

**SEMIPARAMETRIC PROBIT MODEL FOR HIGHDIMENSIONAL  
CLUSTERED DATA**

DANIEL REYES RAGUINDIN

Thesis submitted to the faculty of the  
School of Statistics, University of the Philippines, Diliman Quezon City  
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE in STATISTICS

May 2016

## ABSTRACT

We propose a semiparametric probit model for high dimensional clustered data. The model allows flexibility in the structure to account for lost information in the process of dimension reduction. Principal components are postulated to have nonparametric effect on the dichotomous response, mitigating the lost information due to the selection of just few principal components. On the other hand, the parametric part takes advantage of inherent homogeneity within clusters, hence, a constant random intercept term accounts for clustering. Simulation studies illustrated the advantages of the proposed model over the ordinary probit model in low dimensional cases. It also provides high predictive ability in high dimensional cases especially when the distribution of the response to the two categories is balance even in the presence of misspecification error.

*Keywords:* probit model; dimension reduction; nonparametric regression; local scoring algorithm.