Because of advances in data collection and storage, statistical analysis in modern scientific research and practice now has opportunities to utilize external information such as summary statistics from similar studies. A likelihood approach based on a parametric model assumption has been developed in the literature to utilize external summary information when the populations for external data and the main internal data are assumed to be the same. In this article we instead consider the generalized estimation equation (GEE) approach for statistical inference, which is semiparametric or nonparametric, and show how to utilize external summary information even when internal and external data populations are not the same. Our approach is coupling the internal data and external summary information to form additional estimation equations, and then applying the generalized method of moments (GMM). We show that the proposed GMM estimator is asymptotically normal and, under some conditions, is more efficient than the GEE estimator without using external summary information. Estimators of asymptotic covariance matrix of the GMM estimators are also proposed. Simulation results are obtained to confirm our theory and to quantify the improvements from utilizing external data. An example is also included for illustration.

报告人简介：王磊，南开大学统计与数据科学学院副研究员，博士生导师。研究方向是复杂数据分析和统计学习，已在 *Biometrika*、*Science China Mathematics*、*Bernoulli*、*Statistica Sinica* 等统计学杂志发表学术论文 60 多篇，主持 3 项国家自然科学基金和 1 项天津市自然科学基金项目。