

# Sieve Maximum Likelihood Estimation for Regression Models With Covariates Missing at Random

by

*Qingxia Chen*

*Department of Biostatistics, Vanderbilt University, Nashville, TN 37232*

`cindy.chen@vanderbilt.edu`

## Abstract

Missing covariates are common in regression problems. We propose a new semiparametric method based on a fully nonparametric distribution for the missing covariates that are assumed to be missing at random. The method of sieve maximum likelihood estimation is used to obtain the estimators of the regression coefficients. These estimators are shown to be consistent and asymptotically normal with their asymptotic covariance matrix that achieves the semiparametric efficiency bound. A bootstrap approach is used to estimate the asymptotic covariance matrix. Some practical modeling approaches for high-dimensional covariates are proposed. Extensive simulation studies are conducted to examine the finite-sample properties of the estimates, and a real data set from a liver cancer clinical trial is analyzed using the proposed method.

This is joint work with Donglin Zeng and Joseph G. Ibrahim (University of North Carolina).