Theoretical Statistics and Mathematics Unit

Indian Statistical Institute

203 B.T. Road, Kolkata - 700108

Thesis Pre-Submission Seminar

Date: November 16, 2026

Time: 04:15 PM

VENUE:

L-Infinity, 5th Floor, A.N. Kolmogorov Bhavan, ISI Kolkata

SPEAKER:

Subhodeep Dey

Stat-Math Unit, ISI Kolkata

TITLE:

Inference Beyond Pooled Ordinary Least Squares for Clustered Data

ABSTRACT:

We study linear regression with clustered cross-sectional data, where dependence within clusters may be strong, semi-strong or weak, and when cluster sizes may be finite or grow to infinity. In this talk, we present the four main contributions of the thesis.

First, we revisit the asymptotic properties of the pooled ordinary least squares (POLS) estimator under different regimes of within-cluster dependence. POLS is shown to be inconsistent in the presence of dominant clusters under strong within-cluster dependence. We also derive regime-specific conditions for the asymptotic normality of POLS, the consistency of the cluster-robust variance estimator and the validity of the associated Wald test.

Second, to overcome this problem, we propose a simple estimator based on cluster averages, which remains consistent for any kind of within-cluster dependence and any cluster sizes, including dominant clusters. We establish its asymptotic normality, the consistency of a heteroskedasticity-robust variance estimator and the validity of a Wald-type test based on it. The proposed estimator is shown to be more efficient than POLS in many practical situations. Also, it remains consistent under classical measurement error in the regressors, which avoids the search for valid instruments. Simulation studies support the theoretical findings, and an application to the Vietnam Living Standards Survey indicates that the proposed estimator yields a better fit.

Third, we allow the regression coefficients to vary across clusters within a random coefficients framework. We propose an alternative estimator based on cluster specific estimates. We show that it is consistent and asymptotically normal, irrespective of cluster sizes and the strength of within-cluster dependence. In contrast, POLS is inconsistent in the random coefficients model for any dependence regimes when a few clusters dominate the sample. We further propose a consistent test of parameter stability across higher-level groupings of clusters, referred to as superblocks. Simulation studies and an application to Engel curves from India's Household Consumption Expenditure Survey (HCES) 2022--23 support the proposed methods.

Finally, since the regime of within-cluster dependence is unknown in practice, we propose a simple test to identify it. The test exploits the variation in cluster sizes through an auxiliary regression of normalized squared cluster scores on cluster size. Simulation studies and an application to the HCES 2022--23 data reveal the efficacy of the proposed test..

ALL ARE CORDIALLY INVITED

Click here for More Seminar Information

Stat-Math Unit, ISI Kolkata