Abstract

The error-in-covariates problem has received great attention among researchers who study semiparametric and nonparametric inference for regression models over the past two decades. Without correction for the measurement error in covariates, the resultant estimators usually contain bias. To account for measurement error, much research have been done in mean regression (Liang et al., 1999; Fuller, 2009; Carroll et al., 2006) and quantile regression (He and Liang, 2000; Hardle et al., 2000; Wei and Carroll, 2009). In contrast, there is little research in mode regression and this motivates us to propose a semiparametric method to address this error-in-covariates problem in Chapters 2, 4.

Chapter 2 considers estimating the mode of a response given an error-prone covariate by assuming that the mode of Y given X is a linear function of X. It is first shown that ignoring measurement error typically leads to inconsistent inference for the mode of the response given the true covariate, as well as misleading inference for regression coefficients in the conditional mode model. To account for measurement error, the Monte Carlo corrected score method (Novick and Stefanski, 2002) is employed to numerically obtain an unbiased score function based on which the regression coefficients is estimated consistently. To relax the normality assumption on measurement error the first method requires, another corrected kernel method is proposed. In this method, an objective function constructed by deconvoluting kernels is maximized to obtain consistent estimators of the regression coefficients. Besides rigorous investigation on large sample properties of the new estimators, we study their finite sample performance via extensive simulation experiments, and find that

the proposed methods substantially outperform a naive inference method that ignores measurement error.

In Chapter 3, we assume that the mode of Y is a linear function of a covariate X and it also depends on another covariate T in an unspecified functional form. This leads to a partially linear model for the conditional mode. We employ B-splines to approximate the unspecified function that relates Y and T. To estimate the covariate effects explaining the association between Y and X, and at the same time, estimate the unspecified function linking Y and T, we develop two methods for inferring these two parts of the partially linear model for the conditional mode. Then, the tuning parameters selection is illustrated and a simulation study is designed to show the performance of two proposed methods. Chapter 4 considers estimating the mode of a response in partially linear models when the aforementioned X is error-prone. To account for measurement error, we incorporate the corrected kernel method proposed in Chapter 2 and the proposed estimation methods in Chapter 4 to infer the parametric part and nonparametric part of the conditional mode accounting for measurement error in X. Results from simulation studies suggest that the proposed method substantially outperform a naive inference method that ignores measurement error. Besides error-prone covariates, in Chapter 5, we consider a scenario where the response is contaminated by Berkson measurement error. In particular, we tackle the regression analysis for a pooled continuous response with error-prone covariates. Finally, Chapter 6 discusses future research in my dissertation.