Report 1 for ST810 Measurement Error

Na Zhang

09-22-2011

References


1 Introduction and Motivation

Liang and Li (2009) focus on variable selection for partially linear models when the covariates are measured with additive errors. They proposed two classes of variable selection procedures, which are penalized least squares and penalized quantile regression using the nonconvex penalized principle. Since measurement error data are very common in many fields, including engineering, economics, physics, biology, biomedical sciences, and epidemiology, it is necessary to consider models with measurement error covariates. Moreover, in linear regression models containing several covariates, some of which are measured with error and some are not, the ordinary least squares method will result in inconsistent estimate for the regression coefficients without considering the measurement errors, because the loss function contains the error-prone covariates and the expected value of the corresponding estimating function does not equal zero. Then, the inconsistency nullifies the theoretical property of the ordinary penalized least squares estimate without considering the measurement error.
Thus, in the presence of measurement error, the existing penalized least squares variable selection procedures ignoring measurement error may not work properly. Therefore, Liang and Li (2009) propose two new variable selection procedures for partial linear measurement error models to deal with the above mentioned concerns.

2 Penalized Least Squared Method

Suppose \((W_i, Z_i, Y_i), i = 1, \ldots, n\) is a random sample from the following partial linear measurement error model (PLMeM),

\[
\begin{align*}
Y &= X^T \beta + \nu(Z) + \epsilon, \\
W &= X + U,
\end{align*}
\]  

(2.1)

where \(Z\) is a univariate observed error-free covariate, \(X\) is a \(d\)-dimension vector of unobserved latent covariates, which is measured in an error-prone way, \(\epsilon\) is the model error with \(E(\epsilon|X,Z) = 0\), \(U\) is the measurement error with mean zero and covariance matrix \(\Sigma_{uu}\). From the second part of (2.1), this is a classical measurement error model and \(W\) is the observed surrogate of \(X\).

When \(\Sigma_{uu}\) is known, the penalized least squared function based on partial residuals is defined as

\[
L_p(\Sigma_{uu}, \beta) = \frac{1}{2} \sum_{i=1}^{n} [Y_i - \hat{m}_y(Z_i) - W_i - \hat{m}_w(Z_i) \beta]^2 - \frac{n}{2} \beta^T \Sigma_{uu} \beta + n \sum_{j=1}^{d} p_{\lambda_j n}(|\beta_j|),
\]  

(2.2)

where \(p_{\lambda_j n}(\cdot)\) is a penalty function with tuning parameter \(\lambda_j n\), \(m_w(Z) = E(W|Z)\), \(m_y(Z) = E(Y|Z)\). The smoothly clipped absolute deviation (SCAD) penalty is preferred here. Liang and Li (2009) further proved the sampling property and oracle properties of the resulting penalized least squared estimate. The estimate has an asymptotic normal distribution. This algorithm is computationally easier.
3 Penalized Quantile Regression

In the presence of outliers, this penalized quantile regression techniques based on the orthogonal regression provide a robust estimator to outliers. The PLMeM model is still considered here. The orthogonal regression has been used to correct estimation bias due to measurement error of the least squares estimate of regression coefficients in linear measurement error models. The estimate also has an asymptotic normal distribution.

4 Simulation Results

In the simulation study, Liang and Li (2009) compared the procedures with respect to estimation accuracy and model complexity in two examples. In the first example, SCAD procedure performs the best in terms of estimation accuracy. Moreover, the SCAD and the $L_{0.5}$ procedures that consider measurement error outperform other procedures in terms of model complexity. In the second example, the penalized quantile regression procedure outperforms the corresponding penalized least squares procedure in terms of median of squared error (MedSE). Furthermore, the SCAD and $L_{0.5}$ procedures have almost the same average number of zero coefficients in the selected models. Both the SCAD and $L_{0.5}$ procedures perform better than the Hard procedure in this example.

5 Conclusions and Discussions

In this paper, Liang and Li (2009) proposed two classes of variable selection methods, the penalized least squares and the penalized quantile regression with nonconvex penalty, for partially linear models with error-prone linear covariates and possibly contaminated response variable. The sampling properties of the proposed methods are studied. The proposed methods perform well with moderate sample sizes. Thus, further research is needed on variable selection for large dimension and small sample size data with measurement error.