A SELECTION PROBLEM IN MEASUREMENT ERROR MODELS

by

Shanti S. Gupta and Xun Lin
Purdue University

Technical Report #97-18C

Department of Statistics
Purdue University
West Lafayette, IN USA

October 1997

A Selection Problem in Measurement Error Models*

by

Shanti S. Gupta and Xun Lin

Department of Statistics, Purdue University, West Lafayette, IN 47907

Department of Statistics, Purdue University, West Lafayette, IN 47907

Abstract

This paper deals with a selection problem for linear measurement error models. A selection procedure is constructed, and its corresponding asymptotic optimality is also investigated. It is shown that with the assumption of the existence of the α-th ($\alpha > 2$) moment, the expected risk of the proposed selection procedure converges to 0 with the rate of order $o(n^{-(\alpha/2-1)})$. It is further shown that when the moment generating functions of the corresponding variables exist, the expected risk of the proposed selection procedure converges with exponential order $O(e^{-c^* n})$ under mild conditions, where c^* is a positive constant.

AMS Classification: primary 62F07; secondary 62J99.

Keywords: asymptotically optimal; measurement error model; selection procedure; rate of convergence.

*This research was supported in part by US Army Research Office, Grant DAAH04-95-1-1065 at Purdue University.

1 Introduction

Measurement error models commonly begin with an underlying model where one or more of the independent variables are measured with error. The distinguishing feature of a measurement error problem is that we cannot observe those variables which are measured with error directly. The goal of measurement error modeling is to obtain understanding from the model. Attainment of this goal requires careful analysis.

This selection problem is from statistical consulting. When I (Xun Lin) was a statistical consultant in the summer of 1996, one of my clients came up with a problem which can be simplified as follows.

Suppose we have k treatments $\Pi_i, i = 1, \ldots, k$ and n observations from each treatment. For each treatment $\Pi_i, i = 1, \ldots, k$ and each observation $j = 1, \ldots, n$, we have the following model:

$$Y_{ij} = \beta_{0i} + \beta_{1i} X_{ij} + \epsilon_{ij}, \qquad W_{ij} = X_{ij} + U_{ij}, \tag{1}$$

where $\{(X_{ij}, U_{ij}, \epsilon_{ij}), 1 \le j \le n\}$ are independently but not necessarily identically distributed random vectors with means $(0, 0, 0)$ and variances $(\sigma_{xxi}, \sigma_{uui}, \sigma_{\epsilon\epsilon i})$. For $i = 1, \ldots, k$ and $j = 1, \ldots, n$, $(X_{ij}, U_{ij}, \epsilon_{ij})$ are independent to each other. But X_{ij} cannot be observed, instead we can only observe (W_{ij}, Y_{ij}). We assume that for each i, σ_{uui} is known and $\sigma_{xxi} > 0$.

A treatment Π_i is said to be the best if the associated slope parameter β_{1i} is the largest among the k slope parameters, otherwise the treatment is said to be nonbest. The goal of this selection problem is to select the best treatment from the k treatments.

Let $\Omega = \{\underset{\sim}{\beta_1} = (\beta_{11}, \beta_{12}, \ldots, \beta_{1k}) | \beta_{i1} \in R, i = 1, \ldots, k\}$ be the parameter space. Let $\underset{\sim}{a} = (a_1, \ldots, a_k)$ be an action, where $a_i = 0, 1,\ i = 1, \ldots, k$. When action $\underset{\sim}{a}$ is taken, $a_i = 1$ means that treatment Π_i is selected as the best treatment; otherwise $a_i = 0$ and Π_i is excluded as the nonbest. For $i = 1, \ldots, k$, let $W_i = (W_{i1}, \ldots, W_{in})$, $Y_i = (Y_{i1}, \ldots, Y_{in})$, $\underset{\sim}{X} = (X_1, \ldots, X_k)$, and $\underset{\sim}{Y} = (Y_1, \ldots, Y_k)$. Let χ be the sample space generated by $(\underset{\sim}{W}, \underset{\sim}{Y})$. Since the true order of $\beta_{11}, \ldots, \beta_{1k}$ is unknown, we denote $\beta_{1[1]} \le \beta_{1[2]} \le \ldots < \beta_{1[k]}$. For simplicity, we assume that $\beta_{1[k]} - \beta_{1[k-1]} = 2\delta > 0$.

A selection rule $d(\underset{\sim}{w}, \underset{\sim}{y}) = (d_1(\underset{\sim}{w}, \underset{\sim}{y}), \ldots, d_k(\underset{\sim}{w}, \underset{\sim}{y}))$ is a mapping defined on χ, where $d_i(\underset{\sim}{w}, \underset{\sim}{y})$ is the probability that given $\underset{\sim}{W} = \underset{\sim}{w}$ and $\underset{\sim}{Y} = \underset{\sim}{y}$, Π_i is selected as the best. Also, $\sum_{i=1}^{k} d_i(\underset{\sim}{w}, \underset{\sim}{y}) = 1$, for all $(\underset{\sim}{w}, \underset{\sim}{y}) \in \chi$.

We consider the following loss function:

$$L(\underset{\sim}{\beta_1}, \underset{\sim}{a}) = \begin{cases} 1, & \text{if a nonbest treatment is selected,} \\ 0, & \text{if the best treatment is selected.} \end{cases} \tag{2}$$

2 Formulation of the Selection Procedure

The population moments of (W_{ij}, Y_{ij}) satisfy

$$(\mu_{wi}, \mu_{yi}) = (0, \beta_{0i}), \tag{3}$$

and

$$(\sigma_{wwi}, \sigma_{wyi}, \sigma_{yyi}) = (\sigma_{xxi} + \sigma_{uui}, \beta_{1i}\sigma_{xxi}, {\beta_{1i}}^2\sigma_{xxi} + \sigma_{\epsilon\epsilon i}). \tag{4}$$

The sample means $(\bar{W}_i, \bar{Y}_i)$ and the sample covariates $(S_{wwi}, S_{wyi}, S_{yyi})$, where, for example,

$$\bar{W}_i = \frac{1}{n}\sum_{j=1}^{n} W_{ij}, \tag{5}$$

$$S_{wyi} = \frac{1}{n-1}\sum_{j=1}^{n}(W_{ij} - \bar{W}_i)(Y_{ij} - \bar{Y}_i), \tag{6}$$

will be the bases of our selection procedure.

We use estimators of the unknown parameters by replacing the unknown population moments with their sample estimators. For the quantities defined above to be proper estimators, $\hat{\sigma_{xxi}}$ must be positive. We have $\hat{\sigma_{xxi}} = S_{wwi} - \sigma_{uui}$ when $S_{wwi} - \sigma_{uui}$ is positive, otherwise we set $\hat{\sigma_{xxi}} = S_{yyi}^{-1} S_{wyi}^2$. We also have

$$\hat{\beta_{1i}} = \begin{cases} (S_{wwi} - \sigma_{uui})^{-1} S_{wyi}, & \text{if } S_{wwi} - \sigma_{uui} > 0, \\ S_{wyi}^{-1} S_{yyi}, & \text{otherwise.} \end{cases} \tag{7}$$

Since we take n samples from each treatment, our selection procedure will be $d_n(\underline{w}, \underline{y}) = (d_{1n}(\underline{w}, \underline{y}), d_{2n}(\underline{w}, \underline{y}), \ldots, d_{kn}(\underline{w}, \underline{y}))$, where

$$d_{in}(\underline{w}, \underline{y}) = \begin{cases} 1, & \text{if } \hat{\beta_{1i}} \text{ is the largest among the } k \text{ slope estimates,} \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

when $\underline{W} = \underline{w}$ and $\underline{Y} = \underline{y}$ are observed.

3 Performance of the Selection Procedures

In this section we study the performance of the selection procedures. We first analyze the expected risk of the proposed procedure.

Definition 1. A sequence of selection procedures $\{d_n(\underline{w},\underline{y})\}_{n=1}^{\infty}$ is said to be asymptotically optimal of order e_n if $E^{(\underline{W},\underline{Y})}L(\underline{\beta}, d_n(\underline{w},\underline{y})) = O(e_n)$, where e_n is a sequence of positive numbers such that $\lim_{n\to\infty} e_n = 0$.

Denote P_n to be the probability measure generated by the random observations $(\underline{W},\underline{Y})$, and for each $(\underline{w},\underline{y}) \in \chi$, let

$$i^* = \{i|\beta_{1i} = \max_{1\le j\le k} \beta_{1j} = \beta_{1[k]},\ i = 1,\dots,k\}, \tag{9}$$

and

$$i_n^* = \{i|\hat{\beta}_{1i} = \max_{1\le j\le k} \hat{\beta}_{1j},\ i = 1,\dots,k\}. \tag{10}$$

Then, the expected risk of the proposed selection procedure is

$$\begin{aligned}
& E^{(\underline{W},\underline{Y})}L(\underline{\beta}, d_n(\underline{x},\underline{y})) \qquad (11)\\
= & \sum_{i=1}^{k}\sum_{j=1,j\ne i}^{k} P_n\{i^*=i, i_n^*=j\}\\
\le & \sum_{i=1}^{k}\sum_{j=1,j\ne i}^{k} P_n\{i^*=i, i_n^*=j, S_{wwi}-\sigma_{uui} > \frac{\sigma_{xxi}}{2}, S_{wwj}-\sigma_{uuj} > \frac{\sigma_{xxj}}{2}\}\\
& +\sum_{i=1}^{k}\sum_{j=1,j\ne i}^{k} P_n\{i^*=i, i_n^*=j, S_{wwi}-\sigma_{uui} \le \frac{\sigma_{xxi}}{2}\}\\
& +\sum_{i=1}^{k}\sum_{j=1,j\ne i}^{k} P_n\{i^*=i, i_n^*=j, S_{wwj}-\sigma_{uuj} \le \frac{\sigma_{xxj}}{2}\}\\
\le & \sum_{i=1}^{k}\sum_{j=1,j\ne i}^{k} P_n\{\beta_{1i}-\hat{\beta}_{1i} > \delta, S_{wwi}-\sigma_{uui} > \frac{\sigma_{xxi}}{2}, S_{wwj}-\sigma_{uuj} > \frac{\sigma_{xxj}}{2}\}\\
& +\sum_{i=1}^{k}\sum_{j=1,j\ne i}^{k} P_n\{\hat{\beta}_{1j}-\beta_{1j} > \delta, S_{wwi}-\sigma_{uui} > \frac{\sigma_{xxi}}{2}, S_{wwj}-\sigma_{uuj} > \frac{\sigma_{xxj}}{2}\}\\
& +\sum_{i=1}^{k}\sum_{j=1,j\ne i}^{k} P_n\{S_{wwi}-\sigma_{uui} \le \frac{\sigma_{xxi}}{2}\}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k} P_n\{S_{wwj}-\sigma_{uuj}\leq \frac{\sigma_{xxj}}{2}\} \\
\leq\ & \sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k} P_n\{\beta_{1i}-\hat{\beta}_{1i}>\delta, S_{wwi}-\sigma_{uui}>\frac{\sigma_{xxi}}{2}\} \\
&+\sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k} P_n\{\hat{\beta}_{1j}-\beta_{1j}>\delta, S_{wwj}-\sigma_{uuj}>\frac{\sigma_{xxj}}{2}\} \\
&+2k\sum_{i=1}^{k} P_n\{S_{wwi}-\sigma_{uui}\leq \frac{\sigma_{xxi}}{2}\} \\
\leq\ & 2k\sum_{i=1}^{k} P_n\{|\beta_{1i}-\hat{\beta}_{1i}|>\delta, S_{wwi}-\sigma_{uui}>\frac{\sigma_{xxi}}{2}\} \\
&+2k\sum_{i=1}^{k} P_n\{S_{wwi}-\sigma_{uui}\leq \frac{\sigma_{xxi}}{2}\}.
\end{aligned}
$$

From above we observe that it suffices to analyze the convergence rates of the followings two sequences:

$$
P_n\{S_{wwi}-\sigma_{uui}\leq \frac{\sigma_{xxi}}{2}\},\quad P_n\{|\hat{\beta}_{1i}-\beta_{1i}|\geq \delta, S_{wwi}-\sigma_{uui}>\frac{\sigma_{xxi}}{2}\}. \tag{12}
$$

We analyze the rate of convergence under the following conditions.

3.1 When The α-th Moment Exists ($\alpha > 2$)

In this subsection we suppose that the α-th ($\alpha > 2$) moments of $(X_{ij}, U_{ij}, \epsilon_{ij})$ exist, that is,

$$
E|X_{ij}|^\alpha < \infty, \qquad E|U_{ij}|^\alpha < \infty, \qquad E|\epsilon_{ij}|^\alpha < \infty. \tag{13}
$$

We will show that the expected risk of the proposed selection procedure converges to 0 at the rate of $o(n^{-(\alpha/2-1)})$.

We introduce some useful lemmas. The first lemma is well known, a similar result can be found in Baum and Katz (1965).

Lemma 1. Let $X_1, \ldots, X_n$ be independent random variables with mean 0. Suppose for a fixed number $\alpha > 1$, $E|X_i|^\alpha < \infty$, for $i = 1, \ldots, n$, then for any $\epsilon > 0$,

$$P\{|\sum X_i/n| \geq \epsilon\} = o(n^{-(\alpha-1)}). \tag{14}$$

As a consequence of Lemma 1, we have

Lemma 2. Let $X_1, \ldots, X_n$ be independent random variables, with mean $EX_i = \mu$ and variance $\mathrm{Var}X_i = \sigma^2$, for $i = 1, \ldots, n$. Also let $\bar{X} = \frac{1}{n}\sum X_i$ and $S_n^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$. Suppose for $i = 1, \ldots, n$ and a fixed number $\alpha > 2$, $E|X_i|^\alpha < \infty$, then for any $\epsilon > 0$,

$$P\{|S_n^2 - \sigma^2| \geq \epsilon\} = o(n^{-(\alpha/2-1)}). \tag{15}$$

Proof.

$$\begin{aligned}
P\{|S_n^2 - \sigma^2| \geq \epsilon\} &= P\{|\frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\bar{X}^2 - \sigma^2| \geq \epsilon\} && (16)\\
&\leq P\{|\frac{1}{n-1}\sum X_i^2 - \frac{n}{n-1}(\mu^2+\sigma^2)| \geq \frac{\epsilon}{2}\} \\
&\quad + P\{|\frac{n}{n-1}\bar{X}^2 - \frac{n\mu^2+\sigma^2}{n-1}| \geq \frac{\epsilon}{2}\} \\
&= P\{|\frac{1}{n}\sum(X_i^2 - (\mu^2+\sigma^2))| \geq \frac{n-1}{n}\frac{\epsilon}{2}\} \\
&\quad + P\{|\bar{X}^2 - \mu^2| \geq \frac{\epsilon}{2} - \frac{1}{n}\sigma^2\} \\
&= P\{|\frac{1}{n}\sum(X_i^2 - (\mu^2+\sigma^2))| \geq \frac{\epsilon}{4}\} \\
&\quad + P\{|\bar{X}^2 - \mu^2| \geq \frac{\epsilon}{4}\} \\
&:= I_1 + I_2, && (17)
\end{aligned}$$

for n large enough, that is, when $n \geq \max(2, [\frac{4\sigma^2}{\epsilon}] + 1)$, we have $\frac{n-1}{n}\frac{\epsilon}{2} \geq \frac{\epsilon}{4}$ and $\frac{\epsilon}{2} - \frac{1}{n}\sigma^2 \geq \frac{\epsilon}{4}$. From Lemma 1, we have

$$\begin{aligned}
I_1 &= P\{|\frac{1}{n}\sum(X_i^2 - (\mu^2+\sigma^2))| \geq \frac{\epsilon}{4}\} && (18)\\
&= o(n^{-\alpha/2-1}) && (19)
\end{aligned}$$

and

$$
\begin{aligned}
I_2 &= P\{|\bar{X}^2-\mu^2|\geq\frac{\epsilon}{4}\} \qquad (20)\\
&= P\{|(\bar{X}+\mu)(\bar{X}-\mu)|\geq\frac{\epsilon}{4}\ \text{ and }\ (\bar{X}+\mu)>(2\mu+1)\}\\
&\quad +P\{|(\bar{X}+\mu)(\bar{X}-\mu)|\geq\frac{\epsilon}{4}\ \text{ and }\ (\bar{X}+\mu)\leq(2\mu+1)\}\\
&\leq P\{(\bar{X}-\mu)>1\}+P\{4(2\mu+1)|(\bar{X}-\mu)|\geq\epsilon\}\\
&= o(n^{-(\alpha-1)}).
\end{aligned}
$$

From Lemma 2, we can see that

$$
\begin{aligned}
P\{S_{wwi}-\sigma_{uui}\leq\frac{\sigma_{xxi}}{2}\} &= P\{S_{wwi}-\sigma_{wwi}\leq-\frac{\sigma_{xxi}}{2}\} \qquad (21)\\
&= o(n^{-(\alpha/2-1)}).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&P_n\{|\hat{\beta_1}_i-\beta_{1i}|\geq\delta, S_{wwi}-\sigma_{uui}>\frac{\sigma_{xxi}}{2}\} \qquad (22)\\
&= P_n\{|\frac{S_{wyi}}{S_{wwi}-\sigma_{uui}}-\beta_{1i}|\geq\delta, S_{wwi}-\sigma_{uui}>\frac{\sigma_{xxi}}{2}\}\\
&\leq P_n\{|S_{wyi}-\beta_{1i}(S_{wwi}-\sigma_{uui})|\geq\delta\frac{\sigma_{xxi}}{2}\}\\
&= P_n\{|\frac{1}{n-1}\sum_{j=1}^{n}(W_{ij}-\bar{W}_i)(Y_{ij}-\bar{Y}_i)-\beta_{1i}(\frac{1}{n-1}(\sum_{j=1}^{n}(W_{ij}-\bar{W}_i)^2-\sigma_{uui})|\geq\delta\frac{\sigma_{xxi}}{2}\}\\
&= P_n\{|\frac{1}{n-1}\sum_{j=1}^{n}W_{ij}Y_{ij}-\frac{n}{n-1}\bar{W}_i\bar{Y}_i-\frac{1}{n-1}\beta_{1i}\sum_{j=1}^{n}W_{ij}^2+\frac{n}{n-1}\beta_{1i}\bar{W}_i^2\\
&\qquad +\beta_{1i}\sigma_{uui}|\geq\delta\frac{\sigma_{xxi}}{2}\}\\
&\leq P_n\{|\frac{1}{n-1}\sum_{j=1}^{n}W_{ij}Y_{ij}-\frac{n}{n-1}\beta_{1i}\sigma_{xxi}|\geq\delta\frac{\sigma_{xxi}}{6}\}\\
&\quad +P_n\{|\frac{n}{n-1}\bar{W}_i\bar{Y}_i-\frac{n}{n-1}\beta_{1i}\bar{W}_i^2-\frac{1}{n-1}\beta_{1i}\sigma_{uui}|\geq\delta\frac{\sigma_{xxi}}{6}\}\\
&\quad +P_n\{|\frac{1}{n-1}\beta_{1i}\sum_{j=1}^{n}W_{ij}^2-\frac{n}{n-1}\beta_{1i}(\sigma_{xxi}+\sigma_{uui})|\geq\delta\frac{\sigma_{xxi}}{6}\}\\
&:= J_1+J_2+J_3.
\end{aligned}
$$

For any $i=1,\ldots,k$, $\{W_{ij}Y_{ij}, j=1,\ldots,n\}$ are independent random variables with mean $E(W_{ij}Y_{ij})=\beta_{1i}\sigma_{xxi}$. By Holder's inequality,

$$E|W_{ij}Y_{ij}|^{\alpha/2} \leq \sqrt{E|W_{ij}|^{\alpha}E|Y_{ij}|^{\alpha}} < \infty, \tag{23}$$

therefore, we have

$$\begin{aligned} J_1 = P_n\{|\frac{1}{n}\sum_{j=1}^{n}(W_{ij}Y_{ij} - \beta_{1i}\sigma_{xxi})| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{6}\} & \quad (24) \\ = o(n^{(-\alpha/2-1)}). & \quad (25) \end{aligned}$$

Since

$$\begin{aligned} & \bar{W}_i\bar{Y}_i - \beta_{1i}\bar{W}_i^2 \quad (26) \\ = & \bar{W}_i(\bar{Y}_i - \beta_{1i}\bar{W}_i) \\ = & \bar{W}_i(\beta_{0i} + \bar{\epsilon}_i + \beta_{1i}\bar{U}_i) \\ = & \beta_{0i}\bar{W}_i + \bar{\epsilon}_i\bar{W}_i + \beta_{1i}\bar{X}_i\bar{U}_i + \beta_{1i}\bar{U}_i^2, \end{aligned}$$

we observe that

$$\begin{aligned} J_2 = & P_n\{|\beta_{0i}\bar{W}_i + \bar{\epsilon}_i\bar{W}_i + \beta_{1i}\bar{X}_i\bar{U}_i + \beta_{1i}\bar{U}_i^2 - \frac{1}{n}\beta_{1i}\sigma_{uui}| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{6}\} \quad (27) \\ \leq & P_n\{|\beta_{0i}\bar{W}_i| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} \\ & +P_n\{|\bar{\epsilon}_i\bar{W}_i| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} \\ & +P_n\{|\beta_{1i}\bar{X}_i\bar{U}_i| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} \\ & +P_n\{|\beta_{1i}\bar{U}_i^2 - \frac{1}{n}\beta_{1i}\sigma_{uui}| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} \\ \leq & P_n\{|\beta_{0i}\bar{W}_i| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} \\ & +P_n\{|\bar{\epsilon}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} + P_n\{|\bar{W}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} \\ & +P_n\{|\sqrt{\beta_{1i}}\bar{X}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} + P_n\{|\sqrt{\beta_{1i}}\bar{U}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} \\ & +P_n\{|\beta_{1i}(\bar{U}_i^2 - \frac{1}{n}\sigma_{uui})| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\}. \end{aligned}$$

Then by Lemma 1, we have

$$P_n\{|\beta_{0i}\bar{W}_i| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} = \begin{cases} o(n^{-(\alpha-1)}), & \text{if } \beta_{0i} \neq 0, \\ 0, & \text{if } \beta_{0i} = 0, \end{cases} = o(n^{-(\alpha-1)}), \tag{28}$$

$$P_n\{|\bar{\epsilon}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} = o(n^{-(\alpha-1)}), \tag{29}$$

$$P_n\{|\bar{W}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} = o(n^{-(\alpha-1)}), \tag{30}$$

$$P_n\{|\sqrt{\beta_{1i}}\bar{X}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} = o(n^{-(\alpha-1)}), \tag{31}$$

$$P_n\{|\sqrt{\beta_{1i}}\bar{U}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} = o(n^{-(\alpha-1)}), \tag{32}$$

$$\begin{aligned} & P_n\{|\beta_{1i}(\bar{U}_i^2 - \frac{1}{n}\sigma_{uui})| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} \\ \leq\ & P_n\{|\beta_{1i}\bar{U}_i^2| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24} - \frac{1}{n}\beta_{1i}\sigma_{uui}\} \\ =\ & P_n\{|\sqrt{\beta_{1i}}\bar{U}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24} - \frac{1}{n}\beta_{1i}\sigma_{uui}}\} \\ =\ & o(n^{-(\alpha-1)}), \end{aligned} \tag{33}$$

$$\begin{aligned} J_3 = & P_n\{|\frac{1}{n}\beta_{1i}\sum_{j=1}^{n} W_{ij}^2 - \beta_{1i}(\sigma_{xxi} + \sigma_{uui})| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{6}\} \\ = & o(n^{-(\alpha/2-1)}). \end{aligned} \tag{34}$$

Hence, by combining the above arguments, we have the following theorem.

Theorem 1. The selection procedure $d_n(\underline{w}, \underline{y})$, defined in (8), is asymptotically optimal with convergence rate of order $o(n^{-(\alpha/2-1)})$ under condition (13). That is,

$$E^{(\underline{w},\underline{y})} L(\underline{\beta}, d_n(\underline{w}, \underline{y})) = o(n^{-(\alpha/2-1)}). \tag{35}$$

3.2 When The Moment Generating Function Exists

In this subsection we suppose that the moment generating functions of $\{X_{ij}^2, U_{ij}^2, \epsilon_{ij}^2\}$ exist in a neighbourhood of the origin, that is, for $-T \le t \le T$,

$$Ee^{tX_{ij}^2} < \infty, \quad Ee^{tU_{ij}^2} < \infty, \quad Ee^{t\epsilon_{ij}^2} < \infty. \tag{36}$$

where T is a positive constant.

We first introduce the following lemma, which can be found in Petrov (1995).

Lemma 3. Let $\{X_1, \ldots, X_n\}$ be independent random variables with mean $EX_i = 0, i = 1, \ldots, n$. Suppose there exist positive constants $g_1, \ldots, g_n$ and T such that

$$Ee^{tX_i} \le e^{g_i t^2/2} \qquad (i = 1, \ldots, n) \tag{37}$$

for $-T \le t \le T$. Let $G_n = \sum_{i=1}^n g_i$, then

$$P(|\sum_{i=1}^n X_i| \ge x) \le \begin{cases} e^{-(x^2/2G_n)}, & \text{if } 0 \le x \le G_n T, \\ e^{-(Tx/2)}, & \text{if } x > G_n T. \end{cases} \tag{38}$$

The following lemma clarifies the probabilistic meaning of the conditions of Lemma 3.

Lemma 4. Let X be a random variable with mean $EX = 0$. The following two assertions are equivalent:

(I) There exist positive constants g and H such that

$$Ee^{tX} \le e^{gt^2/2} \qquad \text{for } -H \le t \le H, \tag{39}$$

(II) There exists a positive constant T such that

$$Ee^{tX} < \infty \qquad \text{for } -T \le t \le T. \tag{40}$$

Proof. It is clear that (I) implies (II). We now prove that (II) also implies (I). If (II) is fulfilled, then the random variable X has the moments of all orders, and the following relation holds:

$$\log Ee^{tX} = \frac{1}{2}\sigma^2 t^2 + o(t^2) \tag{41}$$

as $t \to 0$, where $\sigma^2 = EX^2$. For any constant $g > \sigma^2$, the inequalities $\log Ee^{tX} \le gt^2/2$ and $Ee^{tX} \le e^{gt^2/2}$ hold for all sufficiently small t, that is, (I) is true. This completes the proof of Lemma 4. As we can see in the proof, we can always set $g = 2\sigma^2$.

We further assume that the 4-th moments of $\{X_{ij}, U_{ij}, \epsilon_{ij}\}$ are bounded, that is, there exists a positive constant C such that

$$EX_{ij}^4 < C, \qquad EU_{ij}^4 < C, \qquad E\epsilon_{ij}^4 < C. \tag{42}$$

We can see from (42) that EW_{ij}^4, EY_{ij}^4 and $E(W_{ij}Y_{ij})^2$ are all bounded.

We analyze $P_n\{S_{wwi} - \sigma_{uui} \le \frac{\sigma_{xxi}}{2}\}$ first.

$$\begin{aligned}
& P_n\{S_{wwi} - \sigma_{uui} \le \frac{\sigma_{xxi}}{2}\} & (43)\\
\le\ & P_n\{|S_{wwi} - \sigma_{wwi}| \ge -\frac{\sigma_{xxi}}{2}\} \\
\le\ & P_n\{|\frac{1}{n}\sum_{j=1}^{n} W_{ij}^2 - \sigma_{wwi}| \ge \frac{n-1}{n}\frac{\epsilon}{2}\} \\
& +P_n\{|\bar{W}_i^2| \ge \frac{\epsilon}{2} - \frac{1}{n}\sigma^2\} \\
=\ & P\{|\frac{1}{n}\sum_{j=1}^{n}(W_{ij}^2 - \sigma_{wwi})| \ge \frac{\epsilon}{4}\} \\
& +P\{|\bar{W}_i| \ge \sqrt{\frac{\epsilon}{4}}\} \\
:=\ & K_1 + K_2, & (44)
\end{aligned}$$

for n large enough, that is, when $n \ge \max(2, [\frac{4\sigma^2}{\epsilon}] + 1)$, we have $\frac{n-1}{n}\frac{\epsilon}{2} \ge \frac{\epsilon}{4}$ and $\frac{\epsilon}{2} - \frac{1}{n}\sigma^2 \ge \frac{\epsilon}{4}$. Since for $j = 1, \ldots, n$, $E(W_{ij} - \sigma_{wwi}) = 0$ and for $-T/2 \le t \le T/2$,

$$Ee^{tW_{ij}^2} \leq Ee^{|t(X_{ij}+U_{ij})^2|} \leq E(e^{2|t|X_{ij}^2}e^{2|t|U_{ij}^2}) = E(e^{2|t|X_{ij}^2})E(e^{2|t|U_{ij}^2}) < \infty. \quad (45)$$

By Lemma 3 and Lemma 4, we have

$$\begin{aligned} K_1 &= P\{|\frac{1}{n}\sum_{j=1}^{n}(W_{ij}^2 - \sigma_{wwi})| \geq \frac{\epsilon}{4}\} \\ &\leq \begin{cases} e^{-(n^2\epsilon^2/32G_n)}, & \text{if } \epsilon \leq 2TG_n/n, \\ e^{-(T\epsilon/8)n}, & \text{if } \epsilon > 2TG_n/n, \end{cases} \end{aligned} \quad (46)$$

where G_n is twice the sum of the n variances of $(W_{ij}^2 - \sigma_{wwi})$, $j = 1, \ldots, n$. Since $(EW_{ij}^4, j = 1, \ldots, n)$ are bounded, $G_n = O(n)$. Therefore,

$$\begin{aligned} K_1 &= P\{|\frac{1}{n}\sum_{j=1}^{n}(W_{ij}^2 - \sigma_{wwi})| \geq \frac{\epsilon}{4}\} \\ &\leq \begin{cases} e^{-(n^2\epsilon^2/32G_n)}, & \text{if } \epsilon \leq 2TG_n/n, \\ e^{-(T\epsilon/8)n}, & \text{if } \epsilon > 2TG_n/n, \end{cases} \\ &= O(e^{-c_{K_1}^* n}), \end{aligned} \quad (47)$$

where $c_{K_1}^*$ is a positive constant. Similarly, for $-T \leq t \leq T$,

$$Ee^{tW_{ij}} \leq Ee^{|tW_{ij}|} \leq Ee^{|t|(W_{ij}^2+1)} < \infty. \quad (48)$$

$$\begin{aligned} K_2 &= P\{|\bar{W}_i| \geq \sqrt{\frac{\epsilon}{4}}\} \\ &= O(e^{-c_{K_2}^* n}), \end{aligned} \quad (49)$$

where $c_{K_2}^*$ is also a positive constant.

Next we consider $P_n\{|\hat{\beta_1}_i - \beta_{1i}| \geq \delta, S_{wwi} - \sigma_{uui} > \frac{\sigma_{xxi}}{2}\}$. We have

$$P_n\{|\hat{\beta_1}_i - \beta_{1i}| \geq \delta, S_{wwi} - \sigma_{uui} > \frac{\sigma_{xxi}}{2}\} \quad (50)$$

$$
\begin{aligned}
&\leq \quad P_n\{|\frac{1}{n-1}\sum_{j=1}^{n} W_{ij}Y_{ij} - \frac{n}{n-1}\beta_{1i}\sigma_{xxi}| \geq \delta\frac{\sigma_{xxi}}{6}\} \\
&\quad +P_n\{|\frac{n}{n-1}\bar{W}_i\bar{Y}_i - \frac{n}{n-1}\beta_{1i}\bar{W}_i^2 - \frac{1}{n-1}\beta_{1i}\sigma_{uui}| \geq \delta\frac{\sigma_{xxi}}{6}\} \\
&\quad +P_n\{|\frac{1}{n-1}\beta_{1i}\sum_{j=1}^{n} W_{ij}^2 - \frac{n}{n-1}\beta_{1i}(\sigma_{xxi}+\sigma_{uui})| \geq \delta\frac{\sigma_{xxi}}{6}\} \\
&:= \quad L_1 + L_2 + L_3.
\end{aligned}
$$

For any $i = 1, \ldots, k$, $\{W_{ij}Y_{ij}, j = 1, \ldots, n\}$ are independently but not necessarily identically distributed random variables. By Cauchy-Schwarz's inequality, we have, for $-T/2 \leq t \leq T/2$,

$$
Ee^{tW_{ij}Y_{ij}} \leq Ee^{|tW_{ij}Y_{ij}|} \leq Ee^{|t|\frac{(W_{ij}^2+Y_{ij}^2)}{2}} \leq \sqrt{Ee^{|t|W_{ij}^2}Ee^{|t|Y_{ij}^2}} < \infty. \tag{51}
$$

Besides, for each i and j, the variance of $W_{ij}Y_{ij}$ is bounded, therefore,

$$
L_1 = P_n\{|\frac{1}{n}\sum_{j=1}^{n}(W_{ij}Y_{ij} - \beta_{1i}\sigma_{xxi})| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{6}\} \tag{52}
$$

$$
= \quad O(e^{-c_{L_1}^* n}), \tag{53}
$$

where $c_{L_1}^*$ is a positive constant. Next we analyze L_2 and L_3. Similarly,

$$
\begin{aligned}
&L_2 \leq P_n\{|\beta_{0i}\bar{W}_i| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} && (54) \\
&+P_n\{|\bar{\epsilon}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} + P_n\{|\bar{W}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} \\
&+P_n\{|\sqrt{\beta_{1i}}\bar{X}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} + P_n\{|\sqrt{\beta_{1i}}\bar{U}_i| \geq \sqrt{\frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}}\} \\
&+P_n\{|\beta_{1i}(\bar{U}_i^2 - \frac{1}{n}\sigma_{uui})| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{24}\} \\
&= \quad O(e^{-c_{L_2}^* n}), && (55)
\end{aligned}
$$

and

$$
L_3 = P_n\{|\frac{1}{n}\beta_{1i}\sum_{j=1}^{n} W_{ij}^2 - \beta_{1i}(\sigma_{xxi}+\sigma_{uui})| \geq \frac{n-1}{n}\delta\frac{\sigma_{xxi}}{6}\} \tag{56}
$$

$$
= \quad O(e^{-c_{L_3}^* n}), \tag{57}
$$

where $c^*_{L_2}$ and $c^*_{L_3}$ are positive constants. Hence, by the above argument, if we set $c^* = \min(c^*_{K_1}, c^*_{K_2}, c^*_{L_1}, c^*_{L_2}, c^*_{L_3})$, then $c^* > 0$. We have the following theorem.

Theorem 2. The selection procedure $d_n(\underline{w}, \underline{y})$, defined in (8), is asymptotically optimal with convergence rate of order $O(e^{-c^* n})$ under conditions (36) and (42). That is,

$$E^{(\underline{W}, \underline{Y})} L(\underline{\beta}, d_n(\underline{w}, \underline{y})) = O(e^{-c^* n}). \tag{58}$$

We consider two special situations next.

Two special situations.

1. $\{(X_{ij}, U_{ij}, \epsilon_{ij}), 1 \le j \le n\}$ are normally distributed. In this case, $\{(X_{ij}, U_{ij}, \epsilon_{ij})\}$ are i.i.d. $N_3((0,0,0), \mathrm{diag}(\sigma_{xxi}, \sigma_{uui}, \sigma_{\epsilon\epsilon i}))$. Since $(X^2_{ij}/\sigma_{xxi}, U^2_{ij}/\sigma_{uui}, \epsilon^2_{ij}/\sigma_{\epsilon\epsilon i})$ are χ^2 distributed and the moment generating functions of them exist in a neighbourhood of 0, and the 4-th moments of $\{(X_{ij}, U_{ij}, \epsilon_{ij})\}$ are also bounded, by Theorem 2, we have that, in the normal case, the selection procedure $d_n(\underline{w}, \underline{y})$, defined in (8), is asymptotically optimal with the rate of convergence of order $O(e^{-c^* n})$.

2. $\{(X_{ij}, U_{ij}, \epsilon_{ij}), 1 \le j \le n\}$ are bounded. Then conditions (36) and (42) always hold and therefore, the selection procedure $d_n(\underline{w}, \underline{y})$ is asymptotically optimal with convergence rate of order $O(e^{-c^* n})$.

4 Simulations

We carried out a simulation study to investigate the preformance of the selection procedure d_n. The expected risk $E^{(\underline{W}, \underline{Y})} L(\underline{\beta}, d_n(\underline{w}, \underline{y}))$ is used as measure of the performance of the selection rule. For any observations $(\underline{W}, \underline{Y})$, let

$$D(\underline{W}, \underline{Y}) = \begin{cases} 1, & \text{if we make a wrong selection,} \\ 0, & \text{if we make a correct selection.} \end{cases} \tag{59}$$

Then, by the law of large numbers, the sample mean of $D(\underline{W}, \underline{Y})$, based on our observations, can be used as an estimator of the expected risk $E^{(\underline{W}, \underline{Y})} L(\underline{\beta}, d_n(\underline{w}, \underline{y}))$.

The simulation scheme is described as follows:

1. For each $j = 1, \ldots, n$ and each i=1, 2 and 3, we generated independent random observations $(X_{ij}, U_{ij}, \epsilon_{ij})$ from multivariate normal $N_3((0,0,0)^T, \mathrm{diag}(\sigma_{xxi}, \sigma_{uui}, \sigma_{\epsilon\epsilon i}))$.

2. Let $W_{ij} = X_{ij} + U_{ij}$ and $Y_{ij} = \beta_{0i} + \beta_{1i} X_{ij} + \epsilon_{ij}$.

3. Based on (W_{ij}, Y_{ij}), we obtained the estimators of β_{1i}, then made the selection and computed $D(\underline{\mathrm{W}}, \underline{\mathrm{Y}})$.

4. Step 1, 2 and 3 were repeated 4000 times. The average of $D(\underline{\mathrm{W}}, \underline{\mathrm{Y}})$ based on the 4000 repetitions, which is denoted by D_n, is used as an estimator of the expected risk $E^{(\underline{\mathrm{W}},\underline{\mathrm{Y}})} L(\underline{\beta}, d_n(\underline{\mathrm{w}}, \underline{\mathrm{y}}))$.

The results are listed for the case where

$$\begin{gathered}\sigma_{xx1} = \sigma_{xx2} = \sigma_{xx3} = 1, \\ \sigma_{uu1} = \sigma_{uu2} = \sigma_{uu3} = 1, \\ \sigma_{\epsilon\epsilon 1} = \sigma_{\epsilon\epsilon 2} = \sigma_{\epsilon\epsilon 3} = 0.25, \\ \beta_{01} = \beta_{02} = \beta_{03} = 0, \\ \beta_{11} = 0, \beta_{12} = 1, \beta_{13} = 2.\end{gathered}$$

and

$$n = 5, 10, 15, 20, 30, 40, 50, 100.$$

From the results of the simulation (see the last page), we can observe that the values of D_n decrease quite rapidly as n increases, for $n \leq 100$. This supports Theorem 2 that the rate of convergence is of order $o(e^{-c^* n})$.

5 Acknowledgement

The authors are very grateful to Professor Zhengyan Lin of Hangzhou University, Hangzhou, China, for helpful comments and discussion.

References

Baum, L. E. and Katz, M. (1965). Convergence Rates in the Law of Large Numbers, Transactions of the Amer. Math. Society, Vol. 120, 1, 108-123.

Bhattacharya, R. N. and Ghosh J. K. (1978). On the validity of the formal Edgeworth expansion. Ann. Statist. 6, 434-451.

Bhattacharya, R. N. and Ghosh J. K. (1989). On moment conditions for valid formal Edgeworth expansions. In: Multivariate Statistics and Probability: Essays in Memory of P.R. Krishnaiah. (C.R. Rao and M.M. Rao, eds.), Academic, New York, 68-79.

Carroll R. J., Ruppert D. and Stefanski L. A. (1995). Measurement error in onlinear models. Chapman and Hill, New York.

Casella, G. and Berger, R. (1990). Statistical Inference. Duxbury Press, Belmont, California.

Fuller, W. A. (1987). Measurement error models. John Wiley, New York.

Gupta, S. S. and Liang, T. (1996). Selecting good exponential populations compared with a control: a nonparametric empirical Bayes approach. Department of Statistics, Purdue University, Technical Report #96-18C.

Gupta S. S. and Liang T. (1996). On empirical Bayes simultaneous selection procedures for comparing normal populations with a standard. Department of Statistics, Purdue University, Technical Report #96-25C.

Gupta, S. S. and Panchapakesan S. (1996). Design of experiments with selection and ranking goals. *Handbook of Statistics*. (S. Ghosh and C. R. Rao, eds.), Elsevier Science B. V., Vol. 13, 555-584.

Hall, P. (1990). On the relative performance of bootstrap and Edgeworth approximations of a distribution function. J. Muti. Anal. 35, 108-120.

Hall, P. (1991). Edgeworth expansions for nonparametric density estimators with applications. Statistics 22, 215-232.

Hall, P. (1992). The Bootstrap and Edgeworth expansion. Springer-Verlag, New York.

Petrov, V. V. (1995). Limit Theorems of Probability Theory - Sequences of Independent Random Variables. Oxford Studies in Probability Series, Clarendon Press, Oxford.

Weber, N. C. (1984). On resampling techniques for regression models. Statist. Probab. Lett. 2, 275-278.

n	D_n
5	0.250
10	0.090
15	0.030
20	0.003
30	0.001
40	0.001
50	0.000
100	0.000