THE UNIVERSITY OF CHICAGO

Department of Statistics
STATISTICS COLLOQUIUM

ZHENG (TRACY) KE

Department of Operations Research and Financial Engineering

Princeton University

Covariance Assisted Screening and Estimation

MONDAY, January 13, 2014 at 4:00 PM

133 Eckhart Hall, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110

ABSTRACT

We consider the problem of variable selection in the very challenging regime where the signals (non-zero regression coefficients) are rare and weak and columns of the design matrix are heavily correlated. We demonstrate that in the presence of Rare/Weak signals, many classical methods and ambitious contemporary algorithms face pitfalls. The situation is worsen in the presence of heavy correlations among design variables.

We propose a new variable selection approach which we call Covariance Assisted Screen- ing and Estimation (CASE). CASE is a two-stage multivariate Screen and Clean algorithm. CASE has two layers of innovations. In the first layer, we alleviate the heavy correlations of the design variables by linear filtering and use the post-filtering model to construct a sparse graph. In the second layer, we use the sparse graph to guide both the screening and cleaning.

We explain how CASE overcomes the well-known computational hurdle of multivariate screen- ing. We also explain how CASE overcomes the so-called challenge of "signal cancellation", so its success is not tied to strong signals or any types of incoherence/irrepresentable conditions.

We set up a theoretical framework where we show CASE obtains the optimal rate of convergence in terms of Hamming errors. We have successfully applied CASE to the long- memory time series and a change-point model, where the optimality is further investigated with the so-called notion of phase diagram.

This is in collaboration with Jianqing Fan and Jinshun Jin. An arXiv preprint about the work could be found at http://arxiv.org/abs/1205.4645.