



Improved estimation in multiple linear regression models with measurement error and general constraint

Hua Liang^{a,*}, Weixing Song^b

^a Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642, USA

^b Department of Statistics, Kansas State University, Manhattan, KS, 66506, USA

ARTICLE INFO

Article history:

Received 3 December 2007

Available online 20 August 2008

AMS 2000 subject classifications:

primary 62J05

62F30

secondary 62J99

Keywords:

Asymptotic distributional quadratic bias

Asymptotic distributional quadratic risk

Attenuation-correction estimator

James–Stein-type estimator

Positive rule Stein type estimator

Preliminary test estimator

Risk function

ABSTRACT

In this paper, we define two restricted estimators for the regression parameters in a multiple linear regression model with measurement errors when prior information for the parameters is available. We then construct two sets of improved estimators which include the preliminary test estimator, the Stein-type estimator and the positive rule Stein type estimator for both slope and intercept, and examine their statistical properties such as the asymptotic distributional quadratic biases and the asymptotic distributional quadratic risks. We remove the distribution assumption on the error term, which was generally imposed in the literature, but provide a more general investigation of comparison of the quadratic risks for these estimators. Simulation studies illustrate the finite-sample performance of the proposed estimators, which are then used to analyze a dataset from the Nurses Health Study.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Improvement of estimation for regression models is a fundamental and interesting topic. In certain cases, one may have, but is not sure, some prior information for the parameters of interest. By incorporating the information into the estimation procedure, one may obtain more efficient estimators than the estimators obtained when the prior information is ignored. Statistical approaches for developing more efficient estimators can be roughly classified into two categories. The first one focuses on developing a proper test procedure to check the validity of the uncertain prior. If the prior information is confirmed, then the commonly used estimators are modified to accommodate the prior. The second one is to develop a procedure in which test and estimation can be conducted simultaneously. The first procedure is very natural and commonly used for theory and application purposes. For example, consider the multiple linear regression models:

$$Y = X\beta + \varepsilon, \quad (1.1)$$

where X is the $n \times p$ design matrix with rank p , β is the $p \times 1$ regression parameter, and ε is the $n \times 1$ random error vector. Suppose that we have prior information for β , which can be described as $R\beta = r$, where R is a $q \times p$ known matrix of rank q and r is a $q \times 1$ known vector, $q \leq p$. A proper test statistic would be based on a distance between $R\hat{\beta}_n$ and r , where $\hat{\beta}_n$ is a “good” estimator of β , for instance, the least squares estimator, $\hat{\beta}_{LS} = (X'X)^{-1}X'Y$, or the maximum likelihood

* Corresponding author.

E-mail address: HLiang@bst.rochester.edu (H. Liang).

estimator. If the prior information is rejected, one should keep using these “good” estimators, otherwise restricted least squares estimators should be used.

The estimators for the regression parameters, which fall into the second category, include the preliminary test estimator (PTE), the James–Stein-type estimator (JSTE), and the positive rule Stein type estimator (PRSE). See [1,2] for a detailed discussion on these estimators. Bancroft [3] was among the first to consider PTE. Saleh and Sen [4] extended his idea to a nonparametric setup. JSTE was introduced by Stein [5] and James and Stein [6], and expanded by Saleh and Sen [4,7] and Sen and Saleh [8] to nonparametric areas.

The aforementioned estimation techniques have received much attention recently in linear regression model when the covariates are measured with errors. Stanley [9,10] revealed that JSTE can eliminate inconsistency of the classical least squares estimators. Shalabh [11] studied properties of JSTE when the covariance matrix of the measurement errors is known. For the simple linear regression model with measurement error, when the slope parameter may be the null scalar and all the random components are normally distributed, Kim and Saleh [12] compared these estimators in the sense of asymptotic distributional quadratic bias, mean square error matrix, and asymptotic distributional quadratic risk. Their comparisons show that PTE behaves better than the attenuation-correction (AC) estimator if the slope is close to 0, but not uniformly better than the AC estimator over the whole range of the regression parameter. Kim and Saleh [13] further investigated the same question for multiple linear models under the same assumption and setting. Various risk functions based on the asymptotic distribution of the estimators under certain local alternatives are calculated and compared. They also showed that JSTE and PRSE dominate the AC estimator.

This paper mainly focuses on the estimation problem in multiple linear regression models with measurement errors. The contributions we made to the existing literature in this work contain four parts:

- (i) to remove the normality assumption. The normality assumption greatly simplifies the theoretical argument, but it is often violated in a practical study. The removal of the normality assumption will make the theoretical results more general and applicable;
- (ii) to improve estimation under the general constraint $R\beta = r$, which contains the case investigated by Kim and Saleh [12,13], $\beta = 0$, as a special one. The theoretical difficulty lies in the question that how the asymptotic distribution of the AC estimator under the local hypothesis depends on the unknown parameter β . This has a substantial impact on constructing the estimators subject to the constraint, or the restricted estimator (RE), which is the building block for constructing PTE, JSTE and PRSE;
- (iii) to calculate the risk functions for linear combinations of intercept and slope parameters to help estimate the mean response;
- (iv) to explicitly compare risk functions for various proposed estimators under a certain circumstance.

The outline of the paper is as follows. In Section 2, we define two REs, based on which two sets of PTE, JSTE and PRSE are constructed. The risk functions of various estimators for the slope parameters under the null hypothesis and local alternatives are presented in Section 3. Also the risks are compared among proposed estimators in some special cases. Simulation studies are presented in Section 4. The proposed estimators are used to analyze a dataset from the Nurses Health Study in Section 5. Our results provide more appropriate estimates and information for nutritional study. The proofs of the main results are shifted to the Appendix.

2. Improved estimator in multiple linear model

Suppose that (Y_i, X_i) , $i = 1, \dots, n$, constitute an independent and identically distributed sample from the linear regression model:

$$Y_i = \beta_0 + X_i' \beta + \varepsilon_i, \quad (2.1)$$

where X_i is a $p \times 1$ vector-valued covariate. We are interested in the estimation of the unknown parameter β when the covariates X_i are measured with error. Instead of observing X_i , we observe $W_i = X_i + u_i$, where the measurement errors u_i are independent and identically distributed, independent of (Y_i, X_i) , with mean zero and covariance matrix Σ_{uu} which is assumed to be known throughout this paper.

Section 4.4.3 of [2] provides a general road map to construct improved estimators if one has some uncertain information about the unknown parameter, say $\theta \in \Theta$, in a statistical model. First, we obtain an optimal unrestricted estimator, say θ_n , and an optimal RE, say θ_n^* , by likelihood method if the likelihood function is available, or by least squares method if the likelihood function is unavailable. Second, we construct an optimum test statistic, say L_n , for testing the “uncertain prior information”, say $\theta \in \Theta_0$, where Θ_0 is a subset of the parameter space Θ . Third, we construct PTE of θ as $\theta_{n,\text{PTE}} = \theta_n - (\theta_n - \theta_n^*)I(L_n < L_{n,\alpha})$, where $L_{n,\alpha}$ is the α -level critical value of L_n from its distribution under $H_0 : \theta \in \Theta_0$. Finally, we replace the indicator function $I(L_n < L_{n,\alpha})$ by a smooth decreasing function cL_n^{-1} , where c is a suitable constant derived by using empirical Bayesian theory, then JSTE is defined by $\theta_{n,\text{JSTE}} = \theta_n - c(\theta_n - \theta_n^*)L_n^{-1}$, and PRSE by $\theta_{n,\text{PRSE}} = \theta_n^* + (1 - cL_n^{-1})I(L_n > c)(\theta_n - \theta_n^*)$.

To adopt the above general rule in our setting, we need to find an “optimal” test statistic to check $R\beta = r$. The test statistic we use in this paper is $L_n = n(R\hat{\beta}_{\text{AC}} - r)'(R\hat{C}_n R')^{-1}(R\hat{\beta}_{\text{AC}} - r)$, where $\hat{\beta}_{\text{AC}}$ is the AC estimator which is defined by

$$\hat{\beta}_{\text{AC}} = (S_{WW} - \Sigma_{uu})^{-1}S_{WY}, \quad (2.2)$$

where S_{WW} and S_{WY} are the sample covariance matrices of W_i 's, and the sample covariance between W_i 's and Y_i 's, respectively, \hat{G}_n is a consistent estimator of the asymptotic covariance matrix of $\hat{\beta}_{AC}$, denoted by G which is defined by (2.5) in the next section. Under general conditions, L_n has an asymptotic χ^2 -distribution with q degrees of freedom. In fact, L_n is the likelihood ratio statistic under the normality assumptions. The next step towards our goal is to find an RE for β under the general constraint.

2.1. Construction of estimator subject to the constraint

Here is the intuition behind the methods of construction of RE. If (X, ε, U) is normally distributed, then we can directly calculate the conditional expectation $E(Y|W)$, which is a linear function of W and can be used to derive the maximum likelihood estimators of β_0 and β , and the associated RE. If (X, ε, U) is not normally distributed, $E(Y|W)$ may not easily be calculated, or may be a nonlinear function of W . We give another version of RE as follows.

Assume that $(X_i, \varepsilon_i, u_i)$ follow $N_{2p+1}[(\mu'_x, 0, \mathbf{0}')', \text{blockdiag}(\Sigma_{xx}, \sigma^2, \Sigma_{uu})]$. It is easy to see that $E(Y_i|W_i) = \nu_0 + \nu'W_i$, $\text{Var}(Y_i|W_i) = \beta' \Sigma_{xx}(I - K_{xx})\beta + \sigma^2 \triangleq \sigma_{y|w}^2$, where $\nu_0 = \beta_0 + \beta'(I - K_{xx})\mu_x$, $\nu = K_{xx}\beta$, $K_{xx} = \Sigma_{WW}^{-1} \Sigma_{xx}$ is the reliability matrix. Gleser [14] and Kim and Saleh [13] showed that the maximum likelihood estimators of ν_0 , ν and $\sigma_{y|w}^2$ are just the naive least squares estimators, namely $\hat{\nu}_{0n} = \bar{Y} - \bar{v}'_n \bar{W}$, $\hat{\nu}_n = S_{WW}^{-1} S_{WY}$ and $\hat{\sigma}_{y|w}^2 = n^{-1} \sum_{i=1}^n (Y_i - \hat{\nu}_{0n} - \hat{\nu}'_n W_i)^2$ provided $\hat{\sigma}_{y|w}^2 - \hat{\nu}'_n \hat{K}_{xx}^{-1} \Sigma_{uu} \hat{\nu}_n$ is non-negative, where $\hat{K}_{xx} = S_{WW}^{-1} \hat{\Sigma}_{xx} = S_{WW}^{-1} (S_{WW} - \Sigma_{uu})$, \bar{Y} and \bar{W} are the sample means of Y_i 's and W_i 's, respectively. Hence an estimator for the slope β can be obtained from $\nu = K_{xx}\beta$ with ν replaced by $\hat{\nu}_n$ and K_{xx} by \hat{K}_{xx} . This leads to the AC estimator $\hat{\beta}_{AC}$ given by (2.2), and $\hat{\beta}_0 = \bar{Y} - \bar{W}'\hat{\beta}_n$.

To construct REs for β_0 and β , we first give a RE for ν . Note that the general restriction can be written as $RK_{xx}^{-1}\nu = r$. If K_{xx} is known, then using the Lagrangian multiplier, one can show that the restricted maximum likelihood estimators of ν , ν_0 are given by,

$$\hat{\nu}_{RE} = \hat{\nu}_n - S_{WW}^{-1} K_{xx}^{-1} R' (RK_{xx}^{-1} S_{WW}^{-1} K_{xx}^{-1} R)^{-1} (RK_{xx}^{-1} \hat{\nu}_n - r), \quad \hat{\nu}_{0n} = \bar{Y} - \bar{W}'\hat{\nu}_{RE}.$$

If the reliability matrix K_{xx} is unknown, replace it with \hat{K}_{xx} . Then the REs for β , β_0 are defined as

$$\hat{\beta}_{RE} = \hat{K}_{xx}^{-1} \hat{\nu}_{RE} = \hat{\beta}_{AC} - \hat{H}_n R' (R\hat{H}_n R')^{-1} (R\hat{\beta}_{AC} - r), \quad \hat{\beta}_{0,RE} = \bar{Y} - \bar{W}'\hat{\beta}_{RE}, \tag{2.3}$$

where $\hat{H}_n = \hat{K}_{xx}^{-1} S_{WW}^{-1} \hat{K}_{xx}^{-1}$, which can easily be shown to be a consistent estimator of $H = K_{xx}^{-1} \Sigma_{WW}^{-1} K_{xx}^{-1}$.

If the random components are not normal, the conditional expectation $E(Y|W) = \beta_0 + \beta'E(X|W)$ may not be linear in W , and the conditional variance $\text{Var}(Y|W)$ may vary with W . Hence, a linear model is inappropriate. In fact, linearity of $E(X|W)$ in W and homoscedasticity of $\text{Var}(X|W)$ in W imply that (X', U') must be multivariate normal, see [15,16]. Although the estimators given in (2.3) are obtained based on the normal assumption, it is worth investigating whether the estimators $\hat{\beta}_{RE}$, $\hat{\beta}_{0,RE}$ have good properties in non-normal settings. See [14] for the details.

Another way to construct the REs in the measurement error regression models is to mimic the procedure for the RE in model (1.1). That is, $\tilde{\beta}_{RE} = \hat{\beta}_{LS} - SR'(RSR')^{-1}(R\hat{\beta}_{LS} - r)$, where S is the asymptotic covariance matrix of $\hat{\beta}_{LS}$. Then an RE for β in the measurement error setting can be defined as

$$\tilde{\beta}_{RE} = \hat{\beta}_{AC} - \hat{G}_n R' (R\hat{G}_n R')^{-1} (R\hat{\beta}_{AC} - r), \tag{2.4}$$

where \hat{G}_n is a consistent estimator of the asymptotic covariance matrix G of $\hat{\beta}_{AC}$, and

$$G = (\sigma^2 + \beta' \Sigma_{uu} \beta) \Sigma_{xx}^{-1} \Sigma_{WW} \Sigma_{xx}^{-1} + \Sigma_{xx}^{-1} [E(uu' \beta \beta' uu') - \Sigma_{uu} \beta \beta' \Sigma_{uu} - \Sigma_{uu} (\beta' \Sigma_{uu} \beta)] \Sigma_{xx}^{-1}. \tag{2.5}$$

See [17,18,13] how to derive \hat{G}_n under the normal and non-normal setups. The RE for β_0 can be defined accordingly. Clearly, the difference between the REs given in (2.3) and (2.4) comes from the difference between \hat{H}_n and \hat{G}_n .

2.2. Construction of improved estimators

Now we are ready to give PTE, JSTE and PRSE for the regression coefficient β based on Saleh's [2] general rule and the REs defined by (2.3) and (2.4). For the sake of clarity, we shall put a $\hat{\cdot}$ over those estimators based on $\hat{\beta}_{RE}$, and a $\tilde{\cdot}$ over those based on $\tilde{\beta}_{RE}$.

- PTE: $\hat{\beta}_{PTE} = \hat{\beta}_{AC} - (\hat{\beta}_{AC} - \hat{\beta}_{RE})I(L_n < \chi_\alpha^2)$, where χ_α^2 is the upper α -percentile of the χ^2 -distribution with degrees of freedom q ;
If $q \geq 3$, then one can define
- JSTE: $\hat{\beta}_{JSTE} = \hat{\beta}_{AC} - (q - 2)(\hat{\beta}_{AC} - \hat{\beta}_{RE})L_n^{-1}$,
- PRSE: $\hat{\beta}_{PRSE} = \hat{\beta}_{RE} + [1 - (q - 2)L_n^{-1}]I(L_n > q - 2)(\hat{\beta}_{AC} - \hat{\beta}_{RE})$.

In a similar way, we can define $\tilde{\beta}_{PTE}$, $\tilde{\beta}_{JSTE}$, and $\tilde{\beta}_{PRSE}$ by replacing $\hat{\beta}_{RE}$ with $\tilde{\beta}_{RE}$ in the expressions of $\hat{\beta}_{PTE}$, $\hat{\beta}_{JSTE}$ and $\hat{\beta}_{PRSE}$, respectively. It would be interesting to see what differences these estimators based on the two REs may have with respect to the risk comparisons.

From the above definitions, one can see that if the data yield $L_n < \chi_\alpha^2$, then $\widehat{\beta}_{PTE} = \widehat{\beta}_{RE}$, otherwise, $\widehat{\beta}_{PTE} = \widehat{\beta}_{AC}$. So PTE is indeed a simple mixture of the AC estimator and the RE. In the ordinary two-step procedure, one would test the hypothesis $R\beta = r$ first, then based on the testing result decide which estimator will be adopted. PTE simply combines these two steps to form a single one. That is, testing and estimation are done simultaneously, while JSTE replaces the indicator function $I(L_n < \chi_\alpha^2)$ with a continuous function, $(q - 2)L_n^{-1}$, of L_n . In the normal case, one can actually obtain JSTE by the empirical Bayesian estimation approach. The constant appearing in JSTE in the classical multiple regression model should be $(q - 2)(n - p)/(n - p + 2)$ instead of $q - 2$ [2]. Since we take into account the asymptotic risk function, a change of the constant to $q - 2$ will not induce any difference in large sample sense, although this change may have some impacts on the small sample behavior of the estimators. The derivation of PRSE in the current setting is similar to the counterpart in the classical regression case. In fact, if L_n tends to 0, then JSTE may go “past” the estimator $\widehat{\beta}_{RE}$, or JSTE may have a different sign from the RE. As a partial remedy, one may restrict $L_n > q - 2$, which results in PRSE. The corresponding PTE, JSTE and PRSE for the intercept can be defined accordingly.

For the sake of convenience, in what follows, we shall call the estimators $\widehat{\beta}_{RE}$, $\widehat{\beta}_{PTE}$, $\widehat{\beta}_{JSTE}$, and $\widehat{\beta}_{PRSE}$ the “hat” estimators, and the estimators $\widetilde{\beta}_{RE}$, $\widetilde{\beta}_{PTE}$, $\widetilde{\beta}_{JSTE}$, and $\widetilde{\beta}_{PRSE}$ the “tilde” estimators.

2.3. Asymptotic distribution of the new estimators

To begin with, we state regularity conditions associated with model (2.1), which will be used throughout the current and subsequent sections. Some conditions are already mentioned in the previous section. They are listed again for the sake of completeness.

- C1: The measurement errors $\varepsilon_i, i = 1, \dots, n$, are i.i.d. with mean 0 and finite positive variance σ^2 ; the measurement error vectors $u_i, i = 1, \dots, n$, are i.i.d. with mean vector 0 and covariance matrix Σ_{uu} , which is known. The fourth moment of the Euclidean norm for u exists, that is, $E\|u\|^4 < \infty$.
- C2: ε_i and u_i are independent, $i = 1, \dots, n$.
- C3: X_i 's are i.i.d. with mean vector μ_x and finite positive covariance matrix Σ_{xx} , and are independent of ε_i and u_i .

These conditions are quite common in the literature of measurement error models. The existence of the fourth moment of u is needed to ensure the asymptotic normality of the AC estimator. The existence of the fourth moment of u was also assumed in [19].

We begin with considering the asymptotic behavior of the improved estimators under the fixed alternative: $H_a : R\beta - r = \delta \neq 0$, where δ is a vector of length q . We claim that $L_n \rightarrow \infty$ in probability as $n \rightarrow \infty$. This claim is based on the following expression.

$$L_n = n(R\widehat{\beta}_n - R\beta)'(R\widehat{G}_nR')^{-1}(R\widehat{\beta}_n - R\beta) + 2n(R\widehat{\beta}_n - R\beta)'(R\widehat{G}_nR')^{-1}\delta + n\delta'(R\widehat{G}_nR')^{-1}\delta.$$

The first term is the order of $O_p(1)$, the second $O_p(\sqrt{n})$ and the third $O_p(n)$ and positive.

Let β_n^* be a generic notation for the hat estimators and tilde estimators. Under the fixed alternative, it can be seen that $\sqrt{n}(\beta_n^* - \beta) = \sqrt{n}(\widehat{\beta}_{AC} - \beta) + o_p(1)$ if Conditions C1–C3 hold. That is, all the estimators defined above are asymptotically equivalent to $\widehat{\beta}_{AC}$. This implies that the asymptotic risk functions are all the same if H_a is true. Thus we cannot infer any difference among these estimators.

To obtain meaningful risk comparisons among these estimators, we consider a sequence of local alternatives, that is,

$$H_{na} : R\beta_n - r = \delta/\sqrt{n}, \tag{2.6}$$

with fixed R, δ and $r, \lim_{n \rightarrow \infty} \beta_n = \beta, R\beta = r$, where δ is a $q \times 1$ vector. Write

$$J_C = CR'(RCR')^{-1}, \quad C = G \text{ or } H. \tag{2.7}$$

The following theorem states the asymptotic distributions of the AC, hat and tilde estimators under the local alternatives (2.6).

Theorem 2.1. *Suppose Conditions C1–C3 hold. Then, under the local alternatives (2.6), as $n \rightarrow \infty$, we have, in distribution,*

$$\sqrt{n}(\beta_n^* - \beta_n) \rightarrow Z - \phi(L)J_C(RZ + \delta),$$

where $Z \sim N(0, G), L = (RZ + \delta)'(RGR')^{-1}(RZ + \delta)$. For $\beta_n^* = \widehat{\beta}_{AC}, \phi(L) = 0$; for the hat estimators, $C = H$, for $\widehat{\beta}_{RE}, \phi(L) = 1$; for $\widehat{\beta}_{PTE}, \phi(L) = I_{(0, \chi_\alpha^2)}(L)$; for $\widehat{\beta}_{JSTE}, \phi(L) = (q - 2)L^{-1}$; and for $\widehat{\beta}_{PRSE}, \phi(L) = 1 - [1 - (q - 2)L^{-1}]I_{(q-2, \infty)}(L)$. For the tilde estimators, $C = G$, the $\phi(\cdot)$ functions are the same as their counterparts in the hat estimators.

In particular, if $R = I_{p \times p}$, then the asymptotic distribution of the two REs are the same, so are the asymptotic distributions of the two PTEs, JSTEs, and PRSEs.

3. Asymptotic bias and risk functions

The commonly used quantities to evaluate the performance of an estimator are its bias, variance or mean squared error. For the estimators proposed in the last section, we will calculate their asymptotic distributional quadratic bias, asymptotic

distributional quadratic risk function with different weight matrices, and show that all estimators but AC are asymptotically biased. The general comparisons between different estimators based on these asymptotic result turn out to be very difficult, but detailed comparisons may be made under special circumstances. We mainly focus on the discussions of asymptotic bias and risk functions of the estimators for the slope parameters and certain linear combinations of the intercept and slope parameters.

The asymptotic distributional bias function of the estimator β_n^* is defined as the bias of the asymptotic distribution of $\sqrt{n}(\beta_n^* - \beta_n)$, that is, $b(\beta, \beta_n^*) = E[Z - \phi(L)J_C(RZ + \delta)]$ according to Theorem 2.1. To make meaningful comparisons among the asymptotic distributional biases, we define the asymptotic distributional quadratic bias as below:

$$B(\beta, \beta_n^*) = b(\beta, \beta_n^*)'b(\beta, \beta_n^*).$$

Let M be a known positive definite matrix. The asymptotic risk function of the estimator β_n^* of β is defined as

$$\rho(\beta, \beta_n^*) = E\{[Z - \phi(L)J_C(RZ + \delta)]'M[Z - \phi(L)J_C(RZ + \delta)]\}.$$

To concisely state the results, the following notation is used.

$$\begin{aligned} \mu &= G^{1/2}R'(RGR')^{-1}\delta, & \lambda &= \delta'(RGR')^{-1}\delta, & \eta &= \delta'J'_C MJ_C \delta, \\ g_{q+i}(\lambda) &= P(\chi_{q+i,\lambda}^2 < \chi_\alpha^2), & h_{j,q+i}(\lambda) &= (q-2)^j E[(\chi_{q+i,\lambda}^2)^{-j}], \\ k_{j,q+i}(\lambda) &= P(\chi_{q+i,\lambda}^2 < q-2) + (q-2)^j E[(\chi_{q+i,\lambda}^2)^{-j} I(\chi_{q+i,\lambda}^2 > q-2)], \end{aligned}$$

where $i = 2, 4, j = 1, 2$, $\chi_{q+i,\lambda}^2$ is the noncentral χ^2 random variable with degrees of freedom $q+i$ and noncentral parameter λ .

The derivation of the asymptotic distributional bias functions can be done by using the fact $E\phi(L)Z = G^{1/2}\mu[E\phi(\chi_{q+2,\lambda}^2) - E\phi(\chi_{q,\lambda}^2)]$, for any measurable function $\phi(\cdot)$. This can be verified by using (A.8) in the Appendix. The details are omitted for the sake of brevity. The following theorem lists the results for the asymptotic distributional bias. Its proof can be finished by Theorem 2.1. We omit the details.

Theorem 3.1. *Suppose Conditions C1–C3 hold. Then, under the local alternatives, the asymptotic biases and asymptotic distributional quadratic biases for the AC, hat and tilde estimators are given by*

$$b(\beta, \beta_n^*) = -f(\lambda)J_C \delta, \quad B(\beta, \beta_n^*) = f^2(\lambda)\delta'J'_C J_C \delta$$

respectively. For the AC estimator, $f(\lambda) = 0$; for all hat estimators, $C = H$, and for all tilde estimators, $C = G$. For RE, $f(\lambda) = 1$; for PTE, $f(\lambda) = g_{q+2}(\lambda)$; for JSTE, $f(\lambda) = h_{1,q+2}(\lambda)$; and for PRSE, $f(\lambda) = k_{1,q+2}(\lambda)$.

Within each set of estimators, the asymptotic distributional quadratic biases can be compared simply through the quantities $g_{q+2}(\lambda)$, $h_{1,q+2}(\lambda)$, and $k_{1,q+2}(\lambda)$. It is worth mentioning that among each set of estimators, the asymptotic distributional quadratic biases of PTE, JSTE and PRSE are smaller than that of RE, since $g_{q+2}(\lambda)$, $h_{1,q+2}(\lambda)$, and $k_{1,q+2}(\lambda)$ are all not bigger than 1.

The more interesting comparison would be made between the hat estimators and tilde estimators. A natural way to make the comparisons is to compute the difference between the biases or the risk functions. However, the computation is rather complex in general situations. But for some special cases, the comparison can be made easily. In particular, we have the following corollary.

Corollary 3.1. *Suppose $u \sim N_p(0, \sigma_{uu}I)$, $\Sigma_{xx} = \sigma_{xx}I$, then for any δ , the asymptotic distributional quadratic biases of the hat estimators are all less than those of the corresponding tilde estimators.*

For an illustrative purpose, let $p = 5$, $q = 3$ (recall that $q > 2$ is required in constructing JSTE and PRSE), $\alpha = 0.15$, $\beta = (3, 3, 1, 2, 3)'$, $\Sigma_{xx} = \Sigma_{uu} = I_{5 \times 5}$, $u \sim N(0, \Sigma_{uu})$, $\sigma^2 = 1$, and R be a 3×5 matrix with the components in the first two columns all 0 and the last three columns forming an identity matrix. δ is a vector with same elements, specifically, $\delta = \delta \mathbf{1}$, where δ is a scalar, $\mathbf{1}$ is a vector of 1's. The plot (a) in Fig. 1, in which the dotted lines represent the asymptotic distributional quadratic biases for the tilde estimators and the solid lines the hat estimators, delineates the asymptotic distributional quadratic biases. In our case, the asymptotic distributional quadratic bias functions of the hat and tilde estimators for each type of estimator (RE, PTE, JSTE or PRSE) are close to each other, but the asymptotic distributional quadratic biases of the hat estimators are all slightly smaller than those of the corresponding tilde estimators, which confirms our discovery in Corollary 3.1. Now, we change Σ_{xx} to a positive matrix with diagonal elements all 1's and off-diagonal elements all 0.1's, other quantities remain unchanged. In contrast to plot (a), plot (b) in Fig. 1 shows an inverse direction, that is, the asymptotic distributional quadratic biases of the tilde estimators are all slightly smaller than those of the corresponding hat estimators. This phenomenon shows, in the sense of asymptotic distributional quadratic bias, that neither the hat estimators nor the tilde estimators can dominate the others over all scenarios.

The asymptotic weighted risk functions of $\hat{\beta}_{AC}$, the hat and tilde estimators are summarized in the following theorem.

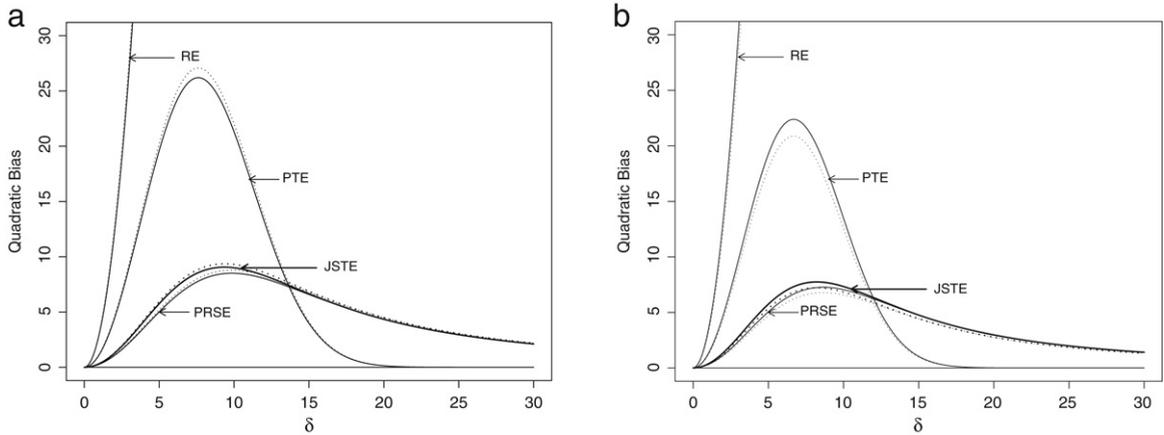


Fig. 1. Bias Plots. The dotted and solid lines are the biases for the tilde estimators and the hat estimators and the horizontal line represents the bias of the AC estimator.

Theorem 3.2. Suppose Conditions C1–C3 hold, then

$$\begin{aligned} \rho(\beta, \widehat{\beta}_{AC}) &= \text{tr}(MG), \\ \rho(\beta, \widehat{\beta}_{RE}) &= \text{tr}(MG) - 2\text{tr}(RGMJ_H) + \text{tr}[J'_H M J_H (RGR')] + \delta' J'_H M J_H \delta, \\ \rho(\beta, \widehat{\beta}_{PTE}) &= \text{tr}(MG) - 2\text{tr}(RGMJ_H)g_{q+2}(\lambda) + 2\delta' J'_G M J_H \delta [g_{q+2}(\lambda) - g_{q+4}(\lambda)] \\ &\quad + \text{tr}[J'_H M J_H (RGR')]g_{q+2}(\lambda) + \delta' J'_H M J_H \delta g_{q+4}(\lambda), \end{aligned}$$

while the risk functions of $\widehat{\beta}_{JSTE}$, $\widehat{\beta}_{PRSE}$ have similar forms as the following

$$\begin{aligned} \rho(\beta, \widehat{\beta}_n^*) &= \text{tr}(MG) - 2\text{tr}(RGMJ_H)f_{1,q+2}(\lambda) + 2\delta' J'_G M J_H \delta [f_{1,q+2}(\lambda) - f_{1,q+4}(\lambda)] \\ &\quad + \text{tr}[J'_H M J_H (RGR')]f_{2,q+2}(\lambda) + \delta' J'_H M J_H \delta f_{2,q+4}(\lambda). \end{aligned}$$

The risk function of $\widehat{\beta}_{JSTE}$ corresponds to $f = h$, and $\widehat{\beta}_{PRSE}$ corresponds to $f = k$.

The risk function comparisons can be done by investigating the matrices G and H , but the actual comparison may be complicated, if possible, because $G = (\sigma^2 + \beta' \Sigma_{uu} \beta)H + \Sigma_{xx}^{-1}[E(uu' \beta \beta' uu') - \Sigma_{uu} \beta \beta' \Sigma_{uu} - \Sigma_{uu}(\beta' \Sigma_{uu} \beta)]\Sigma_{xx}^{-1}$, and the risk function involves inverse operations such as $(RGR')^{-1}$. This comparison can be done in a straightforward way for some special cases. For example, if $M = H^{-1}$, we have the following corollary.

Corollary 3.2. Suppose Conditions C1–C3 hold. Let $M = H^{-1}$. Then

$$\begin{aligned} \rho(\beta, \widehat{\beta}_{AC}) &= \text{tr}(H^{-1}G), \quad \rho(\beta, \widehat{\beta}_{RE}) = \text{tr}(H^{-1}G) - \text{tr}[(RHR')^{-1}RGR'] + \delta'(RHR')^{-1}\delta, \\ \rho(\beta, \widehat{\beta}_{PTE}) &= \text{tr}(H^{-1}G) - \text{tr}[(RHR')^{-1}RGR']g_{q+2}(\lambda) + \delta'(RHR')^{-1}\delta[2g_{q+2}(\lambda) - g_{q+4}(\lambda)], \end{aligned}$$

the risk functions of $\widehat{\beta}_{JSTE}$, $\widehat{\beta}_{PRSE}$ have similar forms as the following

$$\begin{aligned} \rho(\beta, \widehat{\beta}_n^*) &= \text{tr}(H^{-1}G) - \text{tr}[(RHR')^{-1}RGR'] [2f_{1,q+2}(\lambda) - f_{2,q+2}(\lambda)] \\ &\quad + \delta'(RHR')^{-1}\delta [2f_{1,q+2}(\lambda) - 2f_{1,q+4}(\lambda) + f_{2,q+4}(\lambda)]. \end{aligned}$$

The risk function of $\widehat{\beta}_{JSTE}$ corresponds to $f = h$, and $\widehat{\beta}_{PRSE}$ corresponds to $f = k$.

Based on Corollary 3.2, one can make the comparisons among the estimators more specifically. In fact,

- the pretest estimator $\widehat{\beta}_{PTE}$ performs better than the AC estimator $\widehat{\beta}_{AC}$ if and only if δ satisfies

$$\delta'(RHR')^{-1}\delta \leq \frac{g_{q+2}(\lambda)\text{tr}[(RHR')^{-1}RGR']}{2g_{q+2}(\lambda) - g_{q+4}(\lambda)}.$$

In particular, under H_0 , that is $\delta = 0$, then

$$\rho(\beta, \widehat{\beta}_{AC}) - \rho(\beta, \widehat{\beta}_{PTE}) = g_{q+2}(0)\text{tr}[(RHR')^{-1}RGR'] > 0.$$

- For a given α , PTE is not uniformly better than the AC estimator. One may determine an α such that PTE has a minimum guaranteed relative efficiency. Similarly to [13,2], the relative efficiency of $\widehat{\beta}_{PTE}$ to $\widehat{\beta}_{AC}$ is defined as $E(\alpha, \lambda) = \rho(\beta, \widehat{\beta}_{AC})/\rho(\beta, \widehat{\beta}_{PTE})$. For any given R, r, G, H , the relative efficiency is a function of α and λ . Suppose the minimum

Table 1
Maximum and minimum guaranteed efficiencies

α	$M = H^{-1}$			$M = G^{-1}$		
	δ_{\min}	min	max	δ_{\min}	min	max
0.05	13.8207	0.7068	1.9531	13.9085	0.7234	1.9742
0.10	12.9342	0.7788	1.7247	13.0384	0.7935	1.7388
0.15	12.3935	0.8241	1.5724	12.5078	0.8370	1.5825
0.20	11.9986	0.8570	1.4611	12.1204	0.8683	1.4687
0.25	11.6853	0.8826	1.3754	11.8126	0.8924	1.3812
0.30	11.4240	0.9033	1.3073	11.6000	0.9118	1.3118

efficiency required is E_0 , then we can choose α by solving the equation $\min_{\lambda} E(\alpha, \lambda) = E_0$. The explicit solution may not be available, but we can use a numerical method to search for the minimization. That is, we compute the corresponding $\min_{\lambda} E(\alpha, \lambda)$ for several α values, and select one such that $\min_{\lambda} E(\alpha, \lambda)$ is close to but not less than E_0 . Now we use the setup for Fig. 1 to illustrate the choice of α . The weight matrix M is chosen to be H^{-1} and G^{-1} , respectively. Table 1 reports, for each α , the maximum relative efficiency (denoted by max), the minimal relative efficiency (denoted by min) and also the value of δ corresponding to the minimal relative efficiency, denoted by δ_{\min} .

So, in our cases, if PTE of β is chosen with at least 0.80 relative efficiency compared to the AC estimator, in both cases ($M = H^{-1}$ and $M = G^{-1}$), we choose $\alpha = 0.15$ as the level of the test. In practice, G and H are unknown. To implement this procedure, one needs to have some preliminary estimates for these quantities.

- The risk function of $\hat{\beta}_{\text{JSTE}}$ can be written as

$$\rho(\beta, \hat{\beta}_{\text{JSTE}}) = \text{tr}(H^{-1}G) - (q - 2)\text{tr}[(RHR')^{-1}RGR'] \times \left((q - 2)E(\chi_{q+2,\lambda}^{-4}) + 2\lambda \left\{ 1 - \frac{(q + 2)\delta'(RHR')^{-1}\delta}{2\lambda \text{tr}[(RHR')^{-1}RGR']} \right\} E\chi_{q+4,\lambda}^{-4} \right),$$

which can be shown by using the definition of h -function and the following facts on χ^2 -distributions:

$$\begin{aligned} \lambda E[(\chi_{q+4,\lambda}^2)^{-2}] &= E[(\chi_{q+2,\lambda}^2)^{-1}] - (q - 2)E[(\chi_{q+2,\lambda}^2)^{-2}], \\ 2E[(\chi_{q+4,\lambda}^2)^{-2}] &= E[(\chi_{q+2,\lambda}^2)^{-1}] - E[(\chi_{q+4,\lambda}^2)^{-1}]. \end{aligned}$$

See (2.2.13d) and (2.2.13e) in [2].

A sufficient condition that ensures $\hat{\beta}_{\text{JSTE}}$ being superior to $\hat{\beta}_{\text{AC}}$ is given by

$$\frac{\text{tr}[(RHR')^{-1}RGR']}{\Lambda_{\max}[(RHR')^{-1}RGR']} \geq \frac{q + 2}{2},$$

where $\Lambda_{\max}(\cdot)$ is the maximal eigenvalue of its matrix argument. If, in addition, we assume that $u \sim N(0, \sigma_{uu}I)$, $\Sigma_{xx} = \sigma_{xx}I$, then the above sufficient condition can be written as $(c_1 - c_2)q \geq 2c_1$, where $c_1 = \sigma^2 + \sigma_{uu}\beta'\beta$, $c_2 = \beta'R(RR')^{-1}R\beta\sigma_{uu}^2/(\sigma_{xx} + \sigma_{uu})$. Furthermore, if $c_2 = 0$, this sufficient condition is simplified as $q \geq 2$. This, together with the following fact, implies that PRSE and JSTE dominate the AC estimator when $c_2 = 0$, the case studied by Kim and Saleh [13].

- The difference $\rho(\beta, \hat{\beta}_{\text{JSTE}}) - \rho(\beta, \hat{\beta}_{\text{PRSE}})$ is

$$\begin{aligned} &\text{tr}[(RHR')^{-1}RGR'] \cdot E\{[1 - (q - 2)\chi_{q+2,\lambda}^{-2}]^2 I(\chi_{q+2,\lambda}^2 < q - 2)\} \\ &+ \delta'(RHR')^{-1}\delta E[(1 - (q - 2)\chi_{q+4,\lambda}^{-2})^2 I(\chi_{q+4,\lambda}^2 < q - 2)] \\ &+ 2\delta'(RHR')^{-1}\delta E[(q - 2)\chi_{q+2,\lambda}^{-2} - 1] I(\chi_{q+2,\lambda}^2 < q - 2), \end{aligned}$$

which is non-negative for all δ , and implies that the performance of PRSE is uniformly better than that of JSTE.

- The difference $\rho(\beta, \hat{\beta}_{\text{JSTE}}) - \rho(\beta, \hat{\beta}_{\text{PTE}})$ is

$$\begin{aligned} &\text{tr}[(RHR')^{-1}RGR'] \cdot [g_{q+2}(\lambda) - 2h_{1,q+2}(\lambda) + h_{2,q+2}(\lambda)] \\ &+ \delta'(RHR')^{-1}\delta [2h_{1,q+2}(\lambda) - 2h_{1,q+4}(\lambda) + h_{2,q+4}(\lambda) - 2g_{q+2}(\lambda) + g_{q+4}(\lambda)]. \end{aligned}$$

Under the null hypothesis ($\delta = 0$),

$$\rho(\beta, \hat{\beta}_{\text{JSTE}}) - \rho(\beta, \hat{\beta}_{\text{PTE}}) = \text{tr}[(RHR')^{-1}RGR'] \cdot \left[P(\chi_{q+2}^2 < \chi_{\alpha}^2) - \frac{q - 2}{q} \right].$$

Thus, in the case of $R\beta$ being close to r , at the significance level α , PTE should be used if $qP(\chi_{q+2}^2 < \chi_{\alpha}^2) \geq q - 2$, otherwise, JSTE is preferable.

- The difference $\rho(\beta, \hat{\beta}_{\text{PRSE}}) - \rho(\beta, \hat{\beta}_{\text{PTE}})$ is

$$\begin{aligned} &\text{tr}[(RHR')^{-1}RGR'] \cdot [g_{q+2}(\lambda) - 2k_{1,q+2}(\lambda) + k_{2,q+2}(\lambda)] \\ &+ \delta'(RHR')^{-1}\delta [2k_{1,q+2}(\lambda) - 2k_{1,q+4}(\lambda) + k_{2,q+4}(\lambda) - 2g_{q+2}(\lambda) + g_{q+4}(\lambda)]. \end{aligned}$$

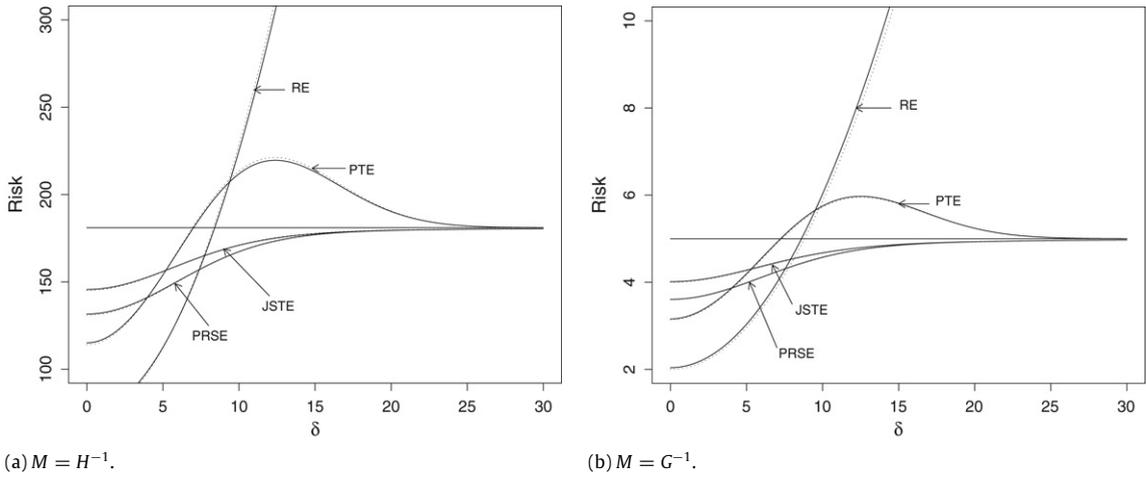


Fig. 2. Risk Plots. The dotted and solid lines are the risks for the tilde estimators and the hat estimators and the horizontal line is the risk of the AC estimator.

Under the null hypothesis ($\delta = 0$), $\rho(\beta, \hat{\beta}_{PRSE}) - \rho(\beta, \hat{\beta}_{PTE})$ is equal to

$$\text{tr}[(RHR')^{-1}RGR'] \cdot \left[P(\chi_{q+2}^2 < \chi_\alpha^2) - \frac{q-2}{q} - E[(1 - (q-2)\chi_{q+2}^{-2})^2 I(\chi_{q+2}^2 < q-2)] \right].$$

Thus, in the case of $R\beta$ being close to r , at the significance level α , PTE should be used if $qP(\chi_{q+2}^2 < \chi_\alpha^2) \geq q-2 + qE[(1 - (q-2)\chi_{q+2}^{-2})^2 I(\chi_{q+2}^2 < q-2)]$, otherwise, PRSE is preferable.

By applying the above theorem with H replaced by G , we can obtain the asymptotic weighted risk functions for all tilde estimators. A similar risk comparison analysis can be done as above, which we summarize in the following theorem.

Theorem 3.3. Suppose Conditions C1–C3 hold, then

$$\begin{aligned} \rho(\beta, \tilde{\beta}_{RE}) &= \text{tr}(MG) - \text{tr}(J'_G MGR') + \delta' J'_G M J_G \delta, \\ \rho(\beta, \tilde{\beta}_{PTE}) &= \text{tr}(MG) - \delta' J'_G M J_G \delta g_{q+4}(\lambda) - [\text{tr}(J'_G MGR') - 2\delta' J'_G M J_G \delta] g_{q+2}(\lambda), \end{aligned}$$

the risk functions of $\tilde{\beta}_{JSTE}, \tilde{\beta}_{PRSE}$ have similar forms as the following

$$\rho(\beta, \tilde{\beta}_n^*) = \text{tr}(MG) - \text{tr}(J'_G MGR') (2f_{1,q+2}(\lambda) - f_{2,q+2}(\lambda)) + \delta' J'_G M J_G \delta [2f_{1,q+2}(\lambda) - 2f_{1,q+4}(\lambda) + f_{2,q+4}(\lambda)].$$

The risk functions of $\tilde{\beta}_{JSTE}$ and correspond to $f = h$, and $f = k$, respectively.

In particular, if $M = G^{-1}$, then the above theorem reduces to the following corollary.

Corollary 3.3. Suppose Conditions C1–C3 hold, then

$$\begin{aligned} \rho(\beta, \tilde{\beta}_{RE}) &= p - q + \lambda, & \rho(\beta, \tilde{\beta}_{PTE}) &= p - (q - 2\lambda)g_{q+2}(\lambda) - \lambda g_{q+4}(\lambda), \\ \rho(\beta, \tilde{\beta}_{JSTE}) &= p - q[2h_{1,q+2}(\lambda) - h_{2,q+2}(\lambda)] + \lambda[2h_{1,q+2}(\lambda) - 2h_{1,q+4}(\lambda) + h_{2,q+4}(\lambda)], \\ \rho(\beta, \tilde{\beta}_{PRSE}) &= p - q[2k_{1,q+2}(\lambda) - k_{2,q+2}(\lambda)] + \lambda[2k_{1,q+2}(\lambda) - 2k_{1,q+4}(\lambda) + k_{2,q+4}(\lambda)]. \end{aligned}$$

For the purpose of illustration, we use the previous setting, that is, $p = 5, q = 3, \alpha = 0.15, \beta = (3, 3, 1, 2, 3)'$, $\Sigma_{xx} = \Sigma_{uu} = I_{5 \times 5}, \sigma^2 = 1$, and R is a 3×5 matrix with the components in the first two columns being 0's and the last three columns an identity matrix. For simplicity, δ is taken to be a vector with same elements. We plot the risk functions of the estimators in Fig. 2, in which plot (a) is for $M = H^{-1}$, and plot (b) is for $M = G^{-1}$. In our cases, the risk functions of the hat and tilde estimators for each type of estimator (RE, PTE, JSTE or PRSE) are close to each other.

Sometimes, we are interested in estimating a linear combination of the intercept and the slope parameters. For example, to predict the response at a specified value of the predictor, say $x_0 \in \mathbb{R}^p$, one needs to calculate the value of $\beta_{0n}^* + x_0' \beta_n^*$, where β_{0n}^* and β_n^* are the generic notation for the AC estimator, hat or tilde estimators of β_0 and β . As we mentioned before, the corresponding estimators of the intercept β_0 are given by $\bar{Y} - \bar{W}' \beta_n^*$. For example, the hat AC estimator of β_0 is defined as $\hat{\beta}_{0,AC} = \bar{Y} - \bar{W}' \hat{\beta}_{AC}$, and the tilde RE of β_0 is given by $\tilde{\beta}_{0,RE} = \bar{Y} - \bar{W}' \tilde{\beta}_{RE}$. We can define the AC estimator and the hat, tilde estimators of the linear combination $\beta_0 + x_0' \beta$ accordingly. The risk function of $\beta_{0n}^* + x_0' \beta_n^*$ is then defined as the mean square error of the asymptotic distribution of $\sqrt{n}(\beta_{0n}^* + x_0' \beta_n^* - \beta_0 - x_0' \beta_n)$ which is denoted as $\rho(\beta_0 - x_0' \beta, \beta_{0n}^* + x_0' \beta_n^*)$, where β_n

Table 2
Risk for the slope when the null hypothesis is true

<i>n</i>		AC	RE	PTE	JSTE	PRSE
50	Hat	8.697	3.220	5.307	5.323	4.857
	Tilde	8.697	3.167	5.251	5.285	4.829
100	Hat	8.578	3.293	5.200	5.234	4.808
	Tilde	8.578	3.264	5.177	5.217	4.794
200	Hat	8.153	3.052	4.689	4.868	4.450
	Tilde	8.153	3.037	4.681	4.858	4.442
500	Hat	8.029	3.090	4.629	4.989	4.394
	Tilde	8.029	3.085	4.623	4.986	4.392

is the same as in (2.6). The following theorem gives the risk functions of various estimators of the linear combination. Let $\tau = G^{-1}\Sigma_{xx}^{-1}Eu\beta'uu'\beta - \mu_x, \sigma_0^2 = \sigma^2 + \beta'\Sigma_{uu}\beta + \mu_x'G\mu_x - 2\mu_x'\Sigma_{xx}Eu\beta'uu'\beta$.

Theorem 3.4. Suppose Conditions C1–C3 hold. Then $\rho(\beta_0 - x_0'\beta, \beta_{0n}^* + x_0'\beta_n^*) = \sigma_0^2 + x_0'Gx_0 + 2x_0'G\tau + (x_0 - \mu_x)'J_C[RGR'E\phi^2(\chi_{q+2,\lambda}^2) + \delta\delta' E\phi^2(\chi_{q+4,\lambda}^2)]J_C'(x_0 - \mu_x) - 2(x_0 - \mu_x)'J_C[RGE\phi(\chi_{q+2,\lambda}^2) - \delta\delta'J_G'E\phi(\chi_{q+2,\lambda}^2) + \delta\delta'J_G'E\phi(\chi_{q+4,\lambda}^2)](x_0 + \tau)$, where $C = H$ corresponds to the estimators based on $\hat{\beta}_{RE}$ and $C = G$ corresponds to the estimators based on $\hat{\beta}_{RE}$, $\phi(x) = 0$ for the AC estimator, $\phi(x) = 1$ for the RE, $\phi(x) = I_{(0, \chi_\alpha^2)}(x)$ for PTE, $\phi(x) = (q - 2)x^{-1}$ for JSTE and $\phi(x) = 1 - [1 - (q - 2)x^{-1}]I_{(q-2, \infty)}(x)$ for PRSE.

When the measurement errors are symmetric around 0, $Eu\beta'uu'\beta = 0$ for all β . As a consequence, we have the following corollary.

Corollary 3.4. Suppose Conditions C1–C3 hold, and $Eu\beta'uu'\beta = 0$. Then

$$\begin{aligned} \rho(\beta_0 - x_0'\beta, \beta_{0n}^* + x_0'\beta_n^*) &= \sigma_0^2 + (x_0 - \mu_x)'J_C RGR'J_C'E\phi^2(\chi_{q+2,\lambda}^2) \\ &\quad + J_C \delta\delta'J_G'E\phi^2(\chi_{q+4,\lambda}^2)(x_0 - \mu_x) - 2(x_0 - \mu_x)'J_C[RGE\phi(\chi_{q+2,\lambda}^2) \\ &\quad - \delta\delta'J_G'E\phi(\chi_{q+2,\lambda}^2) + \delta\delta'J_G'E\phi(\chi_{q+4,\lambda}^2)](x_0 - \mu_x), \end{aligned}$$

where $\sigma_0^2 = \sigma^2 + \beta'\Sigma_{uu}\beta + \mu_x'G\mu_x$, C and $\phi(\cdot)$ are as in Theorem 3.4.

4. Simulation studies

To see the finite-sample performance of the proposed estimators, we conduct simulation experiments under the various scenarios. The data are generated from the following multiple linear regression model with measurement errors

$$\begin{cases} Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + \beta_8X_8 + \varepsilon, \\ W_j = X_j + u_j, \quad j = 1, 2, \dots, 8. \end{cases} \tag{4.1}$$

The true regression parameter is chosen to be $\beta_n = \beta + \delta\mathbf{1}/\sqrt{n}$ with $R\beta = r$. We consider two cases, $\delta = 0$, and $\delta \neq 0$. For each case, we calculate the risk function for sample size $n = 50, 100, 200$, and 500 , and repeat each simulation 1000 times. The values of the risk reported in the tables below are the average of 1000 sample risks. In the simulation, α is chosen to be 0.15. In practice, one can choose the value of α based on the maximin rule given in Table 1. The predictors X are generated from multivariate normal $N(0, I_{8 \times 8})$, and the measurement errors u are generated from multivariate normal $N(0, 0.2^2I_{8 \times 8})$. X and u are independent. The regression error ε follows $N(0, 1)$.

Case 1: When $\delta = 0$ or $R\beta_n = r$ holds.

The true values of the regression parameters are chosen to be $\beta_0 = 1, \beta_1 = \beta_5 = \beta_7 = 0, \beta_2 = 1.5, \beta_3 = 0.75, \beta_4 = -\beta_6 = 2, \beta_8 = 3$. So the β 's satisfy the constraint $R\beta = r$ with

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad r = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{4.2}$$

Table 2 reports the risk values of various slope estimators.

In Table 2, ‘‘hat’’ denotes the hat estimators, and ‘‘tilde’’ denotes the tilde estimators. Each cell gives the risk value with weight matrix G^{-1} . The risks with weight matrix H^{-1} have a similar pattern and therefore are omitted. In this simulation, one can see that the risks are pretty stable within each class of estimators, also the risks of the hat estimators and of the

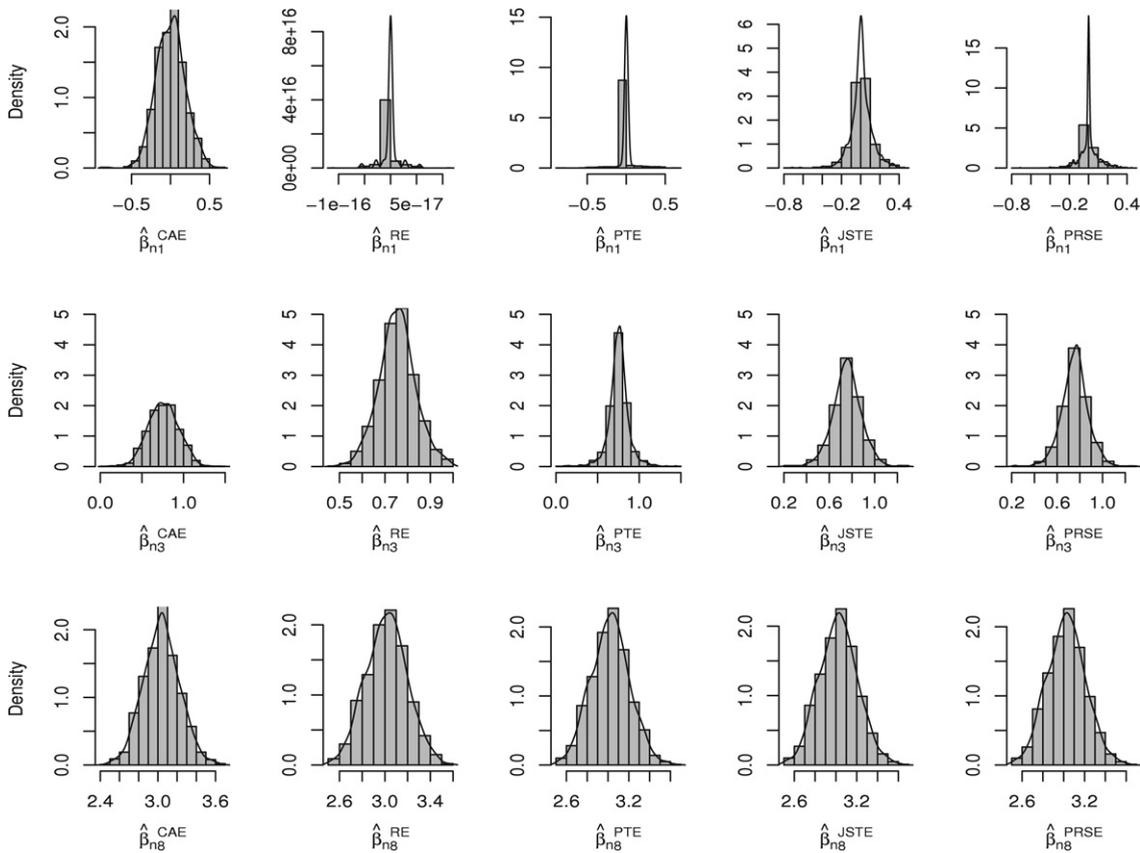


Fig. 3. Histograms for the estimates of the slope parameters.

tilde estimators are very close. The smallest risk is achieved by RE, then PTE, PRSE, JSTE, and the largest risk is obtained by the AC estimator. These results coincide with our theory. For an illustrative purpose, the histograms of the estimates, with the weight matrix $M = H^{-1}$, of the slope parameter for $n = 200$ are given in Fig. 3.

Case 2: When $\delta \neq 0$ or $R\beta_n = r$ does not hold.

In this case, $R\beta_n - r = \delta R\mathbf{1}/\sqrt{n}$, where $\mathbf{1}$ is a 8×1 vector with elements all 1's. Other quantities in the model remain unchanged. Fig. 4 reports the risks of the hat estimators with weight matrix G^{-1} for different sample sizes. The δ value ranges from 0 to 10. The risks of the hat estimators have a similar pattern regardless of the choice of weight matrix.

The risk of the AC estimator for the slope is almost constant for various δ . When the values of δ are close to 0, the RE for both slopes and intercept achieves the smallest risk, but its risk increases quickly when δ gets bigger. PTE for both slopes and intercept has smaller risk when δ is smaller, once δ leaves 0, the risk of PTE begins to increase and exceeds the risk of the AC estimator. After it hits a certain point, it comes down and eventually approaches the risk of the AC estimator. For small δ values, the risks of JSTE and PRSE for the slope estimators are higher than that of the RE. However, when δ gets bigger, JSTE and PRSE begin to dominate all other estimators. The patterns of the risks of the AC estimator, RE and PTE for the intercept are similar to those of the AC estimator, RE and PTE for the slopes, but this is not true for the risks of JSTE and PRSE for the intercept.

Upon the request of one referee, we also conduct a simulation study when X and u follow non-normal distributions. Each component of the predictor X is generated from a uniform distribution on $[1, 2]$, and each component of the measurement error u is generated from a uniform distribution on $[-0.5, 0.5]$, $X_1, \dots, X_8, u_1, \dots, u_8$ are independent. All other entities are the same as in the normal case. The simulation results are similar to the normal cases and are not reported here.

5. Real data example

The assessment of an individual diet is difficult, but fundamental in exploring the relationship between diet and cancer, and in monitoring dietary behavior among individuals and populations. A variety of dietary assessment instruments have been derived, of which three main types are most commonly used in nutritional research. The instrument of choice in large nutritional epidemiology studies is the Food Frequency Questionnaire (FFQ). For proper interpretation of epidemiologic studies that use FFQs as the basic dietary instrument, one needs to know the relationship between reported intakes from the FFQ, usual intake, energy, vitamin A, and other variables such as age and body mass index (bmi).

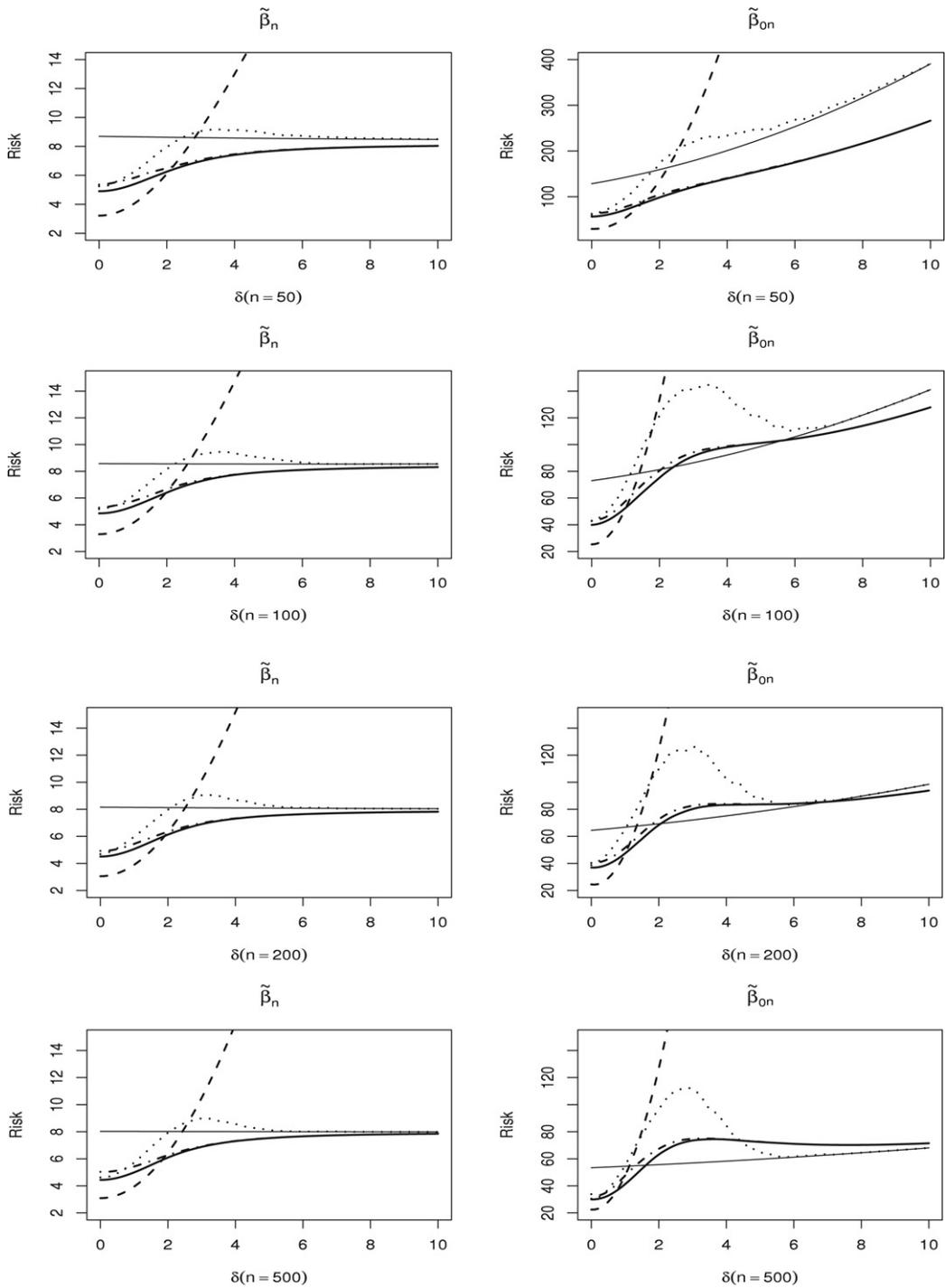


Fig. 4. Risk Functions. Thin solid line: risk of the AC estimator; dashed line: risk of RE; dotted line: risk of PTE; dash-dotted line: risk of JSTE; thick solid line: risk of PRSE.

FFQs are thought to often involve a systematic bias (i.e., under- or over-reporting at the level of the individual). The other records also include measurement errors. To illustrate the proposed method, we analyze a dataset from the Nurses Health Study [20], which has a calibration study of size $n = 168$ women. All of them completed a single FFQ and four multiple-day food diaries. There are 6 variables, age (X_1), bmi (X_2), energy (X_3), vitamin A (X_4), usual intake (X_5), and the calories from fat, FFQ (Y), in this dataset. Among these 6 variables, energy, usual intake, vitamin A are measured with error, but for each subject, these 3 variables are measured four times. A simple variance analysis suggests that the variance of energy is 3.63, the variance of vitamin A is 381.92, and the variance of usual intake is 10.34. For an initial analysis, the following multiple

Table 3
Estimates of the slope parameters with standard errors and associated *p*-values

	Estimate	Estimated standard error	<i>p</i> -value
β_1	0.1386	0.0719	0.0556*
β_2	−0.4412	0.1388	0.0018*
β_3	−0.0362	0.0896	0.6867
β_4	−0.5345	0.7306	0.4655
β_5	1.1141	0.3279	0.0009*

Table 4
Estimates and bootstrap standard deviations

	β_1	β_2	β_3	β_4	β_5
AC	0.1386 (0.0568)	−0.4411 (0.1099)	−0.0362 (0.0643)	−0.5345 (0.5101)	1.1141 (0.2428)
RE	0.1748 (0.0393)	−0.4034 (0.0858)			1.2495 (0.1096)
PTE	0.1748 (0.0415)	−0.4034 (0.0859)			1.2495 (0.1115)

regression model is used to fit this dataset.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \varepsilon.$$

The averages of these four replications in X_3, X_4, X_5 are used in the design matrix. The covariance matrix of the measurement errors is estimated by using the formula in [21].

With all variables in the model and without constraint, Table 3 lists the estimated values based on the AC estimation for the slope, the estimated standard errors based on the procedure developed by Liang et al. [21], and the associated *p*-values calculated from *t*-distribution with degrees of freedom $n - 6 = 168 - 6 = 162$.

Table 3 shows that X_3 and X_4 are not significant at 0.1 significant level, while variables X_1, X_2 , and X_5 are significant. Recall that X_5 represents usual intake, which is strongly related to intakes from the FFQ. On the other hand, vitamin A should not be a good predictor of food composition. Thus, using advanced statistical methods for getting a reasonable estimate but weighting towards $\beta_3 = \beta_4 = 0$ makes a lot of sense for nutritional research.

Now we impose a constraint on our model $\beta_3 = \beta_4 = 0$. In this case $q = 2$. Table 4 reports the estimated values, obtained by calculating the estimators we are studying, for the regression parameters. To measure the variation of these estimates, the bootstrap standard errors are also reported here. For each estimation procedure, 1000 bootstrap samples are drawn. Accordingly, 1000 bootstrap estimates for the slope parameters are obtained. To get a robust estimation for the standard errors, only the middle 80%, or 800 estimates are used in the calculation. The resulting bootstrap standard errors are shown in the brackets.

Note that the bootstrap standard errors of AC in Table 4 are close to the standard errors of AC in Table 3. Also, the standard errors of RE, PTE are all smaller than their counterparts of AC. These results clearly show that the proposed estimators improve upon the usual AC estimator.

6. Discussion

We have introduced two classes of estimators, the hat and tilde estimators, and made a comprehensive comparison of their risk functions in some special cases. The comparison in general cases is more complicated, and it is difficult to say which estimators should be used in practice, unless further information is available, such as the values of R, r, G, H etc. Usually, the practitioner may have some prior information about R and r . Also, once the sample is obtained, one can estimate G and H . Based on our comparison (see Figs. 1 and 2), the quadratic bias function and the risk functions of the hat estimators and the tilde estimators for the slope parameters do not differ substantially, even for different weight matrices. Our analysis indicates that if the prior knowledge about the regression parameters is true, RE, PTE, JSTE and PRSE all have smaller risks than the AC estimator. When the regression parameters deviate from the prior information, RE becomes useless eventually, while PTE behaves quite well, except for some medium departure from the null hypothesis. If the slope parameters are of interest, JSTE and PRSE are highly recommended in that they possess smaller risk than the AC estimator, RE and PTE. In particular, PRSE dominates JSTE in some special cases, see the discussions following Corollary 3.1.

The procedure developed in this paper can easily be extended to the case where Σ_{uu} is unknown, but we have a consistent estimator of Σ_{uu} .

It is also possible to extend the procedure to the partially linear models $Y = X'\beta + \nu(z) + \varepsilon$ with error-prone linear covariate X . The major work can be regarded as a combination of this paper and the work of Liang et al. [21], because the latter has already derived a root- n consistent estimator of the parameter β .

In principle, the method proposed in this paper can also be extended to linear or partially linear models for longitudinal data. The derivation should be straightforward except for more complex notation.

Acknowledgments

The authors thank two referees for their helpful comments which improved the presentation of this manuscript. They also thank Dr. Raymond J. Carroll for very helpful comments. This research was partially supported by NIH/NIAID grants AI62247 and AI059773.

Appendix. Proofs of the main results

Proof of Theorem 2.1. A direct derivation yields $\hat{\beta}_{AC} - \beta_n = (S_{WW} - \Sigma_{uu})^{-1}(S_{WY} - S_{WW}\beta_n + \Sigma_{uu}\beta_n)$, and $S_{WY} - S_{WW}\beta_n = \frac{1}{n} \sum_{i=1}^n (W_i - \mu_x)(\varepsilon_i - u_i'\beta_n) - (\mu_x - \bar{W})\bar{u}'\beta_n + O_p(1/n)$. Let $\xi_{in} = (X_i - \mu_x)(\varepsilon_i - u_i'\beta_n) + u_i\varepsilon_i + (u_i u_i' - \Sigma_{uu})\beta_n$. It follows that

$$\sqrt{n}(\hat{\beta}_{AC} - \beta_n) = (S_{WW} - \Sigma_{uu})^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_{in} - \sqrt{n}(\mu_x - \bar{W})\bar{u}'\beta_n + O_p(1/\sqrt{n}) \right].$$

Note that the last two terms approach zero and $S_{WW} - \Sigma_{uu} \rightarrow \Sigma_{xx}$ in probability. **Theorem 2.1** follows for the AC estimator by applying the central limit theorem, together with the fact that $\beta_n \rightarrow \beta$.

For all other hat estimators, we have the general form $\hat{\beta}^* = \hat{\beta}_{AC} - (\hat{\beta}_{AC} - \hat{\beta}_{RE})\phi(L_n)$. Recall $\hat{\beta}_{RE} = \hat{\beta}_{AC} - \hat{H}_n R' (R\hat{H}_n R')^{-1} (R\hat{\beta}_{AC} - r)$. It follows that $\hat{\beta}^* = \hat{\beta}_{AC} - \phi(L_n)\hat{H}_n R' (R\hat{H}_n R')^{-1} (R\hat{\beta}_{AC} - r)$. Under the local alternative $R\beta_n = r + \delta/\sqrt{n}$, we know that

$$\sqrt{n}(\hat{\beta}^* - \beta_n) = \sqrt{n}(\hat{\beta}_{AC} - \beta_n) - \phi(L_n)\hat{H}_n R' (R\hat{H}_n R')^{-1} [R\sqrt{n}(\hat{\beta}_{AC} - \beta_n) + \delta].$$

So the result for hat estimators follows from the fact that $\hat{H}_n R' (R\hat{H}_n R')^{-1} \rightarrow J_H$ in probability, and $\sqrt{n}(\hat{\beta}_{AC} - \beta_n) \rightarrow Z$ in distribution. In the same way, we can prove the theorem for all tilde estimators. \square

To prove **Corollary 3.1** and compute the risk function of other estimators, we need the following lemmas.

Lemma A.1 ([2]). *If the $p \times 1$ vector Y is distributed normally with mean vector μ and covariance matrix $I_{p \times p}$, then for any measurable function ϕ ,*

$$E[\phi(Y'Y)YY'] = E[\phi(\chi_{(p+2, \mu'\mu)}^2)]I_{p \times p} + E[\phi(\chi_{(p+4, \mu'\mu)}^2)]\mu\mu' \quad (\text{A.1})$$

$$E[\phi(Y'Y)Y] = \mu E[\phi(\chi_{(p+2, \mu'\mu)}^2)] \quad (\text{A.2})$$

provided that the expectations exist.

Lemma A.2. *If the $p \times 1$ vector Y is distributed normally with mean vector μ and covariance matrix $I_{p \times p}$, and A is an idempotent matrix with rank $q \leq p$, then for any measurable function ϕ , we have*

$$E[\phi(Y'AY)YY'] = (I - A)(I + \mu\mu')(I - A)h_0 + (A + A\mu\mu' + \mu\mu'A - 2A\mu\mu'A)h_2 + A\mu\mu'Ah_4 \quad (\text{A.3})$$

and

$$E\phi(Y'AY)Y = A\mu h_2 + (I - A)\mu h_0 \quad (\text{A.4})$$

provided the expectations exist, where $h_i = E\phi(\chi_{(q+i, \lambda)}^2)$, $\lambda = \mu'A\mu$.

Proof. Let P be the $p \times p$ orthogonal matrix such that $PAP' = \text{Blockdiag}(I_{q \times q}, 0_{(p-q) \times (p-q)})$, and $P = (P_1', P_2')'$, P_1 is $q \times p$, and P_2 is $(p - q) \times p$. Let $Z = PY$, then $Y'AY = Z'\text{Blockdiag}(I, 0)Z$. Partition Z into two blocks $Z = (Z_1', Z_2')'$. It is easy to see $Z_1 \sim N(P_1\mu, I_{q \times q})$, $Z_2 \sim N(P_2\mu, I_{(p-q) \times (p-q)})$, Z_1 and Z_2 are independent. Hence $Y'AY = Z_1'Z_1$, $YY' = P'ZZ'P$, and

$$E[\phi(Y'AY)YY'] = E[\phi(Z_1'Z_1)P'ZZ'P] = P'E \left[\phi(Z_1'Z_1) \begin{pmatrix} Z_1Z_1' & Z_1Z_2' \\ Z_2Z_1' & Z_2Z_2' \end{pmatrix} \right] P.$$

By (A.1), $E[\phi(Z_1'Z_1)Z_1Z_1'] = E[\phi(\chi_{(q+2, \mu'P_1'\mu)}^2)]I_{q \times q} + E[\phi(\chi_{(q+4, \mu'P_1'\mu)}^2)]P_1\mu\mu'P_1'$. By the independence of Z_1 and Z_2 and (A.2),

$$E[\phi(Z_1'Z_1)Z_1Z_2'] = E[\phi(Z_1'Z_1)Z_1]EZ_2' = P_1\mu\mu'P_2'E[\phi(\chi_{(q+2, \mu'P_1'\mu)}^2)],$$

$$E[\phi(Z_1'Z_1)Z_2Z_2'] = E[\phi(Z_1'Z_1)]E(Z_2Z_2') = (I + P_2\mu\mu'P_2')E[\phi(\chi_{(q, \mu'P_1'\mu)}^2)].$$

Hence $E[\phi(Y'AY)YY']$ is equal to

$$\begin{aligned} & (P_2'P_2 + P_2'P_2\mu\mu'P_2')E[\phi(\chi_{(q,\mu'P_1\mu)}^2)] + (P_1'P_1 + P_1'P_1\mu\mu'P_2'P_2 \\ & + P_2'P_2\mu\mu'P_1'P_1)E[\phi(\chi_{(q+2,\mu'P_1\mu)}^2)] + P_1'P_1\mu\mu'P_1'P_1E[\phi(\chi_{(q+4,\mu'P_1\mu)}^2)]. \end{aligned} \tag{A.5}$$

Thus (A.3) is obtained by using the facts that $P_1'P_1 = A$, $P_2'P_2 = I - A$.

To prove (A.4), one can show that $E\phi(Y'AY)Y = P_1'P_1\mu h_2 + P_2'P_2\mu h_0$. Again $P_1'P_1 = A$, and $P_2'P_2 = I - A$ imply the desired result. \square

Proof of Corollary 3.1. It is sufficient to show that $J_H'J_H \leq J_G'J_G$. From the normality of u and Lemma A.2, one can show that

$$G = (\sigma^2 + \beta' \Sigma_{uu} \beta) \Sigma_{xx}^{-1} \Sigma_{WW} \Sigma_{xx}^{-1} + \Sigma_{xx}^{-1} \Sigma_{uu} \beta \beta' \Sigma_{uu} \Sigma_{xx}^{-1}.$$

Let $c_1 = (\sigma^2 + \sigma_{uu}\beta'\beta)(\sigma_{xx} + \sigma_{uu})/\sigma_{xx}^2$, $c_2 = \sigma_{uu}^2/\sigma_{xx}^2$. The diagonal form of the covariance matrices of x and u implies $G = c_1I + c_2\beta\beta'$. Then

$$RGGR' = R[c_1I + c_2\beta\beta']^2R' = c_1^2RR' + 2c_1c_2R\beta\beta'R' + c_2^2(\beta'\beta)R\beta\beta'R'. \tag{A.6}$$

Note that $H = (\sigma_{xx} + \sigma_{uu})I/\sigma_{xx}^2$, one can obtain that $(RHR')^{-1}RHHR'(RHR')^{-1} = (RR')^{-1}$.

It follows from a direct calculation that

$$\begin{aligned} (RGR')(RHR')^{-1}RHHR'(RHR')^{-1}(RGR') &= R(c_1I + c_2\beta\beta')R'(RR')^{-1}R(c_1I + c_2\beta\beta')R' \\ &= c_1^2RR' + 2c_1c_2R\beta\beta'R' + c_2^2(\beta'R'(RR')^{-1}R\beta)R\beta\beta'R'. \end{aligned} \tag{A.7}$$

Since $R'(RR')^{-1}R$ is an idempotent matrix with rank $q \leq p$, $I - R'(RR')^{-1}R$ is non-negative definite, and then $\beta'R'(RR')^{-1}R\beta \leq \beta'\beta$, (A.6) and (A.7) imply

$$(RGR')(RHR')^{-1}RHHR'(RHR')^{-1}(RGR') \leq RGGR'$$

or

$$(RHR')^{-1}RHHR'(RHR')^{-1} \leq (RGR')^{-1}RGGR'(RGR')^{-1}.$$

Note that the left-hand side is $J_H'J_H$, and the right-hand side is $J_G'J_G$. We complete the proof. \square

Proof of Theorems 3.2 and 3.3. From Theorem 2.1, one can see that the asymptotic distributions for all estimators are the same as that of $Z - \phi(L)CR'(RCR')^{-1}(RZ + \delta)$, where $C = H$ or G , and $\phi(\cdot)$ is defined in Theorem 2.1. We now compute the risk function for this general form. The results in Theorems 3.2 and 3.3 can be obtained by replacing C with H and G , respectively.

Let $Y = G^{-1/2}Z + \mu$, $\mu = G^{1/2}R'(RGR')^{-1}\delta$, $A = G^{1/2}R'(RGR')^{-1}RG^{1/2}$. Then $Y \sim N_p(\mu, I)$, and $L = Y'AY$. We first show that

$$E\phi(L)Z = G^{1/2}\mu[E\phi(\chi_{(q+2,\lambda)}^2) - E\phi(\chi_{(q,\lambda)}^2)], \tag{A.8}$$

and

$$E\phi(L)ZZ' = G^{1/2}[(I - A + \mu\mu')E\phi(\chi_{(q,\lambda)}^2) + (A - 2\mu\mu')E\phi(\chi_{(q+2,\lambda)}^2) + \mu\mu'E\phi(\chi_{(q+4,\lambda)}^2)]G^{1/2}. \tag{A.9}$$

In fact, $E\phi(L)Z$ can be written as $G^{1/2}E\phi(L)Y - G^{1/2}\mu E\phi(L)$. By (A.4) and $E\phi(L) = E\phi(\chi_{(q,\lambda)}^2)$, we obtain that

$$E\phi(L)Z = G^{1/2}A\mu E\phi(\chi_{(q+2,\lambda)}^2) + G^{1/2}(I - A)\mu E\phi(\chi_{(q,\lambda)}^2) - G^{1/2}\mu E\phi(\chi_{(q,\lambda)}^2).$$

Then (A.8) is obtained by noticing that $A\mu = \mu$.

Note that $E\phi(L)ZZ' = G^{1/2}E[\phi(L)(Y - \mu)(Y - \mu)']G^{1/2}$, and $E[\phi(L)(Y - \mu)(Y - \mu)'] = E[\phi(L)YY'] - E[\phi(L)Y]\mu' - \mu E[\phi(L)Y'] + \mu\mu'E[\phi(L)]$. Using (A.3) and (A.4), after some algebra, we prove (A.9).

Now we are ready to compute the risk functions. For any positive definite matrix M , we have $\rho(\beta, \beta_n^*) = E[Z - \phi(L)]_C(RZ + \delta)]'M[Z - \phi(L)]_C(RZ + \delta)]$, which equals

$$E(Z'MZ) - 2E[\phi(L)Z'M]_C(RZ + \delta)] + E[\phi^2(L)]_C(RZ + \delta)]'M]_C(RZ + \delta)].$$

The first term equals $\text{tr}(MG)$. To compute the second and third terms, note that

$$E[\phi(L)(RZ + \delta)Z'] = RGE[E\phi(\chi_{(q+2,\lambda)}^2) - \delta\mu'G^{1/2}E\phi(\chi_{(q+2,\lambda)}^2)] + \delta\mu'G^{1/2}E[\phi(\chi_{(q+4,\lambda)}^2)] \tag{A.10}$$

and

$$E[\phi^2(L)(RZ + \delta)(RZ + \delta)'] = RGR'E[\phi^2(\chi_{(q+2,\lambda)}^2)] + \delta\delta'E[\phi^2(\chi_{(q+4,\lambda)}^2)] \tag{A.11}$$

by (A.8) and (A.9). Correspondingly,

$$\begin{aligned} \rho(\beta, \beta_n^*) &= \text{tr}(MG) - 2\text{tr}\{M_J C E[\phi(L)(RZ + \delta)Z']\} + \text{tr}\{J'_C M_J C E[\phi^2(L)(RZ + \delta)(RZ + \delta)']\} \\ &= \text{tr}MG - 2\text{tr}(RGM_J C)E[\phi(\chi_{q+2,\lambda}^2)] + 2\mu'G^{1/2}M_J C \delta E[\phi(\chi_{q+2,\lambda}^2)] \\ &\quad - 2\mu'G^{1/2}M_J C \delta E[\phi(\chi_{q+4,\lambda}^2)] + \text{tr}(J'_C M_J C RGR')E[\phi^2(\chi_{q+2,\lambda}^2)] + \delta'J'_C M_J C \delta E[\phi^2(\chi_{q+4,\lambda}^2)]. \end{aligned}$$

This completes the proof. \square

Proof of Theorem 3.4. Recall that the AC estimator for β is given by $\widehat{\beta}_{AC} = (S_{WW} - \Sigma_{uu})^{-1}S_{WY}$, and the estimator of β_0 is given by $\widehat{\beta}_0 = \bar{Y} - \bar{W}'\widehat{\beta}_{AC}$. Note that

$$\sqrt{n} \begin{pmatrix} \widehat{\beta}_{0n} - \beta_0 \\ \widehat{\beta}_{AC} - \beta_n \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{bmatrix} \varepsilon_i - u'_i \beta_n - \mu'_x \Sigma_{xx}^{-1}(W_i - \mu_x)(\varepsilon - u'_i \beta_n) - \mu'_x \Sigma_{xx}^{-1} \Sigma_{uu} \beta_n \\ \Sigma_{xx}^{-1}(W_i - \mu_x)(\varepsilon - u'_i \beta_n) + \Sigma_{xx}^{-1} \Sigma_{uu} \beta_n \end{bmatrix} + o_p(1).$$

Let $Z_n = \sqrt{n}(\widehat{\beta}_{AC} - \beta_n)$, $Z_{0n} = \sqrt{n}(\widehat{\beta}_{0n} - \beta_0)$. The multivariate central limit theorem and the fact that $\lim_{n \rightarrow \infty} \beta_n = \beta$ indicate that (Z_{0n}, Z'_n) converges to (Z_0, Z') , a normal vector with mean $(0, 0)'$ and covariance matrix being

$$\begin{pmatrix} \sigma_0^2 & E\beta'uu'\beta u' \Sigma_{xx}^{-1} - \mu'_x G \\ \Sigma_{xx}^{-1} E u \beta' u u' \beta - G \mu_x & G \end{pmatrix}, \tag{A.12}$$

where G is given in (2.5) and β here is subject to $R\beta = r$.

Note that all the proposed estimators for β have a common form $\beta_n^* = \widehat{\beta}_{AC} - \phi(L_n)J_{C_n}(\widehat{R}\widehat{\beta}_{AC} - r)$ with $C_n = \widehat{H}_n$ or \widehat{G}_n . So $\sqrt{n}(\beta_n^* - \beta_n) = Z_n - \phi(L_n)J_{C_n}(RZ_n + \delta)$. From (A.12), we know that $\sqrt{n}(\beta_n^* - \beta_n) \rightarrow Z - \phi(L)J_C(RZ + \delta)$ in distribution. Therefore, for any real vector x_0 ,

$$\begin{aligned} nE[x'_0(\beta_n^* - \beta_n)]^2 &\rightarrow E[x'_0(Z - \phi(L)J_C(RZ + \delta))]^2 \\ &= x'_0 G x_0 - 2x'_0 J_C E[\phi(L)(RZ + \delta)Z']x_0 + x'_0 J_C E\phi^2(L)(RZ + \delta)(RZ + \delta)'J'_C x_0. \end{aligned} \tag{A.13}$$

The estimators for the intercept β_0 , $\beta_{0n}^* = \bar{Y} - \bar{W}'\beta_n^*$, can be expressed as

$$\sqrt{n}(\beta_{0n}^* - \beta_0) = \sqrt{n}(\widehat{\beta}_{0n} - \beta_0) + \sqrt{n}\phi(L_n)\bar{W}'J_{C_n}(\widehat{R}\widehat{\beta}_{AC} - r),$$

which can be written as

$$\sqrt{n}(\beta_{0n}^* - \beta_0) = Z_{0n} + \phi(L_n)\bar{W}'J_{C_n}(RZ_n + \delta),$$

by recalling the notation Z_{0n} and Z_n . From (A.12), we have $\sqrt{n}(\beta_{0n}^* - \beta_0) \rightarrow Z_0 + \phi(L)\mu'_x J_C(RZ + \delta)$, in distribution. Therefore,

$$\begin{aligned} nE(\beta_{0n}^* - \beta_0)^2 &\rightarrow E[Z_0 + \phi(L)\mu'_x J_C(RZ + \delta)]^2 \\ &= \sigma_0^2 + \mu_x J_C E[\phi^2(L)(RZ + \delta)(RZ + \delta)']J'_C \mu_x + 2\mu'_x J_C E[\phi(L)Z_0(RZ + \delta)]. \end{aligned}$$

To deal with the last term, denote $\tau = G^{-1}\Sigma_{xx}^{-1}Eu\beta'uu'\beta - \mu_x$, then $E(Z_0|Z) = \tau'Z$, $Z_0 - E(Z_0|Z)$ and Z are independent, and L depends on Z only. We have

$$\begin{aligned} E[\phi(L)Z_0(RZ + \delta)] &= E\{\phi(L)[Z_0 - E(Z_0|Z) + E(Z_0|Z)](RZ + \delta)\} \\ &= E[\phi(L)(RZ + \delta)Z']\tau. \end{aligned} \tag{A.14}$$

Therefore, $nE(\beta_{0n}^* - \beta_0)^2$ tends to

$$\sigma_0^2 + \mu_x J_C E[\phi^2(L)(RZ + \delta)(RZ + \delta)']J'_C \mu_x + 2\mu'_x J_C E[\phi(L)(RZ + \delta)Z']\tau. \tag{A.15}$$

Finally, we have

$$\begin{aligned} nE[(\beta_{0n}^* - \beta_0)x'_0(\beta_n^* - \beta_n)] &\rightarrow E\{[Z_0 + \phi(L)\mu'_x J_C(RZ + \delta)][Z - \phi(L)J_C(RZ + \delta)]\}x_0 \\ &= x'_0 E(Z_0 Z) - E\phi(L)(RZ + \delta)'Z_0 J'_C x_0 + \mu'_x J_C E\phi(L)(RZ + \delta)Z'x_0 - \mu'_x J_C E\phi^2(L)(RZ + \delta)(RZ + \delta)'J'_C x_0. \end{aligned}$$

From $E(Z_0|Z) = \tau'Z$, we have $E(Z_0 Z) = E(ZZ')\tau = G\tau$. Then by (A.14),

$$\begin{aligned} nE[(\beta_{0n}^* - \beta_0)x'_0(\beta_n^* - \beta_n)] &\rightarrow x'_0 G \tau - \tau' E\phi(L)Z(RZ + \delta)'J'_C x_0 \\ &\quad + \mu'_x J_C E\phi(L)(RZ + \delta)Z'x_0 - \mu'_x J_C E\phi^2(L)(RZ + \delta)(RZ + \delta)'J'_C x_0. \end{aligned} \tag{A.16}$$

A combination of (A.15), (A.13) and (A.16) yields

$$\begin{aligned} \rho(\beta_0 - x'_0 \beta, \beta_{0n}^* + x'_0 \beta_n^*) &= \sigma_0^2 + x'_0 G x_0 + 2x'_0 G \tau + (x_0 - \mu_x)'J_C E\phi^2(L)(RZ + \delta)(RZ + \delta)'J'_C (x_0 - \mu_x) \\ &\quad - 2(x_0 - \mu_x)'J_C E\phi(L)(RZ + \delta)Z'(x_0 + \tau). \end{aligned}$$

The theorem is proved by this expression, (A.10) and (A.11). \square

References

- [1] G.G. Judge, M.E. Bock, *The Statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics*, North Holland Publication, New York, 1978.
- [2] A.K.Md.E. Saleh, *Theory of Preliminary Test and Stein-Type Estimation with Application*, Wiley & Sons, New York, 2006.
- [3] T.A. Bancroft, On biases in estimation due to the use of preliminary tests of significance, *Ann. Math. Statist.* 15 (1944) 190–204.
- [4] A.K.Md.E. Saleh, P.K. Sen, On shrinkage least-squares estimation in a parallelism problem, *Ann. Statist.* 6 (1978) 154–168.
- [5] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, in: *Proceedings of the Third Berkeley Symposium on Mathematics and Statistical Probability*, vol. 1, 1956, pp. 197–206.
- [6] W. James, C. Stein, Estimation with quadratic loss, in: *Proceedings of the Fourth Berkeley Symposium on Mathematics and Statistical Probability*, vol. 1, 1961, pp. 361–379.
- [7] A.K.Md.E. Saleh, P.K. Sen, Non-parametric estimation of location parameter after a preliminary test regression, *Commun. Statist.* 15 (1986) 1451–1466.
- [8] P.K. Sen, A.K.Md.E. Saleh, On preliminary test and shrinkage M-estimation in linear models, *Ann. Statist.* 15 (1987) 1580–1592.
- [9] T.D. Stanley, Stein rule least squares estimation, a heuristic for fallible data, *Econ. Lett.* 20 (1986) 147–150.
- [10] T.D. Stanley, Improved estimators in some linear errors-in-variables models in finite samples, *J. Forecasting* 7 (1988) 103–113.
- [11] Shalabh, Improved estimation in measurement error models through Stein rule procedure, *J. Mult. Anal.* 67 (1998) 35–48.
- [12] H.M. Kim, A.K.Md.E. Saleh, Preliminary test estimators of the parameters of simple linear model with measurement error, *Metrika* 57 (2003) 223–251.
- [13] H.M. Kim, A.K.Md.E. Saleh, Improved estimation of regression parameters in measurement error models, *J. Mult. Anal.* 95 (2005) 273–300.
- [14] L.J. Gleser, The importance of assessing measurement reliability in multivariate regression, *J. Am. Statist. Assoc.* 87 (1992) 696–707.
- [15] P.K. Goel, M.H. DeGroot, Only normal distributions have linear posterior expectations in linear regression, *J. Am. Statist. Assoc.* 75 (1980) 895–900.
- [16] C.R. Rao, Characterization of prior distributions and solutions to a compound decision problem, *Ann. Statist.* 4 (1976) 823–835.
- [17] W.A. Fuller, *Measurement Error Models*, Wiley & Sons, New York, 1987.
- [18] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, *Measurement Error in Nonlinear Models*, 2nd ed., Chapman and Hall, New York, 2006.
- [19] Schneeweiss, Consistent estimation of a regression with errors in the variables, *Metrika* 23 (1976) 101–115.
- [20] B. Rosner, D. Spiegelman, W.C. Willett, Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error, *Am. J. Epidemiol.* 132 (1990) 734–735.
- [21] H. Liang, W. Härdle, R.J. Carroll, Estimation in a semiparametric partially linear errors-in-variables model, *Ann. Statist.* 27 (1999) 1519–1535.