## Model Selection of Correlation Structure for Clustered Data

Annie Qu Dept. of Stat, University of Illinois at Urbana-Champaign 725 S. Wright St. Champaign, IL 61820 anniequ@illinois.edu

## Abstract

ABSTRACT Model selection of correlation structure is a challenging problem because it involves a higher order of moments than model selection of covariates only. However, the correct specification of the correlation structure plays an important role in improving estimation efficiency for clustered data. Our strategy is to approximate the empirical estimator of the correlation matrix as closely as possible using a pool of basis matrices candidates, and penalize models involving too many basis matrices. The proposed method has the advantages of not requiring the likelihood function and of being computationally efficient. In addition, it can identify complex correlation structures through the group-wise selection strategy from a large number of basis matrices, and is applicable for both continuous and discrete response data. In theory, we show that the proposed method enjoys the oracle property of selecting the true correlation structure consistently and estimating the correlation parameters with the same asymptotic normal distribution as if the true structure is known. Our simulation studies and data example show that the proposed method works effectively to select the true structure. This is joint work with Jianhui Zhou of University of Virginia.