Theoretical Statistics and Mathematics Unit, Kolkata
INDIAN STATISTICAL INSTITUTE

SEMINAR

Date: December 02, 2025
Time: 04:30 PM

VENUE:

L- Infinity

(5th Floor, A.N. Kolmogorov Bhavan), ISI Kolkata

TITLE:

Cross-Sectional Regression with Cluster Dependence

SPEAKER:

Subhodeep Dey

Stat-Math Unit, ISI Kolkata

ABSTRACT:

The interdependence of model error in linear regression can cause the least squares estimator (LSE) to become inconsistent, and the CLT may no longer hold, complicating hypothesis testing for regression parameters. However, when data are structured in clusters or blocks, valid inference may still be attainable using robust standard errors. We have considered cross-sectional dependence within clusters and independence between clusters, both when the cluster size is finite and when it is infinite. In this talk, we will describe some problems under the above setup.

First, we propose a new estimator that is consistent and asymptotically normal under such cross-sectional dependence. The estimator performs well even with some dominating clusters, and we derive appropriate standard errors. A detailed simulation study examines the efficacy of the proposed estimator.

Second, we propose an alternative estimator that is consistent and asymptotically normal under such cross-sectional dependence and for infinite cluster sizes. Here, we can also allow the regression parameter to vary across the clusters. We have proposed an appropriate standard error based on super blocks for different testing purposes.

Third, we propose an estimator for the least absolute deviation regression parameter for cross-sectional data with cluster dependence, which is shown to be consistent and asymptotically normal. In this context, we want to propose a consistent estimate for the covariance matrix of the asymptotic distribution and propose a test procedure based on our proposed estimator.

Fourth, we show that failure to control for within-cluster error correlation can lead to both underestimated and overestimated standard errors, producing correspondingly misleading confidence intervals, large/small t-statistics, and low/high p-values. We also demonstrate that the famous Moulton factor is often misunderstood.

These are based on joint work with Prof. Gopal Krishna Basak and Prof. Samarjit Das.

ALL ARE CORDIALLY INVITED