Теорія ймовірностей
та математична статистика
Вип. 97, 2017, с. 196–219

Teoriya Ĭmovirnosteĭ
ta Matematychna Statystyka
No. 97, 2017, pp. 196–219

A COMPARATIVE STUDY FOR TWO NEWLY DEVELOPED ESTIMATORS FOR THE SLOPE IN FUNCTIONAL EIV LINEAR MODEL

A. A. AL-SHARADQAH

Abstract. Two estimators were recently developed in [1] for the slope of a line in the functional EIV model. Both are unbiased, up to order σ^4, where σ is the error standard deviation. One estimator was constructed as a function of the maximum likelihood estimator (MLE). Therefore, it was called Adjusted MLE (AMLE). The second estimator was constructed in a completely different approach. Although both the estimators are unbiased, up to the order σ^4, the latter estimator is much more accurate than the AMLE. We study here these two estimators more rigorously, and we show why one estimator outperforms the other one.

Key words and phrases. Simple linear regression, Errors-in-Variables models, small-noise model, maximum likelihood estimator, bias correction, mean squared errors.

2000 *Mathematics Subject Classification.* 68T10, 68K45, 68K40, 62P30.

1. Introduction

Regression models in which all the variables in the model are subject to errors are known as Errors-In-Variables (EIV) models [8, 10, 14]. In the EIV linear model, the n observed points $\{\mathbf{m}_i = (x_i, y_i)\}_{i=1}^n$ are considered as random perturbations of the *true points* $\tilde{\mathbf{m}}_1 = (\tilde{x}_1, \tilde{y}_1)^\top, \ldots, \tilde{\mathbf{m}}_n = (\tilde{x}_n, \tilde{y}_n)^\top$, i. e.,

$$x_i = \tilde{x}_i + \delta_i, \qquad y_i = \tilde{y}_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where δ_i and ε_i, for each $i = 1, \ldots, n$, are i. i. d. normal random variables with zero mean and variances σ_x^2 and σ_y^2, respectively. The true points are lying on the true line and are defined by

$$\tilde{y}_i = \tilde{\alpha} + \tilde{\beta}\tilde{x}_i, \qquad i = 1, \ldots, n, \tag{1.2}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are the true values of the intercept α and the slope β. This paper is a continuation of our work in [1]; therefore, we will adopt the same assumptions about the true points. That is, we will use the functional model, in which the true points are unknown but fixed. Here we will assume that the ratio $\lambda = \sigma_\varepsilon^2/\sigma_\delta^2$ is known. For simplicity, we write $\sigma_\delta^2 = \sigma^2$ and $\sigma_\varepsilon^2 = \lambda\sigma^2$. In this case, the MLE of (α, β) in the functional model is equivalent to the orthogonal distance regression that minimizes the following:

$$\mathcal{F}_1(\alpha, \beta) = \frac{1}{\beta^2 + \lambda} \sum_{i=1}^n d_i^2, \quad d_i = y_i - \alpha - \beta x_i. \tag{1.3}$$

To minimize this objective function, we first differentiate $\mathcal{F}_1$ with respect to α and substitute its resulting expression $\hat{\alpha}_1 = \bar{y} - \beta\bar{x}$ back into the objective function (1.3). Therefore, we obtain the following:

$$\mathcal{F}_1(\beta) = \frac{1}{\beta^2 + \lambda} \sum d_i^{*2}, \tag{1.4}$$

where $d_i^* = y_i^* - \beta x_i^*$. The notations x_i^* and y_i^* refer to the 'centered' coordinates of x_i and y_i, i. e.,

$$x_i^* = x_i - \bar{x}, \qquad y_i^* = y_i - \bar{y}, \qquad i = 1, \ldots, n. \tag{1.5}$$

Here, we use the standard notation for sample means $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, while for the components of the so-called 'scatter matrix' we use the following:

$$s_{xx} = \sum (x_i - \bar{x})^2, \quad s_{yy} = \sum (y_i - \bar{y})^2, \quad s_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}). \tag{1.6}$$

Therefore, differentiating Eq. (1.4) with respect to β gives the following quadratic equation

$$s_{xy}\beta^2 - (s_{yy} - \lambda s_{xx})\beta - \lambda s_{xy} = 0. \tag{1.7}$$

Equation (1.7) has two distinct roots, but the one that minimizes $\mathcal{F}_1(\beta)$ is

$$\hat{\beta}_1 = \frac{s_{yy} - \lambda s_{xx} + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}}, \tag{1.8}$$

if $s_{xy\neq 0}$ (which is true almost surely). Then, we find $\hat{\alpha}_1 = \bar{y} - \hat{\beta}_1\bar{x}$ [8].

A family of objective functions. Instead of restricting ourselves with only one objective function that led to the MLE, we considered in [1] a general class of objective functions

$$\mathcal{F}(\alpha, \beta) = g(\beta) \sum d_i^2, \quad d_i = y_i - \alpha - \beta x_i, \tag{1.9}$$

where $g(\beta)$ is an arbitrary, smooth positive function of β. This class of objective functions produces two popular estimators. The first estimator is the MLE that minimizes (1.9) whenever $g(\beta) = (\beta^2 + \lambda)^{-1} =: g_1(\beta)$ (say) and the second estimator is the least squares estimator (LS) that minimizes $\mathcal{F}$ in (1.9) whenever the weight function $g(\beta) = 1 =: g_0(\beta)$ (say). It should be clear that $\alpha = \bar{y} - \beta\bar{x}$ and

$$\mathcal{F}(\beta) = g(\beta) \sum d_i^{*2}. \tag{1.10}$$

Another estimator, $\hat{\beta}_2$ that will be discussed in this paper was developed in [1]. The estimator $\hat{\beta}_2$ is the solution that minimizes the objective function $\mathcal{F}$ in (1.10) with the weight

$$g(\beta) = \left(\beta^2 + \lambda\right)^{-\frac{n-3}{n-2}} =: g_2(\beta).$$

With this weight, the new objective function leads to a new estimator $\hat{\beta}_2$. That is, it is the solution, that minimizes (1.10) with the weight $g_2(\beta)$, and it is one of the roots of the cubic equation

$$s_{xx}\beta^3 + (n-4)s_{xy}\beta^2 - \left[(n-3)s_{yy} - \lambda(n-2)s_{xx}\right]\beta - \lambda(n-2)s_{xy} = 0. \tag{1.11}$$

The development of $g_2(\beta)$ comes after deriving the expression of the bias (up to the second-order term) for the estimator that minimizes $\mathcal{F}$ in (1.10). The second-order bias formula depends on g and its derivative, g'. Equating the second-order bias with zero gives us an ordinary first-order linear differential equation (presented shortly). The solution of this differential equation yields $g_2(\beta)$. We called this bias-correction by '*pre-bias elimination technique*', because we choose $g = g_2$ that eliminates the second-order bias in advance.

Moreover, we addressed another bias-correction technique, where the bias is eliminated by subtracting the noisy version of the bias from the estimator itself. This bias correction technique is well known in the literature, but we refer it here by '*post-bias elimination technique*'. The new estimator comes as an adjustment of the MLE for its second-order bias. Indeed, it is a function of the MLE and has the following form

$$\check{\beta}_1 = \left(1 - \frac{\hat{\sigma}^2}{\|\mathbf{x}^*\|^2}\right)\hat{\beta}_1, \tag{1.12}$$

where

$$\hat{\sigma}^2 = \frac{1}{(\hat{\beta}_1^2 + \lambda)(n-2)}\left\|\mathbf{y}^* - \hat{\beta}_1\mathbf{x}^*\right\|_2^2,$$

where $\mathbf{x}^* = (x_1 - \bar{x}, \ldots, x_n - \bar{x})^\top$, $\mathbf{y}^* = (y_1 - \bar{y}, \ldots, y_n - \bar{y})^\top$. Since it is a modified version of the MLE, we called $\check{\beta}_1$ the '*Adjusted Maximum Likelihood Estimator*' (AMLE).

Even though we have derived the general formulas for the bias and the MSE in [1], those formulas work only for estimators minimizing the objective function given in (1.9). The AMLE does not minimize any objective function, those general formulas cannot be applied. Therefore, the higher order bias and the higher order MSE shall be derived for the AMLE in a completely different approach.

The numerical experiments in [1] show that the AMLE, $\check{\beta}_1$, outperforms the MLE, but it still falls behind $\hat{\beta}_2$, although both $\hat{\beta}_2$ and $\check{\beta}_1$ were developed to eliminate the second-order bias. They behave differently in practice. This motivates us to study them further. To do so, we will derive the higher order terms for the bias and the MSE of $\check{\beta}_1$ and $\hat{\beta}_2$, and then we will compare them.

This paper is organized as follows. Section 2 states some statistical assumptions and presents previous results. Those results paves the road for Section 3, where we present the higher order expansions of the bias as well as the MSE of the AMLE, and then we compare between $\hat{\beta}_2$ and the AMLE, $\check{\beta}_1$. This paper involves many technical derivations that are deferred to the appendix.

In this paper, we use vector notations. Accordingly, (1.1) can be expressed as $\mathbf{x} = \tilde{\mathbf{x}} + \boldsymbol{\delta}$ and $\mathbf{y} = \tilde{\mathbf{y}} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ represent the vectors of all noisy errors that corrupt the first and the second coordinates of the true vectors $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)^\top$ and $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top$, respectively.

2. Previous results

Anderson [5] proved that the MLEs of (α, β), i.e. $\hat{\alpha}_1$ and $\hat{\beta}_1$, do not have finite moments, i.e. $\mathsf{E}(|\hat{\alpha}_1|) = \infty$ and $\mathsf{E}(|\hat{\beta}_1|) = \infty$, see also [13]. The infinite first moment's phenomenon is very common in EIV models. For instance, Chernov [11] proved that the most accurate estimator, the MLEs, for the center and the radius of a circle in the circle fitting problem have infinite moments too, while Zelniker and Clarkson [21] proved that the 'awkward' *Delogne–Kåsa* method returns estimators with finite first moments. Moreover, Al-Sharadqan *et al.* show that the first moment for several accurate estimators do not exist [3] either. The infinite first moment problem also appears in other EIV models, such as ellipse fitting [3] and multivariate EIV linear model [9].

Therefore, there is no direct approach to study the statistical properties of these estimators. Traditionally, statisticians investigate the properties of estimators if their moments are finite. If the moments are finite but have complicated formulas, statisticians use the first few terms of the Taylor expansions of their means and their variances. That is, before Anderson's discovery, statisticians employed the Taylor expansion of $(\hat{\alpha}_1, \hat{\beta}_1)$ in order to derive some 'approximate' formulas for the moments of $\hat{\alpha}_1$ and $\hat{\beta}_1$ (including their means and variances). Anderson demonstrated that all those formulas should be regarded as *moments of some approximations* rather than 'approximate moments'.

The MLE of (α_1, β_1) here have infinite variances and infinite mean squared errors! This poses immediate methodological questions: (1) How can we characterize, in practical terms, the accuracy of estimates whose theoretical MSE is infinite (and whose bias is undefined)? (2) Is there any precise meaning to the widely accepted notion that the MLE $\hat{\alpha}_1$ and $\hat{\beta}_1$ are best? To answer these questions, we would rather study the moments of their approximations rather than the approximations of their moments. With the aid of Taylor approximation, we will take advantage of the first few terms of $\hat{\beta}_1$. These few terms have finite moments because they are either quadratic or cubic form of Gaussian vectors. That is, we can write $\hat{\beta}_1$ as $\hat{\beta}_1 = \hat{\beta}_{\text{Approx}} + \mathcal{O}_{\mathsf{P}}(\sigma^4)$. Here $\hat{\beta}_{\text{Approx}}$ has finite moments while the reminder $\mathcal{O}_{\mathsf{P}}$ does not.

The issue of the infinite moments for the MLE was ignored by practitioners because of the excellent behavior of these estimators in real-life applications. Indeed, the infinite moment of the estimators is barely seen in practice [2, 6, 7, 10] except when the noise level is relatively large. Therefore, Chernov [10, pp. 17] experimentally investigated this issue and discovered that if a set of n observations are distributed around a line segment of length L, then the infinite first-moment occurs whenever $\frac{\sigma}{L}$ is greater than or equal to 0.24. This value is unrealistically high for computer vision and image processing; therefore, the infinite first moment's issue is rarely observed.

To investigate how close $\hat{\beta}_{\text{approx}}$ is to $\hat{\beta}_1$, one could imagine an artificial example, where the probability distribution function (CDF) of the MLE could be expressed as a mixture distribution of the two distributions F_X and F_Y with weight $1-p$ and p for some $p \in (0,1)$, respectively. Here, the random variable X has finite moments, and Y has an infinite first moment. That is, $F_{\hat{\beta}_1} = (1-p)F_X + pF_Y$, thus $\mathsf{E}(|\hat{\beta}_1|) = \infty$ even if $p = 10^{-6}$. Here $p = 10^{-6}$ means that if one million samples were generated and the MLE was computed for each sample, then, on average, only one sample would come from the 'bad' distribution (as the Cauchy distribution) Y, while all other samples would come from the 'good' distribution X. This justifies how a very accurate estimator, as the MLE, has infinite first moment.

Al-Sharadqah and Chernov [2] investigated the issue of having an *accurate estimator with infinite moments* in EIV models. They experimentally investigated the MLE for both linear and circular regressions using this criterion. That is, the probability distribution function of its approximation, say $F_{\text{Approx}}(x)$, is good enough, if it accounts for 'almost all' of $F_{\hat{\beta}_1}(x)$ that can be represented as

$$F_{\hat{\beta}_1}(x) = (1-p)F_{\text{Approx}}(x) + pF_{\text{R}}(x) \qquad -\infty < x < \infty, \tag{2.1}$$

where $F_{\text{R}}(x)$ is some other probability distribution function (the 'remainder') and p is sufficiently small positive real number. According to Eq. (2.1), the realizations of $\hat{\beta}_1$ are taken from the 'good' distribution F_{Approx} with probability $1-p$ and from the 'bad' distribution f_{R} with probability p.

Thorough intensive numerical experiments have been conducted and it was found that the values of p for both linear and quadratic approximations are indeed very small as long as σ/L lies below some typical values, such as 0.1. Therefore, under the small-noise model adopted here, the MLE and its approximations are 'virtually' equal.

This paper is tailored for image processing applications, where the number of observable points (pixels) is limited, and the noise is small. The typical value of the noise level σ does not exceed $0.05L$. Accordingly, we will study estimators whenever $\sigma \to 0$, which is known as the *small-sigma model*.

The small-sigma model has a great impact on many research topics in image processing, signal processing, computer vision, and many other research topics [10]. Its importance stems from the following reason. On an image, the number of observed points (pixels on a computer screen) n is usually strictly limited, but the noise level σ is small. The small-noise model was firstly introduced by Kadane in the early 1970s and used later by Anderson [7] and Kanatani [15] (see also [17] for a more persuasive discussion). Such models were also studied by Amemiya, Fuller and Wolter [4, 20] ,who made a more rigid assumption that $n \sim \sigma^{-a}$ for some $0 < a < 2$.

This paper focuses on comparing the two estimators according to their order of magnitudes. We distinguish between terms of order σ^2 and σ^2/n. In typical computer vision and imaging processing applications, the number of points typically lies between 10–20 ($\sim \frac{1}{\sigma}$) up to few hundreds ($\sim \frac{1}{\sigma^2}$). Table 1 classifies terms according to their dependence on n. For example, terms with order of magnitude σ^2/n are comparable with terms of order σ^3 or even σ^4 (for relatively large n). Therefore, we will call the second-order

Table 1. The order of magnitudes of the four terms in the MSE

	σ^2/n	σ^4	σ^4/n	σ^6
Small samples ($n \sim 1/\sigma$)	σ^3	σ^4	σ^5	σ^6
Large samples ($n \sim 1/\sigma^2$)	σ^4	σ^4	σ^6	σ^6

bias of order σ^2 and σ^2/n by the *'essential second-order bias'* and the *'non-essential second-order bias'*, respectively. Indeed, the non-essential bias vanishes for large n while the essential bias persists.

In the analog of consistency of an estimator, we call an estimator *geometrically consistent* if it returns the true values of the parameters whenever all the points are observed without error (i.e., the data set is noiseless). Informally, $\lim_{\sigma\to 0}\hat{\theta}(\mathbf{m}_1,\ldots,\mathbf{m}_n) = \tilde{\theta}$, where $\tilde{\theta}$ is the true value of the parameter vector. We should mention here that geometric consistency requirement is considered as the minimal requirement for any estimator in geometric estimation problems.

This paper is a continuation of our work in [1], where the error analysis has been developed to study the statistical properties for any geometrically consistent estimator minimizing $\mathcal{F}$. Firstly, for an estimator, say $\hat{\beta}$, we have used its Taylor expansion around the true value $\tilde{\beta}$, i.e.,

$$\hat{\beta} = \tilde{\beta} + \Delta_1\hat{\beta} + \Delta_2\hat{\beta} + \Delta_3\hat{\beta} + \Delta_4\hat{\beta} + \mathcal{O}_{\mathsf{P}}\big(\sigma^5\big), \tag{2.2}$$

where $\tilde{\beta}$ is the true value of β and

$$\Delta_1\hat{\beta} = \sum \beta_{x_i}\delta_i + \sum \beta_{y_i}\varepsilon_i,$$

$$\Delta_2\hat{\beta} = \frac{1}{2}\left[\sum_{i,j}\beta_{x_ix_j}\delta_i\delta_j + \sum_{i,j}\beta_{x_iy_j}\delta_i\varepsilon_j + \sum_{i,j}\beta_{y_iy_j}\varepsilon_i\varepsilon_j\right]$$

are the first- and the second-order errors, respectively. Also, the formal expressions of the higher order error terms, i.e., $\Delta_3\hat{\beta}$ and $\Delta_4\hat{\beta}$, will be presented later. The symbol β_{x_i} represents the first partial derivative of the estimator $\hat{\beta}$ with respect to x_i, i.e., $\hat{\beta}_{x_i} = \frac{\partial\hat{\beta}}{\partial x_i}$, evaluated at $\tilde{\beta}$ and $(\tilde{x}_k, \tilde{y}_k)$, for all $k = 1,\ldots,n$. Similarly, $\beta_{x_iy_j}$ is the second partial derivatives of $\hat{\beta}$ with respect to x_i and y_j, i.e., $\hat{\beta}_{x_iy_j} = \frac{\partial^2\hat{\beta}}{\partial x_i\partial y_j}$, evaluated at $\tilde{\beta}$ and $(\tilde{x}_k, \tilde{y}_k)$, for all $k = 1,\ldots,n$. Accordingly, the following results have been established in [1].

Theorem 2.1. *Let* $\kappa(\beta) = (\beta^2 + \lambda)g(\beta)$ *and* $S = \frac{\|\tilde{\mathbf{x}}^*\|^2}{n}$, *then*

$$\mathsf{E}\big(\hat{\beta}\big) = \tilde{\beta} + \mathsf{E}\big(\Delta_2\hat{\beta}\big) + \mathsf{E}\big(\Delta_4\hat{\beta}\big) + \mathcal{O}\big(\sigma^6\big),$$

where

$$\mathsf{E}\big(\Delta_2\hat{\beta}\big) = \frac{-\tilde{\kappa}'\sigma^2}{2\tilde{g}S} + \frac{\big(\tilde{\kappa}' + \tilde{\beta}\tilde{g}\big)\sigma^2}{\tilde{g}nS} + \mathcal{O}\big(\sigma^4\big), \tag{2.3}$$

$$\mathsf{E}\big(\Delta_4\hat{\beta}\big) = \frac{\sigma^4\tilde{\kappa}'}{4\tilde{g}^2S^2}\left[2\tilde{\kappa}'' - \frac{3\tilde{g}'\tilde{\kappa}'}{\tilde{g}}\right] + \mathcal{O}\big(\sigma^4/n\big), \tag{2.4}$$

where $\tilde{g} = g\big(\tilde{\beta}\big)$, *and* $\tilde{\kappa}'$ *and* $\tilde{\kappa}''$ *be the first and the second derivatives of* κ *evaluated at the true values of the set of observations and the true parameter* $\tilde{\beta}$.

In the same analog, we called here every term of order of magnitude σ^4 by '*the fourth-order essential bias*', while we called all other terms of order of magnitude σ^4/n^a by the fourth-order nonessential bias for any $a > 0$.

Next, we turn our attention to the leading term of the bias. If we split the $\mathcal{O}(\sigma^2)$ terms into the *essential second-order bias* of order σ^2 and non-essential terms of order $\mathcal{O}(\sigma^2/n)$, we obtain

$$\text{bias}_{\text{ess}}(\hat{\beta}) = \frac{-\tilde{\kappa}'\sigma^2}{2\tilde{g}S}. \tag{2.5}$$

One might be interested in eliminating the *essential* second-order bias. This problem can be accomplished by solving the ordinary differential equation (ODE) $\kappa'(\tilde{\beta}) = 0$, i. e.,

$$(\tilde{\beta}^2 + \lambda)\tilde{g}' + 2\tilde{\beta}\tilde{g} = 0, \tag{2.6}$$

where $n \geq 3$. Solving the ODE given in (2.6) yields $g = g_1$. Accordingly, the minimum value of the corresponding objective function can be achieved at the MLE, $\hat{\beta}_1$.

Furthermore, one can obtain a more accurate estimator whose its entire second-order bias equals zero (i. e., its bias terms of magnitudes σ^2 and σ^2/n are both zero), then we need to find the weight that solves the ordinary differential equation (ODE)

$$(n-2)(\tilde{\beta}^2 + \lambda)g' + 2(n-3)\tilde{\beta}g = 0, \tag{2.7}$$

which leads to

$$g(\tilde{\beta}) = (\tilde{\beta}^2 + \lambda)^{-\frac{n-3}{n-2}} =: g_2(\tilde{\beta})$$

as a solution of the ODE (2.7). This justifies the rationale of choosing g_2. Based on that, $\hat{\beta}_2$ is an estimator of β and it minimizes the objective function $\mathcal{F}_{g_2}$ (i. e., $\mathcal{F}$ when $g = g_2$). This gives us an estimator with a zero second-order bias. It is the only estimator that eliminates the second-order bias. To compute this estimator, we solve $\frac{\partial \mathcal{F}_{g_2}}{\partial \beta} = 0$, which is reduced to solving the cubic equation given in (1.11). It is worth mentioning here that the MLE given in (1.8) is the solution of the quadratic equation (1.7). Therefore, we can consider (1.11) as a 'correction' of (1.7), and as such we can solve (1.11) numerically by using the solution (1.8) of (1.7) as an initial guess. Alternatively, we might just solve the cubic equation (1.11) by exact formulas, and select the root that minimizes the objective function. We summarize these results in the following theorem.

Theorem 2.2. *Up to an irrelevant scalar factor, the fit* (1.10) *has a zero-essential bias if and only if* $g = g_1(\beta) = \frac{1}{\beta^2+\lambda}$*. Moreover, for* $n \geq 4$*, the fit given in* (1.10) *has a zero second-order bias if and only if* $g = g_2$ *(up to an irrelevant scalar factor). Furthermore, without loss of generality, if we set* $S = 1$*, then*

- *For* $g(\beta) = g_1(\beta)$*;* $\hat{\beta} = \hat{\beta}_1$ *and*

$$\mathsf{E}(\Delta_2\hat{\beta}_1) = \frac{\sigma^2}{n}\tilde{\beta} \quad \textit{and} \quad \mathsf{E}(\Delta_4\hat{\beta}_1) = 0 + \mathcal{O}(\sigma^4/n). \tag{2.8}$$

- *For* $g(\beta) = g_2(\beta)$ *and* $n \geq 3$*;* $\hat{\beta} = \hat{\beta}_2$ *and* $nS = \|\tilde{\mathbf{x}}^*\|^2 = s_{\tilde{x}\tilde{x}}$*.*

$$\mathsf{E}(\Delta_2\hat{\beta}_2) = 0 \quad \textit{and} \quad \mathsf{E}(\Delta_4\hat{\beta}_2) = \frac{2\big(n-2+(2n-5)\tilde{\beta}^2\big)\tilde{\beta}}{(n-2)^3(1+\tilde{\beta}^2)}\sigma^4 + \mathcal{O}(\sigma^4/n^2). \tag{2.9}$$

Note here that $\hat{\beta}_2$ has zero second-order bias, while the MLE, $\hat{\beta}_1$, has a non-zero essential second-order bias. This demonstrates why $\hat{\beta}_2$ outperforms $\hat{\beta}_1$ for intermediate values of n, while both the estimators are comparably equal for large n. On the other hand, we have derived a general formula for the MSE of all estimators solving Eq. (1.10). The formula depends on the weight function g and its formal expression is

$$\text{MSE}(\hat{\beta}) = \frac{(\tilde{\beta}^2+\lambda)\sigma^2}{nS} + \frac{\sigma^4}{nS^2}\left(\lambda + \frac{2\tilde{\beta}^2 - \lambda}{n}\right) + \frac{\sigma^4}{4\tilde{g}^2 nS^2} \times \Bigg[(\tilde{\tau}' - 2\tilde{\beta}\tilde{g})\tilde{\kappa}' +$$

$$+ 4\left(1 - \frac{2}{n}\right)\Big(-(\tilde{\beta}^2+\lambda)\tilde{g}\tilde{\kappa}'' - \Big(2\tilde{\beta}\tilde{g} - (\tilde{\beta}^2+\lambda)\tilde{g}'\Big)\tilde{\kappa}'\Big) + \frac{8\tilde{\beta}\tilde{g}\tilde{\kappa}' + 4(7\tilde{\beta}^2 + 2\lambda)\tilde{g}^2}{n}\Bigg],$$

up to order σ^6.

As a standard statistical measure, the efficiency of any unbiased estimator can be determined by the *Cramér–Rao lower bound* (CRB). Kanatani [16] in 1998 derived a general CRB for arbitrary curves for *any unbiased estimators*. In geometric fitting problem, however, all estimators are biased. This makes the natural bound, CRB, is not helpful.

In the early 2000's, Chernov and Lesort [12] realized that Kanatani's formula does not work for any practical estimator in curve fitting problem. To overcome of this situation, Chernov and Lesort [12] employed first-order analysis for any geometrically consistent estimators. They showed that Kanatani's formula work for all geometrically consistent estimators, up to the leading order. Thus, Chernov and Lesort called it the *Kanatani–Cramér–Rao lower bound* (KCR). From that time, the KCR has been used as a measure for the efficiency for any meaningful estimator.

In the course of linear regression, the KCR lower bound means that the first leading term of the 'approximate' covariance matrix has a natural bound given by

$$\mathbf{V} \geq \sigma^2 \mathbf{V}_{\min}, \quad \mathbf{V}_{\min} = \frac{\lambda + \tilde{\beta}^2}{s_{xx}} \begin{bmatrix} \overline{\tilde{x}\tilde{x}} & -\bar{\tilde{x}} \\ -\bar{\tilde{x}} & 1 \end{bmatrix}, \tag{2.10}$$

and hence, $\mathbf{V}_{\min}(\hat{\beta}) = \frac{\lambda + \tilde{\beta}^2}{s_{xx}} = \frac{\lambda + \tilde{\beta}^2}{nS}$.

This general formula for the MSE produces the MSE of the MLE $\hat{\beta}_1$ and $\hat{\beta}_2$. Their MSE can be simply computed in terms of $\|\tilde{\mathbf{x}}^*\|^2 = nS$. For the MLE, since $\tilde{\kappa}' = \tilde{\kappa}'' = 0$, one obtains

$$\mathrm{MSE}(\hat{\beta}_1) = \frac{(\tilde{\beta}^2 + \lambda)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \frac{(n\lambda + 9\tilde{\beta}^2 + \lambda)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4} \overset{\lambda=1}{=} \frac{(\tilde{\beta}^2 + 1)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \frac{(n + 9\tilde{\beta}^2 + 1)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}. \tag{2.11}$$

While the MSE of $\hat{\beta}_2$ is

$$\begin{aligned} \mathrm{MSE}(\hat{\beta}_2) &= \frac{(\tilde{\beta}^2 + \lambda)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \frac{\sigma^4}{nS^2}\left(\lambda + \frac{2\tilde{\beta}^2 - \lambda}{n}\right) + \frac{2\sigma^4\tilde{\beta}^2}{n^2(n-2)S^2} = \\ &\overset{\lambda=1}{=} \frac{(\tilde{\beta}^2 + 1)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \left(1 + \frac{2\tilde{\beta}^2}{n-2}\right)\frac{(n-1)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}, \end{aligned} \tag{2.12}$$

where we used here $\tilde{\tau}' = 0$ and $\tilde{\kappa}' = \frac{2\tilde{\beta}\tilde{g}_2}{n-2}$.

It is worth mentioning here that the MSE of any estimator can be decomposed into

$$\mathrm{MSE}(\hat{\beta}) = \mathsf{E}\left[(\Delta_1\hat{\beta})^2\right] + \mathsf{E}\left[(\Delta_2\hat{\beta})^2\right] + 2\,\mathsf{E}(\Delta_1\hat{\beta}\Delta_3\hat{\beta}).$$

The most significant term in this expansion is $\mathsf{E}\left[(\Delta_1\hat{\beta})^2\right]$ and it is of order σ^2. This term does not depend on g so the leading terms of the MSE for all methods minimizing $\mathcal{F}$ are equal, and they all coincide with the KCR lower bound. Thus, all methods minimizing (1.10) are statistically efficient in the KCR sense.

The second important term in the MSE comes from the essential bias. Its contribution can be seen as part of $\mathsf{E}\left[(\Delta_2\hat{\beta})^2\right]$ (i. e., (essential bias + nonessential bias)^2 + $\mathrm{var}(\Delta_2\hat{\beta})$). These expressions were stated in Table 2. After a careful look at this table, one can easily see why the MLE outperforms the LS, but both estimators still fall behind $\hat{\beta}_2$.

3. Main Results

In [1], it was shown that the newly developed estimator $\hat{\beta}_2$ is the best estimator among all other estimators minimizing $\mathcal{F}$ in Eq. (1.10) including the least squares $\hat{\beta}_0$

Table 2. The components of the mean squared error for each of the three estimators: least-squares estimator $\hat{\beta}_0$, the MLE $\hat{\beta}_1$, and the new proposed estimator $\hat{\beta}_2$.

Method	$\mathsf{E}(\Delta_1\hat{\beta})^2$	$\mathsf{E}(\Delta_2\hat{\beta})^2 = \mathrm{Bias}(\Delta_2\hat{\beta}) + \mathrm{Var}(\Delta_2\hat{\beta})$	$2\,\mathsf{E}(\Delta_1\hat{\beta}\Delta_3\hat{\beta})$
$\hat{\beta}_0$	$\frac{(\tilde{\beta}^2+\lambda)\sigma^2}{nS}$	$\frac{\sigma^4}{nS^2}\left[\lambda + \frac{2\tilde{\beta}^2-\lambda}{n}\right] + \frac{\sigma^4\tilde{\beta}^2}{nS^2}\left[n-4+\frac{5}{n}\right]$	$-\frac{2\sigma^4}{nS^2}\left(1-\frac{3}{n}\right)\left(3\tilde{\beta}^2+\lambda\right)$
$\hat{\beta}_1$	$\frac{(\tilde{\beta}^2+\lambda)\sigma^2}{nS}$	$\frac{\sigma^4}{nS^2}\left[\lambda + \frac{2\tilde{\beta}^2-\lambda}{n}\right] + \frac{\sigma^4\tilde{\beta}^2}{n^2S^2}$	$\frac{2\sigma^4(3\tilde{\beta}^2+\lambda)}{n^2S^2}$
$\hat{\beta}_2$ ($\tilde{\tau}'=0$, $\tilde{\kappa}'=\frac{2\beta g_2}{n-2}$)	$\frac{(\tilde{\beta}^2+\lambda)\sigma^2}{nS}$	$\frac{\sigma^4}{nS^2}\left[\lambda + \frac{2\tilde{\beta}^2-\lambda}{n}\right] + \frac{2\sigma^4\tilde{\beta}^2}{n^2(n-2)S^2}$	0

and the MLE $\hat{\beta}_1$. Moreover, the numerical experiments of [1] showed that the AMLE $\check{\beta}_1$ outperforms the MLE, but the AMLE still falls behind $\hat{\beta}_2$.

Although both estimators eliminate the second-order bias, they behave quite differently in numerical experiments and their accuracy are quite different. Therefore, we devoted this paper to investigate why these estimators are quite different. In this paper, we will derive the bias and the MSE of their approximations, then we will discuss our findings.

Even though general formulas for the bias and the MSE have been derived in [1], those formulas work only for an estimator minimizing such an objective function (as we have seen for the geometric fit $\hat{\beta}_1$ and $\hat{\beta}_2$ when substituting $g = g_1$ and $g = g_2$, respectively, in $\mathcal{F}$). However, these general formulas cannot be applied to the AMLE because it does not minimize any objective function. Therefore, we need to derive them directly. To keep our calculations simple, we will only consider $\lambda = 1$.

To understand how this estimator works, we need to study the MLE first. Most of the upcoming expressions in this section can be written in terms of Kronecker product, and as such, we use some of its handy properties. These tools are presented below in Definition 3.1, Proposition 3.1, and Theorem 3.1.

Definition 3.1. Let $\mathbf{A}$ be an $m \times n$ matrix and let $\mathbf{B}$ be a $p \times q$ matrix. Then the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is that $(mp) \times (nq)$ matrix defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}.$$

Furthermore, $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$.

Proposition 3.1. *Let $\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1$ be square matrices of size p and let $\mathbf{A}_2$ and $\mathbf{B}_2$ and $\mathbf{C}_2$ be square matrices of size q. Then*

$$\mathrm{tr}(\mathbf{A}_1 \otimes \mathbf{A}_2(\mathbf{B}_1 \otimes \mathbf{B}_2 + \mathbf{C}_1 \otimes \mathbf{C}_2)) = \mathrm{tr}(\mathbf{A}_1\mathbf{B}_1)\mathrm{tr}(\mathbf{A}_2\mathbf{B}_2) + \mathrm{tr}(\mathbf{A}_1\mathbf{C}_1)\mathrm{tr}(\mathbf{A}_2\mathbf{C}_2).$$

Theorem 3.1 [19]**.** *Let $\boldsymbol{\zeta}$ be $n-$dimensional random vector with mean μ and covariance matrix $\boldsymbol{\Sigma}$ and let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrices of size n then*

$$\mathsf{E}\Big(\boldsymbol{\zeta}^\top \mathbf{A}\boldsymbol{\zeta}\Big) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \mu^\top \mathbf{A}\mu. \tag{3.1}$$

Moreover, if $\boldsymbol{\zeta} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive definite matrix, then

$$\mathsf{E}\Big(\boldsymbol{\zeta}^\top \mathbf{A}\boldsymbol{\zeta} \cdot \boldsymbol{\zeta}^\top \mathbf{B}\boldsymbol{\zeta}\Big) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})\,\mathrm{tr}(\mathbf{B}\boldsymbol{\Sigma}) + 2\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}).$$

3.1. MLE and its approximations. Here we will derive the first four order error terms of the MLE, namely $\Delta_i\hat{\beta}_1$, for each $i = 1, \dots, 4$. This is a crucial step for studying the AMLE. Moreover, one can use the results obtained in this section to validate Theorem 2.1.

Linear approximation and the KCR lower bound. We start our analysis with the linear approximation, i. e. $\hat{\beta}_{\mathrm{L}} = \tilde{\beta} + \Delta_1\hat{\beta}_1$. Here, the first-order error term $\Delta_1\hat{\beta}_1$ is a linear combination of $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ that represents the vectors of all noisy errors corrupting the first and the second coordinates of the vectors $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$. Using the first-order Taylor expansion of Eq. (1.4) about the true value $\tilde{\beta}$ and keeping only terms of order σ^2 yield

$$\begin{aligned}\mathcal{F}_1(\hat{\beta}) &= \frac{1}{1+\tilde{\beta}^2} \sum\Big(\tilde{y}_i^* + \varepsilon_i^* - \big(\tilde{\beta} + \Delta_1\hat{\beta}_1\big)(\tilde{x}_i^* + \delta_i^*)\Big)^2 + \mathcal{O}_{\mathsf{P}}(\sigma^3) = \\ &= \frac{1}{1+\tilde{\beta}^2} \sum\big(\varepsilon_i^* - \tilde{\beta}\delta_i^* - \tilde{x}_i^*\,\Delta_1\hat{\beta}_1\big)^2 + \mathcal{O}_{\mathsf{P}}(\sigma^3),\end{aligned} \tag{3.2}$$

where both $\delta_i^* = \delta_i - \bar{\delta}$ and $\varepsilon_i^* = \varepsilon_i - \bar{\varepsilon}$ denote the 'centered' errors. Accordingly, $\mathcal{F}_1$ attains its minimum at

$$\Delta_1\hat{\beta}_1 = \frac{(\tilde{\mathbf{x}}^*)^\top\big(\boldsymbol{\varepsilon}^* - \tilde{\beta}\boldsymbol{\delta}^*\big)}{\|\tilde{\mathbf{x}}^*\|^2}, \tag{3.3}$$

where $\boldsymbol{\delta}^*$ and $\boldsymbol{\varepsilon}^*$ denote the vectors of δ_i^*'s and ε_i^*'s, respectively. Let $\mathbf{h}^*$ denote the combined vector of δ_i^*'s and ε_i^*'s; i. e., $\mathbf{h}^* = \big(\boldsymbol{\delta}^{*\top}, \boldsymbol{\varepsilon}^{*\top}\big)^\top$. The components of $\mathbf{h}^*$ are not independent random variables, but $\mathbf{h}$ and $\mathbf{h}^*$ are related by $\mathbf{h}^* = \mathbf{N}\mathbf{h}$, where $\mathbf{N}$ is a $(2n) \times (2n)$ matrix defined by

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{N}_n \end{bmatrix}, \qquad \mathbf{N}_n = \mathbf{I}_n - \tfrac{1}{n}\,\mathbf{1}_n. \tag{3.4}$$

(Here $\mathbf{0}_n$ and $\mathbf{1}_n$ denote $n \times n$ matrices consisting zeroes and ones, respectively.) Thus, we can express $\Delta_1\hat{\beta}_1$ as a linear function of the random vector $\mathbf{h}$ whose components are independent.

$$\Delta_1\hat{\beta}_1 = \frac{\big(-\tilde{\beta}(\tilde{\mathbf{x}}^*)^\top, (\tilde{\mathbf{x}}^*)^\top\big)^\top\mathbf{N}_n\mathbf{h}}{\|\tilde{\mathbf{x}}^*\|^2} = \frac{\big(-\tilde{\beta}(\tilde{\mathbf{x}}^*)^\top, (\tilde{\mathbf{x}}^*)^\top\big)^\top\mathbf{h}}{\|\tilde{\mathbf{x}}^*\|^2} = \mathbf{G}^\top\mathbf{h}, \tag{3.5}$$

where we used the relation $(\tilde{\mathbf{x}}^*)^\top\mathbf{N}_n = (\tilde{\mathbf{x}}^*)^\top$ and the i^{th} component of $\mathbf{G}$ is

$$G_i = \begin{cases} -\tilde{\beta}\tilde{x}_i^*/\|\tilde{\mathbf{x}}^*\|^2 & \text{for } 1 \le i \le n, \\ \tilde{x}_{i-n}^*/\|\tilde{\mathbf{x}}^*\|^2 & \text{for } n+1 \le i \le 2n. \end{cases} \tag{3.6}$$

Therefore, the linear approximation is

$$\hat{\beta}_{\mathrm{L}} = \tilde{\beta} + \Delta_1\hat{\beta}_1. \tag{3.7}$$

Example 3.1 **(variance and bias of linear approximation)**. From (3.5), we can find the variance of $\hat{\beta}_{\mathrm{L}}$. Since $\mathsf{E}(\Delta_1\hat{\beta}_1) = 0$, the linear approximation $\hat{\beta}_{\mathrm{L}}$ is an unbiased estimator of $\tilde{\beta}$ (i. e. $\mathsf{E}(\hat{\beta}_{\mathrm{L}}) = \tilde{\beta}$). Thus,

$$\operatorname{Var}\big(\hat{\beta}_{\mathrm{L}}\big) = \mathsf{E}\big(\mathbf{h}^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}\big) = \sigma^2\operatorname{tr}\big(\mathbf{G}\mathbf{G}^\top\big) = \frac{\big(1+\tilde{\beta}^2\big)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2}. \tag{3.8}$$

This follows from writing $\mathbf{G}\mathbf{G}^\top$ as

$$\mathbf{G}\mathbf{G}^\top = \mathbf{a}_1 \otimes \mathbf{B}_n, \text{ where } \mathbf{a}_1 = \frac{1}{\|\tilde{\mathbf{x}}^*\|^4}\big(-\tilde{\beta}, 1\big)^\top\big(-\tilde{\beta}, 1\big), \quad \mathbf{B}_n = \tilde{\mathbf{x}}^*(\tilde{\mathbf{x}}^*)^\top, \tag{3.9}$$

and using Definition 3.1 and the fact $\operatorname{tr}(\mathbf{B}_n) = \|\tilde{\mathbf{x}}^*\|^2 = s_{\tilde{x}\tilde{x}}$.

Example 3.1 shows that $\mathrm{Var}(\hat{\beta}_{\mathrm{L}})$ is of order σ^2/n, which also attains the KCR. It also indicates that the linear approximation is unbiased estimator of $\hat{\beta}_1$! Only in 1976, explicit formulas for the density functions of the estimators $\hat{\alpha}_1$ and $\hat{\beta}_1$ were derived; see [5, 7]. It turns out that those densities are not normal and do not belong to any standard family of probability densities. Those formulas are overly complicated and involve double-infinite series. It was promptly noted [5] that they were not very useful for practical purposes. Moreover, the probability density function of $\hat{\beta}_1$ is skewed except when $\tilde{\beta} = 0$. Therefore, the linear approximation $\hat{\beta}_{\mathrm{L}}$ (whose pdf is normal!) is not a good approximation for the MLE. Accordingly, we will go further in our analysis by considering the quadratic and the cubic approximations.

Quadratic and cubic approximations. The quadratic and the cubic approximations of the MLE are given by the following general formulas

$$\hat{\beta}_{\mathrm{Q}} = \tilde{\beta} + \Delta_1\hat{\beta}_1 + \Delta_2\hat{\beta}_1, \ \text{ and } \hat{\beta}_{\mathrm{C}} = \hat{\beta}_{\mathrm{Q}} + \Delta_3\hat{\beta}_1, \tag{3.10}$$

where $\Delta_3\hat{\beta}_1$ involves all random terms of order $\mathcal{O}_{\mathrm{P}}(\sigma^3)$.

Before presenting the formal expressions of these approximations, we introduce the following terms:

$$\alpha_1 = \frac{-r_1 + 2\tilde{\beta}}{2\|\tilde{\mathbf{x}}^*\|^4}, \qquad \gamma_1 = -\frac{r_1\tilde{\beta}^2 + 2\tilde{\beta}}{2\|\tilde{\mathbf{x}}^*\|^2}, \tag{3.11}$$

$$\alpha_2 = -\frac{r_1\tilde{\beta} + 2}{2\|\tilde{\mathbf{x}}^*\|^4}, \qquad \gamma_2 = \frac{r_1\tilde{\beta} + 1}{2\|\tilde{\mathbf{x}}^*\|^2}, \tag{3.12}$$

$$\alpha_3 = \frac{r_1}{2\|\tilde{\mathbf{x}}^*\|^4}, \qquad \gamma_3 = \frac{-r_1}{2\|\tilde{\mathbf{x}}^*\|^2}, \tag{3.13}$$

where

$$r_1 = \frac{-2\tilde{\beta}}{1 + \tilde{\beta}^2}, \qquad r_2 = \frac{3\tilde{\beta}^2 - 1}{(1 + \tilde{\beta}^2)^2}. \tag{3.14}$$

Also, let $\mathbf{a}$ and $\mathbf{b}$ be 2-by-2 symmetric matrices defined as

$$\mathbf{a} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & \gamma_3 \end{bmatrix}. \tag{3.15}$$

Theorem 3.2. *Let $\hat{\beta}_1$ be the MLE of the slope β of the line $y = \alpha + \beta x$ and $\mathbf{B}_n = \tilde{\mathbf{x}}^*(\tilde{\mathbf{x}}^*)^\top$. Then, the second order term $\Delta_2\hat{\beta}_1$ is a quadratic form of the combined error $\mathbf{h}$ and it takes the formal expression*

$$\Delta_2\hat{\beta}_1 = \mathbf{h}^\top\mathbf{N}^\top\mathbf{Q}\mathbf{N}\mathbf{h}\,, \tag{3.16}$$

where $\mathbf{N} = \mathbf{I}_2 \otimes \mathbf{N}_n$ (cf. Eq. (3.4)) and $\mathbf{Q}$ is a $(2n) \times (2n)$ matrix defined by

$$\mathbf{Q} = \mathbf{a} \otimes \mathbf{B}_n + \mathbf{b} \otimes \mathbf{I}_n\,. \tag{3.17}$$

Furthermore, the third-order term $\Delta_3\hat{\beta}_1$ is a tensor product of the centered-combined error $\mathbf{h}^*$ and it takes the form

$$\begin{aligned}\Delta_3\hat{\beta}_1 = -\frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\Big[&\mathbf{G}^\top\mathbf{h}^*(\mathbf{h}^*)^\top\mathbf{A}\mathbf{h}^* - r_2\|\tilde{\mathbf{x}}^*\|^2(\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*\mathbf{G}^\top\mathbf{h}^* + \\ &+ r_1\|\tilde{\mathbf{x}}^*\|^2\mathbf{G}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^* + 3r_1\mathbf{E}^\top\mathbf{h}^*(\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^* + 2\mathbf{E}^\top\mathbf{h}^*(\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^*\Big],\end{aligned} \tag{3.18}$$

where $\mathbf{E} = (\tilde{\mathbf{x}}^\top, \mathbf{0}^\top)^\top$, $\mathbf{A} = \hat{\boldsymbol{\gamma}}_1 \otimes \mathbf{I}_n$, and

$$\hat{\boldsymbol{\gamma}}_1 = \begin{bmatrix} r_2\tilde{\beta}^2 + 2r_1\tilde{\beta} + 1 & -(r_2\tilde{\beta} + r_1) \\ -(r_2\tilde{\beta} + r_1) & r_2 \end{bmatrix}. \tag{3.19}$$

Example 3.2 **(Variance and bias of the quadratic approximation)**. As a consequence of Theorem 3.2, we find the components of the MSE of $\hat{\beta}_Q$. Since $\mathbf{N}$ is an idempotent matrix,

$$\operatorname{tr}(\mathbf{N}^\top \mathbf{Q}\mathbf{N}) = \operatorname{tr}(\mathbf{Q}\mathbf{N}), \text{ where } \mathbf{Q}\mathbf{N} = \mathbf{a} \otimes (\mathbf{B}_n \mathbf{N}_n) + \mathbf{b} \otimes \mathbf{N}_n, \tag{3.20}$$

then, by Definition 3.1, we have

$$\operatorname{tr}(\mathbf{Q}\mathbf{N}) = \operatorname{tr}(\mathbf{a})\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) + \operatorname{tr}(\mathbf{b})\operatorname{tr}(\mathbf{N}_n) = \frac{\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}. \tag{3.21}$$

Here we used $\operatorname{tr}(\mathbf{N}_n) = n-1$, $\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) = \|\tilde{\mathbf{x}}^*\|^2$, $\operatorname{tr}(\mathbf{b}) = 0$, and $\operatorname{tr}(\mathbf{a}) = \frac{\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^4}$. Consequently, Theorems 3.1 and 3.2 show that

$$\operatorname{bias}(\hat{\beta}_Q) = \sigma^2 \operatorname{tr}(\mathbf{N}^\top \mathbf{Q}\mathbf{N}) = \frac{\tilde{\beta}\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2}. \tag{3.22}$$

To compute the variance of $\hat{\beta}_Q$, we only need to compute the variance of $\Delta_2\hat{\beta}_1$. Being a quadratic form of $\mathbf{h}$,

$$\operatorname{Var}(\Delta_2\hat{\beta}) = 2\sigma^4 \operatorname{tr}\big((\mathbf{N}^\top\mathbf{Q}\mathbf{N})^2\big) = 2\sigma^4\operatorname{tr}\big((\mathbf{Q}\mathbf{N})^2\big),$$

where

$$\operatorname{tr}\big[(\mathbf{Q}\mathbf{N})^2\big] = \operatorname{tr}(\mathbf{a}^2)\operatorname{tr}(\mathbf{B}_n^2\mathbf{N}_n) + 2\operatorname{tr}(\mathbf{a}\mathbf{b})\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) + \operatorname{tr}(\mathbf{b}^2)\operatorname{tr}(\mathbf{N}_n) = \frac{n+2\tilde{\beta}^2-1}{2\|\tilde{\mathbf{x}}^*\|^4}.$$

Here we used

$$\operatorname{tr}(\mathbf{B}_n^2\mathbf{N}_n) = \|\tilde{\mathbf{x}}^*\|^4, \quad \operatorname{tr}(\mathbf{a}^2) = \frac{4(\tilde{\beta}^2+2)}{4\|\tilde{\mathbf{x}}^*\|^8}, \quad \operatorname{tr}(\mathbf{a}\mathbf{b}) = \frac{-4}{4\|\tilde{\mathbf{x}}^*\|^6},$$

and

$$\operatorname{tr}(\mathbf{N}_n) = n-1, \qquad \operatorname{tr}(\mathbf{b}^2) = \frac{2}{4\|\tilde{\mathbf{x}}^*\|^4}.$$

By Theorem 3.1, we have

$$\operatorname{Var}(\Delta_2\hat{\beta}_1) = \frac{n+2\tilde{\beta}^2-1}{\|\tilde{\mathbf{x}}^*\|^4}\sigma^4, \qquad \operatorname{Var}(\hat{\beta}_Q) = \frac{(1+\tilde{\beta}^2)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \frac{n+2\tilde{\beta}^2-1}{\|\tilde{\mathbf{x}}^*\|^4}\sigma^4. \tag{3.23}$$

Example 3.2 shows that the leading term of $\operatorname{Var}(\hat{\beta}_Q)$ is of order σ^2/n and attains the KCR. Furthermore, the bias is a linear function of $\tilde{\beta}$. In particular, it overestimates the true value of the parameter if $\tilde{\beta} > 0$ and underestimates $\tilde{\beta}$ if $\tilde{\beta} < 0$. The bias of $\hat{\beta}_Q$ has typical values of order σ^2/n, so its contribution to the MSE is negligible (i. e. σ^4/n^2) when n is large enough. This shows the MLE $\hat{\beta}_1$ has only a *nonessential second-order bias of order* σ^2/n.

Before going further, we need the following lemma.

Lemma 3.1. *We have the following properties:*

$$\operatorname{tr}(\mathbf{G}\mathbf{G}^\top\mathbf{N}) = \frac{1+\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^2}, \qquad \operatorname{tr}(\boldsymbol{\Gamma}_1\mathbf{N}) = \frac{1+3\tilde{\beta}^2}{1+\tilde{\beta}^2}, \qquad \operatorname{tr}(\boldsymbol{\Gamma}_2\mathbf{N}) = -4\tilde{\beta}.$$

Moreover,

$$\operatorname{tr}(\boldsymbol{\Gamma}_1\mathbf{G}\mathbf{G}^\top\mathbf{N}) = \frac{3\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^2}, \qquad \operatorname{tr}(\boldsymbol{\Gamma}_2\mathbf{N}\mathbf{Q}\mathbf{N}) = -\frac{4\tilde{\beta}^2+1}{\|\tilde{\mathbf{x}}^*\|^2}.$$

Proof of Lemma 3.1. Recall that $\mathbf{N} = \mathbf{I}_2 \otimes \mathbf{N}_n$ and $\mathbf{G}\mathbf{G}^\top = \mathbf{a}_1 \otimes \mathbf{B}_n$ (see Eq. (3.9)). Also we will use $(\tilde{\mathbf{x}}^*)^\top\mathbf{N}_n = (\tilde{\mathbf{x}}^*)^\top$ and $\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) = \|\tilde{\mathbf{x}}^*\|^2$. Thus,

$$\operatorname{tr}(\mathbf{G}\mathbf{G}^\top\mathbf{N}) = \operatorname{tr}(\mathbf{a}_1)\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) = \frac{1+\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^2}.$$

This proves the first assertion.

The second assertion follows from the definitions of $\mathbf{\Gamma}_1$ and $\mathbf{N}$. Since $\operatorname{tr}(\hat{\boldsymbol{\gamma}}_1) = 0$ and $\operatorname{tr}(\boldsymbol{\gamma}_1) = \frac{1+3\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^2(1+\tilde{\beta}^2)}$, we obtain

$$\operatorname{tr}(\mathbf{\Gamma}_1\mathbf{N}) = \operatorname{tr}(\boldsymbol{\gamma}_1)\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) + \operatorname{tr}(\hat{\boldsymbol{\gamma}}_1)\operatorname{tr}(\mathbf{N}_n) = \frac{1+3\tilde{\beta}^2}{1+\tilde{\beta}^2}.$$

In the same manner, if one uses $r_1(1+\tilde{\beta}^2) = -2\tilde{\beta}$, then the third assertion follows immediately.

Next, we write $\mathbf{G}\mathbf{G}^\top\mathbf{N}\mathbf{\Gamma}_1\mathbf{N}$ as

$$\mathbf{G}\mathbf{G}^\top\mathbf{N}\mathbf{\Gamma}_1\mathbf{N} = \Big(\mathbf{a}_1 \otimes (\mathbf{B}_n\mathbf{N}_n)\Big)\Big(\boldsymbol{\gamma}_1 \otimes (\mathbf{B}_n\mathbf{N}_n) + \hat{\boldsymbol{\gamma}}_1 \otimes (\mathbf{N}_n)\Big). \tag{3.24}$$

Since $\operatorname{tr}(\mathbf{a}_1\hat{\boldsymbol{\gamma}}_1) = \frac{-1}{\|\tilde{\mathbf{x}}^*\|^4}$ and $\operatorname{tr}(\mathbf{a}_1\boldsymbol{\gamma}_1) = \frac{3\tilde{\beta}^2+1}{\|\tilde{\mathbf{x}}^*\|^6}$, we have

$$\operatorname{tr}(\mathbf{\Gamma}_1\mathbf{G}\mathbf{G}^\top\mathbf{N}) = \operatorname{tr}(\mathbf{a}_1\boldsymbol{\gamma}_1)\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n\mathbf{B}_n) + \operatorname{tr}(\mathbf{a}_1\hat{\boldsymbol{\gamma}}_1)\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) = \frac{3\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^2}.$$

Finally, we prove the last assertion. The matrix $\mathbf{\Gamma}_2\mathbf{N}\mathbf{Q}\mathbf{N}$ can be expressed as

$$\mathbf{\Gamma}_2\mathbf{N}\mathbf{Q}\mathbf{N} = \boldsymbol{\gamma}_2 \otimes (\mathbf{B}_n\mathbf{N}_n)(\mathbf{a}\otimes\mathbf{B}_n + \mathbf{b}\otimes\mathbf{I}_n).$$

Since $\operatorname{tr}(\mathbf{a}\boldsymbol{\gamma}_2) = -\frac{4\tilde{\beta}^2+2}{\|\tilde{\mathbf{x}}^*\|^6}$ and $\operatorname{tr}(\mathbf{b}\boldsymbol{\gamma}_2) = \frac{1}{\|\tilde{\mathbf{x}}^*\|^4}$, we have

$$\operatorname{tr}(\mathbf{\Gamma}_2\mathbf{N}\mathbf{Q}) = \operatorname{tr}(\boldsymbol{\gamma}_2\mathbf{a})\operatorname{tr}(\mathbf{B}_n^2\mathbf{N}_n) + \operatorname{tr}(\boldsymbol{\gamma}_2\mathbf{b})\operatorname{tr}(\mathbf{B}_n\mathbf{N}_n) = -\frac{4\tilde{\beta}^2+1}{\|\tilde{\mathbf{x}}^*\|^2}.$$

This completes the proof of the lemma. □

The following theorem summarizes the final expression of the MSE of $\hat{\beta}_1$, up to order σ^4/n^2. Its proof is deferred to the appendix.

Theorem 3.3. *Let $\hat{\beta}_1$ be the MLE of β for the linear model $y = \alpha + \beta x$ and let $\hat{\beta}_Q$ be its quadratic approximation, then*

$$\operatorname{MSE}(\hat{\beta}_Q) = \frac{(1+\tilde{\beta}^2)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \frac{(n+9\tilde{\beta}^2+1)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}, \tag{3.25}$$

where $\operatorname{bias}^2(\hat{\beta}_Q) = \frac{\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^4}\sigma^4$.

3.2. Comparison between $\check{\beta}_1$ and $\hat{\beta}_2$. To theoretically compare $\check{\beta}_1$ and $\hat{\beta}_2$ in terms of their MSEs, we will first find the MSE of $\check{\beta}_1$. This step involves the expected values of the product of two quadratic forms of $\mathbf{h}^*$, $\boldsymbol{\delta}^*$, and $\boldsymbol{\gamma}^*$ and it leads to many useful identities. The identities are summarized in the following lemma while their derivations are deferred to the appendix.

Lemma 3.2. *Define $\check{\mathbf{a}}_1 = (-\tilde{\beta}, 1)^\top$, $\boldsymbol{\gamma}^* = \boldsymbol{\varepsilon}^* - \tilde{\beta}\boldsymbol{\delta}^*$, and*

$$\mathbf{r}_1 = \begin{bmatrix} -\tilde{\beta} & .5 \\ .5 & 0 \end{bmatrix}.$$

Also, define $\mathbf{\Sigma}_1 = (\mathbf{I}_n \otimes \check{\mathbf{a}}_1)\mathbf{V}_n(\check{\mathbf{a}}_1^\top \otimes \mathbf{I}_n) = (\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\otimes\mathbf{V}_n$ and $\mathbf{\Sigma}_2 = \mathbf{r}_1 \otimes \mathbf{V}_n$. Then, the following identities hold

$$\mathsf{E}\big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)^2\big] = n(n-2)(\tilde{\beta}^2+1)^2\sigma^4, \tag{3.26}$$

$$\mathsf{E}\big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*)\big) = (n-2)\|\tilde{\mathbf{x}}^*\|^{-2}\tilde{\beta}(\tilde{\beta}^2+1)\sigma^4, \tag{3.27}$$

$$\mathsf{E}[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*)] = (n-2)\|\tilde{\mathbf{x}}^*\|^2(\tilde{\beta}^2+1)^2\sigma^4, \tag{3.28}$$

$$\mathsf{E}\big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*)\big] = -(n-2)\|\tilde{\mathbf{x}}^*\|^2\tilde{\beta}(\tilde{\beta}^2+1)\sigma^4, \tag{3.29}$$

$$\mathsf{E}\big(\mathbf{h}^{*\top}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1{}^\top)\otimes\mathbf{B}_n)\mathbf{h}^*\cdot\mathbf{h}^{*\top}\mathbf{\Sigma}_1\mathbf{h}^*\big)=(n-2)\|\tilde{\mathbf{x}}^*\|^2(\tilde{\beta}^2+1)^2\sigma^4, \tag{3.30}$$

$$\mathsf{E}\big(\mathbf{h}^{*\top}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1{}^\top)\otimes\mathbf{B}_n)\mathbf{h}^*\cdot\mathbf{h}^{*\top}\mathbf{\Sigma}_2\mathbf{h}^*\big)=-(n-2)\|\tilde{\mathbf{x}}^*\|^2\tilde{\beta}(\tilde{\beta}^2+1)\sigma^4, \tag{3.31}$$

where $\mathbf{Q}=\mathbf{a}\otimes\mathbf{B}_n+\mathbf{b}\otimes\mathbf{I}_n$.

Now, we turn our attention to derive the MSE of $\check{\beta}_1$, up to order σ^6. For this purpose, we will derive the explicit formulas for the second- and third-order error terms of $\hat{\sigma}^2$. If $\hat{\sigma}^2$ is expanded about the true value $\tilde{\beta}$, then we have

$$\hat{\sigma}^2=\frac{1}{n-2}\sum_i\big(a_i^2+b_i^2+2a_ib_i+2a_ic_i\big)\big(f_0+f_1\Delta_1\hat{\beta}_1+f_1\Delta_2\hat{\beta}_1+f_2\Delta_1\hat{\beta}_1^2\big), \tag{3.32}$$

where

$$f_0=\frac{1}{1+\tilde{\beta}^2},\quad f_1=-\frac{2\tilde{\beta}}{(1+\tilde{\beta}^2)^2},\quad f_2=-\frac{1-3\tilde{\beta}^2}{(1+\tilde{\beta}^2)^3}.$$

Thus, $\hat{\sigma}^2$ can be expressed as

$$\hat{\sigma}^2=\Delta_2\hat{\sigma}^2+\Delta_3\hat{\sigma}^2+\Delta_4\hat{\sigma}^2+\mathcal{O}_{\mathsf{P}}\big(\sigma^5\big). \tag{3.33}$$

The expressions of these terms are summarized in the following lemma, while their derivations are deferred to the appendix.

Lemma 3.3.

$$\Delta_2\hat{\sigma}^2=\frac{f_0(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)}{n-2}. \tag{3.34}$$

$$\Delta_3\hat{\sigma}^2=\frac{\Delta_1\hat{\beta}_1}{n-2}\Big(f_1(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)-2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\Big). \tag{3.35}$$

$$\Delta_4\hat{\sigma}^2=\frac{1}{n-2}(f_1\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*-2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\Delta_2\hat{\beta}_1+(f_2\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*-2f_1\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\Delta_1\hat{\beta}_1^2. \tag{3.36}$$

Based on this lemma, we now write

$$\begin{aligned}\mathrm{MSE}(\check{\beta}_1)&=\mathsf{E}\left(\left[\left(1-\frac{\hat{\sigma}^2}{\|\mathbf{x}^*\|^2}\right)\hat{\beta}_1-\tilde{\beta}\right]^2\right)=\\&=\mathrm{MSE}(\hat{\beta}_1)-\mathsf{E}\left(\frac{2\Delta\hat{\beta}_1\hat{\sigma}^2\hat{\beta}_1}{\|\mathbf{x}^*\|^2}\right)+\mathsf{E}\left(\frac{\hat{\sigma}^4}{\|\mathbf{x}^*\|^4}\hat{\beta}_1^2\right)+\mathcal{O}\big(\sigma^6\big).\end{aligned} \tag{3.37}$$

We will start with

$$\mathsf{E}\left(\frac{\hat{\sigma}^4}{\|\mathbf{x}^*\|^4}\hat{\beta}_1^2\right)=\mathsf{E}\left[\frac{\hat{\sigma}^4\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^4}\right]=\frac{\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^4}\mathsf{E}\big[(\Delta_2\hat{\sigma}^2)^2\big]. \tag{3.38}$$

Substituting $\Delta_2\hat{\sigma}^2$ (see (3.34)) in (3.38) and using Lemma 3.2 give us

$$\mathsf{E}\left[\frac{\hat{\sigma}^4\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^4}\right]=\frac{f_0^2\tilde{\beta}^2\mathsf{E}\big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)^2\big]}{\|\tilde{\mathbf{x}}^*\|^4(n-2)^2}=\frac{\tilde{\beta}^2n}{\|\tilde{\mathbf{x}}^*\|^4(n-2)}+\mathcal{O}\big(\sigma^6\big). \tag{3.39}$$

Next we find $\mathsf{E}\Big(\frac{\hat{\sigma}^2\hat{\beta}_1\Delta\hat{\beta}_1}{\|\mathbf{x}^*\|^2}\Big)$, which can be decomposed into three terms.

$$\mathsf{E}\left(\frac{\hat{\sigma}^2\hat{\beta}_1\Delta\hat{\beta}_1}{\|\mathbf{x}^*\|^2}\right)=\mathrm{I}+\mathrm{II}+\mathrm{III}, \tag{3.40}$$

where

$$\mathrm{I}=\mathsf{E}\left(\frac{\tilde{\beta}\Delta_2\hat{\sigma}^2\Delta_2\hat{\beta}_1}{\|\tilde{\mathbf{x}}^*\|^2}\right),\quad\mathrm{II}=\mathsf{E}\left(\frac{\Delta_2\hat{\sigma}^2\Delta_1\hat{\beta}_1\hat{\beta}_1}{\|\mathbf{x}^*\|^2}\right),\quad\mathrm{III}=\mathsf{E}\left(\frac{\Delta_3\hat{\sigma}^2\Delta_1\hat{\beta}_1\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}\right).$$

The second-order error term of $\hat{\beta}_1$ was expressed in terms of the combined error vector $\mathbf{h}^*$ (cf. (3.16) and (3.17)). Therefore,

$$\mathrm{I} = \mathsf{E}\left(\frac{\tilde{\beta} f_0(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*)}{(n-2)\|\tilde{\mathbf{x}}^*\|^2}\right) = \frac{\tilde{\beta}^2\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}.$$

Next, we find

$$\mathrm{II} = \frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\mathsf{E}\left[\Delta_2\hat{\sigma}^2\Delta_1\hat{\beta}_1\left(\Delta_1\hat{\beta}_1 - \frac{2\tilde{\beta}(\tilde{\mathbf{x}}^*)^\top\boldsymbol{\delta}^*}{\|\tilde{\mathbf{x}}^*\|^2}\right)\right] =$$

$$= \frac{f_0}{(n-2)\|\tilde{\mathbf{x}}^*\|^6}\mathsf{E}\Big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\Big((\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*) - 2\tilde{\beta}(\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*)\Big)\Big]. \tag{3.41}$$

Substituting (3.28) and (3.29) into (3.41) leads to

$$\mathrm{II} = \frac{\sigma^4(3\tilde{\beta}^2+1)}{\|\tilde{\mathbf{x}}^*\|^4}. \tag{3.42}$$

Lastly, we will prove that $\mathrm{III} = 0$. That is, using (3.35) in Lemma 3.3, we can rewrite $\Delta_1\hat{\beta}_1\Delta_3\hat{\sigma}^2$ as

$$\Delta_1\hat{\beta}_1\Delta_3\hat{\sigma}^2 = \frac{\Delta_1\hat{\beta}_1^2}{n-2}\Big(f_1(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*) - 2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\Big), \tag{3.43}$$

which is a product of two quadratic forms of $\mathbf{h}^*$. Indeed, recall that $\boldsymbol{\gamma}^* = (\check{\mathbf{a}}_1^\top \otimes \mathbf{I}_n)\mathbf{h}^*$, then

$$\Delta_1\hat{\beta}_1^2 = \frac{\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*}{\|\tilde{\mathbf{x}}^*\|^4} = \frac{\mathbf{h}^{*\top}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top \otimes \mathbf{B}_n)\mathbf{h}^*}{\|\tilde{\mathbf{x}}^*\|^4}.$$

Also $\boldsymbol{\Sigma}_2 = \mathbf{r}_1 \otimes \mathbf{V}_n$, then $\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^* = \mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*$ and $\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^* = \mathbf{h}^{*\top}\boldsymbol{\Sigma}_2\mathbf{h}^*$. Thus,

$$\mathsf{E}(\Delta_1\hat{\beta}_1\Delta_3\hat{\sigma}^2) = \mathsf{E}\left[\frac{\mathbf{h}^{*\top}((\check{\mathbf{a}_1}\check{\mathbf{a}_1}^\top)\otimes\mathbf{B}_n)\mathbf{h}^*}{(n-2)\|\tilde{\mathbf{x}}^*\|^4}\Big(f_1(\mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*) - 2f_0(\mathbf{h}^{*\top}\boldsymbol{\Sigma}_2\mathbf{h}^*)\Big)\right], \tag{3.44}$$

where $f_1 = \frac{-2\tilde{\beta}}{(\tilde{\beta}^2+1)^2}$.

Now, Substituting Eq. (3.30) and Eq. (3.31) into Eq. (3.44) yields $\mathsf{E}(\Delta_1\hat{\beta}\Delta_3\hat{\sigma}^2) = 0$, and as such, $\mathrm{III} = 0$. Combining I, II, and III leads to

$$\mathsf{E}\left(\frac{\hat{\sigma}^2\hat{\beta}_1\Delta\hat{\beta}_1}{\|\mathbf{x}^*\|^2}\right) = \frac{(4\tilde{\beta}^2+1)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4} + \mathcal{O}\big(\sigma^6\big). \tag{3.45}$$

Lastly,

$$\mathrm{MSE}(\check{\beta}_1) = \frac{(1+\tilde{\beta}^2)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \left(1 + \frac{2\tilde{\beta}^2}{n-2}\right)\frac{(n-1)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}.$$

Elementary calculus can, now, helps us proving the following. For all values of $n \geq 3$ and $\sigma > 0$, $k(\tilde{\beta}) = \mathrm{MSE}(\hat{\beta}_1) - \mathrm{MSE}(\check{\beta}_1)$ is an increasing function in $|\tilde{\beta}|$. This shows that post-bias elimination reduces the variation of the new estimator.

More importantly, the MSEs of both of the estimators $\check{\beta}_1$ and $\hat{\beta}_2$ are equal, up to order σ^6. This means that both estimators differ in their third-order term of their MSEs. Since our analysis shows that indeed both methods eliminate $\mathcal{O}(\sigma^2)$ bias, we will track their higher order terms.

We find the bias of $\check{\beta}_1$, where

$$\check{\beta}_1 = \hat{\beta}_1 - \frac{\hat{\sigma}^2\hat{\beta}_1}{\|\mathbf{x}^*\|^2} := \hat{\beta}_1 - C$$

up to order σ^4/n^2. Then we will compare the biases of $\hat{\beta}_2$ and $\check{\beta}_1$. Smaller bias means a better estimator.

To find the bias of $\check{\beta}_1$, we only need to find $\mathsf{E}(C)$, where C is expressed as

$$C = \tilde{C} + \Delta_1 C + \Delta_2 C + \Delta_3 C + \Delta_4 C + \mathcal{O}_{\mathsf{P}}\big(\sigma^5\big).$$

Here $\tilde{C} = 0$ since $\hat{\sigma}^2 \sim \mathcal{O}_{\mathsf{P}}(\sigma^2)$. The expected values of $\Delta_1 C$ and $\Delta_3 C$ equal to zero, and as such

$$\mathsf{E}(\Delta_2 C) = \mathsf{E}\left(\frac{\hat{\sigma}^2\hat{\beta}}{\|\mathbf{x}^*\|^2}\right) = \frac{\tilde{\beta}\,\mathsf{E}(\Delta_2\hat{\sigma}^2)}{\|\tilde{\mathbf{x}}^*\|^2} = \frac{\sigma^2\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}.$$

The fourth-order error term of C is the complicated one. To track our sequel analysis we present the following lemma that summarizes some needed expressions while their derivations are deferred to the appendix.

Lemma 3.4. *Recall all definitions in Lemma 3.2 and define* $\boldsymbol{\Sigma}_3 = (\mathbf{s}_1\mathbf{s}_1^\top)\otimes\mathbf{I}_n$, *where* $\mathbf{s}_1 = (1,0)^\top$. *Then*

$$\mathsf{E}\big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\|\boldsymbol{\delta}^*\|^2\big) = \sigma^4(n-2)\Big((\tilde{\beta}^2+1)(n-1)+2\tilde{\beta}^2\Big). \tag{3.46}$$

$$\mathsf{E}\Big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*)\Big) = (n-2)\|\tilde{\mathbf{x}}^*\|^2(\tilde{\beta}^2+1)\sigma^4. \tag{3.47}$$

$$\mathsf{E}\Big((\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*)\Big) = (n-2)\|\tilde{\mathbf{x}}^*\|^2(\tilde{\beta}^2+5)\sigma^4. \tag{3.48}$$

$$\mathsf{E}\Big((\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*)\Big) = (n-2)\|\tilde{\mathbf{x}}^*\|^{-2}(1-\tilde{\beta}^2)\sigma^4. \tag{3.49}$$

$$\mathsf{E}\Big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*)\Big) = -\sigma^4(n-2)\|\tilde{\mathbf{x}}^*\|^2\tilde{\beta}(\tilde{\beta}^2+1). \tag{3.50}$$

Now, to simplify our calculations, we write

$$\mathsf{E}(\Delta_4 C) = \mathrm{I}' + \mathrm{II}' + \mathrm{III}',$$

where

$$\mathrm{I}' = \frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\mathsf{E}\left[\Delta_2\hat{\sigma}^2\left(\Delta_2\hat{\beta} - \frac{2\Delta_1\hat{\beta}\langle\tilde{\mathbf{x}}^*,\boldsymbol{\delta}^*\rangle}{\|\tilde{\mathbf{x}}^*\|^2} + \tilde{\beta}\left(-\frac{\|\boldsymbol{\delta}^*\|^2}{\|\tilde{\mathbf{x}}^*\|^2} + 4\frac{\langle\tilde{\mathbf{x}}^*,\boldsymbol{\delta}^*\rangle^2}{\|\tilde{\mathbf{x}}^*\|^4}\right)\right)\right].$$

$$\mathrm{II}' = \frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\mathsf{E}\left(\Delta_1\hat{\beta}\Delta_3\hat{\sigma}^2 - \frac{2\tilde{\beta}\Delta_3\hat{\sigma}^2(\tilde{\mathbf{x}}^{*\top}\boldsymbol{\delta}^*)}{\|\tilde{\mathbf{x}}^*\|^2}\right).$$

$$\mathrm{III}' = \frac{\tilde{\beta}}{\|\tilde{\mathbf{x}}\|^2}\mathsf{E}(\Delta_4\hat{\sigma}^2).$$

We start with I'. Here we have

$$\mathrm{I}' = \frac{f_0}{(n-2)\|\tilde{\mathbf{x}}^*\|^2}\mathsf{E}\left[\left(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\right)\left(\Delta_2\hat{\beta}_1 - \frac{2\Delta_1\hat{\beta}_1(\tilde{\mathbf{x}}^{*\top}\boldsymbol{\delta}^*)}{\|\tilde{\mathbf{x}}^*\|^2} - \frac{\tilde{\beta}\|\boldsymbol{\delta}^*\|^2}{\|\tilde{\mathbf{x}}^*\|^2} + \frac{4\tilde{\beta}(\tilde{\mathbf{x}}^{*\top}\boldsymbol{\delta}^*)^2}{\|\tilde{\mathbf{x}}^*\|^4}\right)\right]. \tag{3.51}$$

After using the definition of $\Delta_2\hat{\beta}_1$, and as such, Eq. (3.27), the first term becomes

$$\frac{f_0}{(n-2)\|\tilde{\mathbf{x}}^*\|^2}\mathsf{E}\Big(\big(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\big)\Delta_2\hat{\beta}_1\Big) = \frac{\sigma^4\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^4}. \tag{3.52}$$

Thus, we need to find

$$\frac{f_0}{(n-2)\|\tilde{\mathbf{x}}^*\|^4}\mathsf{E}\left[\left(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\right)\left(-2\Delta_1\hat{\beta}_1\langle\tilde{\mathbf{x}}^*,\boldsymbol{\delta}^*\rangle - \tilde{\beta}\|\boldsymbol{\delta}^*\|^2 + \frac{4\tilde{\beta}\langle\tilde{\mathbf{x}}^*,\boldsymbol{\delta}^*\rangle^2}{\|\tilde{\mathbf{x}}^*\|^2}\right)\right] =$$

$$= \frac{f_0}{(n-2)\|\tilde{\mathbf{x}}^*\|^4}\mathsf{E}\left[\frac{\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*}{\|\tilde{\mathbf{x}}^*\|^2}\cdot\Big(-2\,\boldsymbol{\gamma}^{*T}\mathbf{B}_n\boldsymbol{\delta}^* + 4\,\tilde{\beta}\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*\Big) - \tilde{\beta}(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\|\boldsymbol{\delta}^*\|^2\right]. \tag{3.53}$$

With the aid of identities (3.29), (3.46), and (3.47), one can show that

$$\frac{f_0}{(n-2)\|\tilde{\mathbf{x}}^*\|^4}\mathsf{E}\left[\left(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\right)\left(-2\Delta_1\hat{\beta}_1\langle\tilde{\mathbf{x}}^*,\boldsymbol{\delta}^*\rangle-\tilde{\beta}\|\boldsymbol{\delta}^*\|^2+\frac{4\tilde{\beta}\langle\tilde{\mathbf{x}}^*,\boldsymbol{\delta}^*\rangle^2}{\|\tilde{\mathbf{x}}^*\|^2}\right)\right]=$$

$$=\frac{\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}\Big((7-n)\tilde{\beta}-2f_0\tilde{\beta}^3\Big). \quad (3.54)$$

Substituting (3.52) and (3.54) in (3.51) yields

$$\mathrm{I}'=\frac{\sigma^4\Big((8-n)\tilde{\beta}-2f_0\tilde{\beta}^3\Big)}{\|\tilde{\mathbf{x}}^*\|^4}. \quad (3.55)$$

Similarly, we will compute II′. Since $\mathsf{E}(\Delta_3\hat{\sigma}^2\Delta_1\hat{\beta}_1)=0$, we have

$$\mathrm{II}'=-2\tilde{\beta}\|\tilde{\mathbf{x}}^*\|^{-4}\mathsf{E}[\Delta_3\hat{\sigma}^2(\tilde{\mathbf{x}}^{*\top}\boldsymbol{\delta}^*)].$$

Substitute (3.35) in II′ and use (3.29) and (3.48) to get

$$\mathrm{II}'=\frac{-2\tilde{\beta}}{(n-2)\|\tilde{\mathbf{x}}^*\|^4}\mathsf{E}\Big[\Delta_1\hat{\beta}_1\Big(f_1(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)-2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\Big)(\tilde{\mathbf{x}}^{*\top}\boldsymbol{\delta}^*)\Big]=$$

$$=\frac{-2\tilde{\beta}}{(n-2)\|\tilde{\mathbf{x}}^*\|^6}\mathsf{E}\Big[\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*\Big(f_1(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)-2f_0\,\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\Big)\Big]=$$

$$=\frac{2\tilde{\beta}f_0\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}. \quad (3.56)$$

Finally, we compute III′. From (3.36) and $\Delta_2\hat{\beta}_1=\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*$ and $\Delta_1\hat{\beta}_1^2=\frac{1}{\|\tilde{\mathbf{x}}^*\|^4}\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*$

$$\mathrm{III}'=\frac{\tilde{\beta}}{(n-2)\|\tilde{\mathbf{x}}\|^2}\mathsf{E}\Big[(f_1\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*-2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\Delta_2\hat{\beta}_1+$$
$$+(f_2\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*-2f_1\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\Delta_1\hat{\beta}^2\Big]=$$
$$=\frac{\tilde{\beta}}{(n-2)\|\tilde{\mathbf{x}}\|^2}\mathsf{E}\Big[(f_1\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*-2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*+$$
$$+\frac{1}{\|\tilde{\mathbf{x}}^*\|^4}(f_2\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*-2f_1\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*\Big].$$

Now using (3.27), (3.28), (3.49), and (3.50), one has

$$\mathrm{III}'=\frac{-\tilde{\beta}(2f_0+1)}{\|\tilde{\mathbf{x}}\|^4}. \quad (3.57)$$

Combining (3.55)–(3.57) gives us

$$\mathsf{E}(\Delta_4C)=-\frac{(2f_0\tilde{\beta}^3+(n-7)\tilde{\beta})\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}. \quad (3.58)$$

Hence,

$$\mathsf{E}(\check{\beta}_1)=\tilde{\beta}+\frac{\sigma^2\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}-\left(\frac{\sigma^2\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}-\frac{(2f_0\tilde{\beta}^3+(n-7)\tilde{\beta})\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}\right)=\tilde{\beta}+\frac{(2f_0\tilde{\beta}^3+(n-7)\tilde{\beta})\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}.$$

Accordingly, our results can be summarized in the following theorem.

Theorem 3.4. *Let* $\hat{\beta}$ *be the MLE of* β *and* $\check{\beta}=(1-\frac{\hat{\sigma}^2}{\|\mathbf{x}^*\|^2})\hat{\beta}$, *then*

$$\mathrm{bias}(\check{\beta}_1)=\frac{(2f_0\tilde{\beta}^3+(n-7)\tilde{\beta})\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}.$$

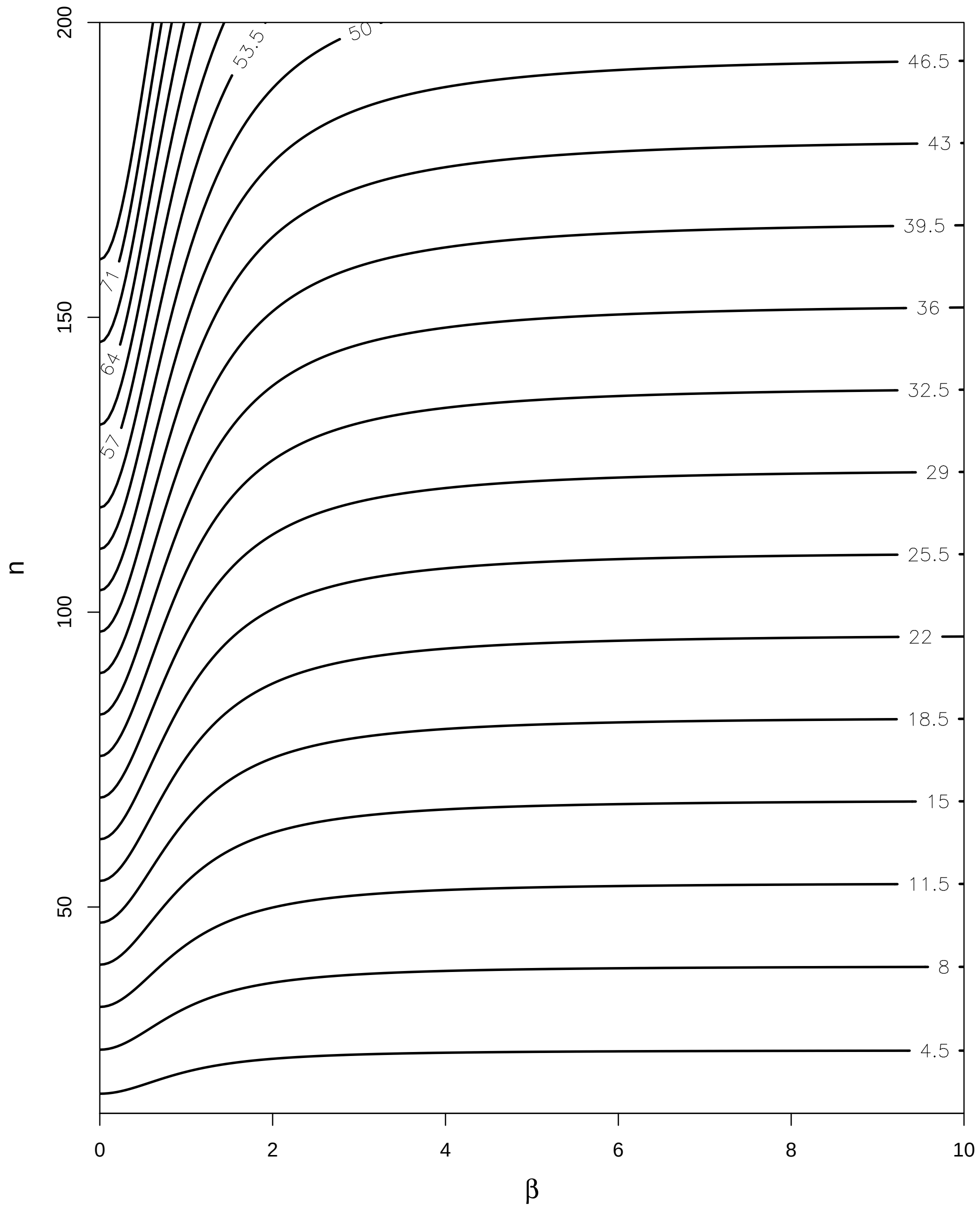

Figure 1. The contour map of the ratio of the fourth-order bias of $\check{\beta}_1$ to the fourth order bias if $\hat{\beta}_2$ over the region $(n, \tilde{\beta}) \in (15, 200) \times (0, 10)$.

$$\mathrm{MSE}(\check{\beta}_1) = \frac{(1+\tilde{\beta}^2)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2} + \left(1 + \frac{2\tilde{\beta}^2}{n-2}\right)\frac{(n-1)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^4}.$$

Note that, here, the bias of $\check{\beta}_1$ presented in Theorem 3.4 depends on $\|\mathbf{x}^*\|^2 = \sum_{i=1}^n (x_i - \bar{x})^2$, and as such, it is of order n. A simple comparison between the results in Theorem 3.4 and Equation (2.9) shows that $\mathrm{bias}(\check{\beta}_1) \sim \sigma^4/n$ while $\mathrm{bias}(\hat{\beta}_2) \sim \sigma^4/n^2$. Therefore, $\mathrm{bias}(\hat{\beta}_2) \sim \mathrm{bias}(\check{\beta}_1)/n$. Figure 1 represents the contour map of the ratio of the fourth-order bias of $\check{\beta}_1$ to the fourth-order bias of $\hat{\beta}_2$ over the region $\tilde{\beta} \in (0, 10)$ and $n \in (15, 200)$. It is clear that this ratio is always more than 2 and it reaches 71 in some parts of this region. These observations demonstrate that the bias of $\hat{\beta}_2$ is much smaller than the bias of $\check{\beta}_1$. Accordingly, $\hat{\beta}_2$ outperforms the AMLE $\check{\beta}_1$.

Finally, we should emphasize here that these results are valid whenever $\frac{\sigma}{L}$ is small.

4. Appendix

Proof of Theorem 3.2. At first, we will expand $\frac{1}{1+\hat{\beta}_1^2}$ about the true value $\tilde{\beta}$ to obtain

$$\frac{1}{1+\hat{\beta}_1^2} = f_0 + f_1\,\Delta_1\hat{\beta}_1 + f_1\,\Delta_2\hat{\beta}_1 + f_2\,\Delta_1\hat{\beta}_1^2 + \mathcal{O}_{\mathsf{P}}\big(\sigma^3\big) \stackrel{\text{def}}{=} h(\hat{\beta}_1) + \mathcal{O}_{\mathsf{P}}\big(\sigma^3\big),$$

where $f_0 = \frac{1}{1+\tilde{\beta}^2}$, $f_1 = -\frac{2\tilde{\beta}}{(1+\tilde{\beta}^2)^2}$, and $f_2 = -\frac{1-3\tilde{\beta}^2}{(1+\tilde{\beta}^2)^3}$. Keeping only terms of order σ^4 yields

$$\mathcal{F}(\hat{\beta}_1) = h(\hat{\beta}_1)\sum z_i^2 + \mathcal{O}_{\mathsf{P}}\big(\sigma^5\big), \tag{4.1}$$

where $z_i = a_i + b_i$ for each $i = 1, \ldots, n$.

The random numbers a_i and b_i have typical values of order σ and σ^2, respectively. Their formal expressions are defined by

$$a_i = \tilde{\beta}\delta_i^* + \tilde{x}_i^*\Delta_1\hat{\beta}_1 - \varepsilon_i^*, \tag{4.2}$$

$$b_i = \tilde{x}_i^*\Delta_2\hat{\beta}_1 + \delta_i^*\Delta_1\hat{\beta}_1\,. \tag{4.3}$$

We can find $\Delta_2\hat{\beta}_1$ by setting the derivative $\frac{\partial\mathcal{F}}{\partial\Delta_2\hat{\beta}_1}$ to zero. Using the fact $\sum a_i\tilde{x}_i^* = 0$ yields

$$\begin{aligned}\frac{d\mathcal{F}}{d\Delta_2\hat{\beta}_1} &= f_1\sum a_i^2 + 2h(\hat{\beta}_1)\sum z_i\tilde{x}_i^* + 2h(\hat{\beta}_1)\sum z_i\delta_i^* = \\ &= f_1\sum a_i^2 + 2f_0\|\tilde{\mathbf{x}}^*\|^2\Delta_2\hat{\beta}_1 + 4f_0\sum \tilde{x}_i^*\,\delta_i^*\,\Delta_1\hat{\beta}_1 + 2f_0\sum \delta_i^*\big(\tilde{\beta}\delta_i^* - \varepsilon_i^*\big)\,,\end{aligned}$$

where we omitted the remainder $\mathcal{O}_{\mathsf{P}}(\sigma^3)$ for brevity.

Let us denote $r_1 = \frac{f_1}{f_0} = \frac{-2\tilde{\beta}}{1+\tilde{\beta}^2}$. Then, we obtain

$$\begin{aligned}-2\|\tilde{\mathbf{x}}^*\|^2\Delta_2\hat{\beta}_1 &= r_1\sum a_i^2 + 4\sum \tilde{x}_i^*\delta_i^*\Delta_1\hat{\beta}_1 + 2\sum \delta_i^*(\tilde{\beta}\delta_i^* - \varepsilon_i^*) = \\ &= r_1\Big(-2\|\tilde{\mathbf{x}}^*\|^2\Delta_1\hat{\beta}_1^2 + \|\tilde{\mathbf{x}}^*\|^2\Delta_1\hat{\beta}_1^2 + \sum\big(\tilde{\beta}\delta_i^* - \varepsilon_i^*\big)^2\Big) + \\ &\quad + 2\sum \delta_i^*\big(\tilde{\beta}\delta_i^* - \varepsilon_i^*\big) + 4\sum \delta_i^*\tilde{x}_i^*\Delta_1\hat{\beta}_1 = \\ &\stackrel{\text{def}}{=} \mathbf{I} + \mathbf{II} + \mathbf{III}.\end{aligned} \tag{4.4}$$

The variables **I**, **II**, and **III** are defined and simplified below. In matrix notations, each term in the above expression can be written as a quadratic form of $\mathbf{h}^*$as follows.

$$\mathbf{I} = -r_1\|\tilde{\mathbf{x}}^*\|^2\Delta_1\hat{\beta}_1^2 = (\mathbf{h}^*)^\top\mathbf{Q}_1\mathbf{h}^*, \tag{4.5}$$

$$\mathbf{II} = r_1\sum\big(\tilde{\beta}\delta_i^* - \varepsilon_i^*\big)^2 + 2\sum \delta_i^*(\tilde{\beta}\delta_i^* - \varepsilon_i^*) = (\mathbf{h}^*)^\top\mathbf{Q}_2\mathbf{h}^*, \tag{4.6}$$

$$\mathbf{III} = 4\sum \delta_i^*\tilde{x}_i^*\,\Delta_1\hat{\beta}_1 = (\mathbf{h}^*)^\top\mathbf{Q}_3\mathbf{h}^*, \tag{4.7}$$

where

$$\mathbf{Q}_1 = \begin{bmatrix} \frac{-r_1\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{B}_n & \frac{r_1\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{B}_n \\ \frac{r_1\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{B}_n & \frac{-r_1}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{B}_n \end{bmatrix}, \qquad \mathbf{Q}_2 = \begin{bmatrix} (r_1\tilde{\beta}^2 + 2\tilde{\beta})\mathbf{I}_n & -(r_1\tilde{\beta}+1)\mathbf{I}_n \\ -(r_1\tilde{\beta}+1)\mathbf{I}_n & r_1\mathbf{I}_n \end{bmatrix}$$

and

$$\mathbf{Q}_3 = \begin{bmatrix} -\frac{4\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{B}_n & \frac{2}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{B}_n \\ \frac{2}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{B}_n & \mathbf{0}_n \end{bmatrix}.$$

Combining (4.4)–(4.7) gives

$$\Delta_2\hat{\beta}_1 = \mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^* = \mathbf{h}^\top\mathbf{N}_n^\top\mathbf{Q}\mathbf{N}_n\mathbf{h},$$

where $\mathbf{Q} = \frac{1}{2\|\tilde{\mathbf{x}}^*\|^2}\Big(\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3\Big)$.

Next, we find $\Delta_3\hat{\beta}_1$ by expanding $\frac{1}{1+\hat{\beta}_1^2}$ about the true value $\tilde{\beta}$, i.e.,

$$\frac{1}{1+\hat{\beta}_1^2} = f_0 + f_1\,\Delta\hat{\beta}_1 + f_2\,\Delta\hat{\beta}_1^2 + f_3\,\Delta\hat{\beta}_1^3 + \mathcal{O}_\mathsf{P}\big(\sigma^4\big) \stackrel{\text{def}}{=} h(\hat{\beta}_1) + \mathcal{O}_\mathsf{P}\big(\sigma^4\big),$$

where $\Delta\hat{\beta}_1 = \Delta_1\hat{\beta}_1 + \Delta_2\hat{\beta}_1 + \Delta_3\hat{\beta}_1 + \mathcal{O}_\mathsf{P}(\sigma^4)$ and $f_3 = \frac{f^{(3)}(\tilde{\beta})}{3!}$, while f_0, f_1, and f_2 are defined earlier. The random variable $h(\hat{\beta}_1)$ can be clearly rewritten as $h(\hat{\beta}_1) = = \tilde{h} + \Delta_1 h + \Delta_2 h + \Delta_3 h$, where $\tilde{h} = f_0$, $\Delta_1 h = f_1\Delta_1\hat{\beta}_1$, $\Delta_2 h = f_1\Delta_2\hat{\beta}_1 + f_2\Delta_1\hat{\beta}_1^2$, and $\Delta_3 h = f_1\Delta_3\hat{\beta}_1 + 2f_2\Delta_1\hat{\beta}_1\Delta_2\hat{\beta}_1 + f_3\Delta_1\hat{\beta}_1^3$. Thus,

$$\mathcal{F}(\hat{\beta}_1) = h(\hat{\beta}_1)\sum z_i^2 + \mathcal{O}_\mathsf{P}\big(\sigma^7\big), \tag{4.8}$$

where

$$z_i = \tilde{\beta}\delta_i^* + \tilde{x}_i^*\Delta_1\hat{\beta}_1 - \varepsilon_i^* + \tilde{x}_i^*\Delta_2\hat{\beta}_1 + \delta_i^*\Delta_1\hat{\beta}_1 + \tilde{x}_i^*\Delta_3\hat{\beta}_1 + \delta_i^*\Delta_2\hat{\beta}_1 + \delta_i^*\Delta_3\hat{\beta}_1. \tag{4.9}$$

Therefore, it is more appropriate to write $z_i = a_i + b_i + c_i + d_i$, where a_i and b_i are defined in (4.2), while

$$c_i = \tilde{x}_i^*\Delta_3\hat{\beta}_1 + \delta_i^*\Delta_2\hat{\beta}_1 \quad \text{and} \quad d_i = \delta_i^*\Delta_3\hat{\beta}_1$$

represent the cubic and the quadric forms of $\mathbf{e}_i^*$, respectively. Note that $\frac{\partial h(\beta)}{\partial\Delta_3\hat{\beta}_1} = f_1 + + 2f_2\Delta_1\hat{\beta}_1 + \mathcal{O}_\mathsf{P}(\sigma^2)$, hence differentiating $\mathcal{F}(\hat{\beta}_1)$ with respect to $\Delta_3\hat{\beta}_1$ yields

$$\begin{aligned}\frac{d\mathcal{F}(\beta)}{d\Delta_3\hat{\beta}_1} &= (f_1 + 2f_2\Delta_1\hat{\beta}_1)\sum z_i^2 + 2h(\beta)\sum z_i(\tilde{x}_i^* + \delta_i^*) + \mathcal{O}_\mathsf{P}\big(\sigma^7\big) = \\ &= 2f_2\Delta_1\hat{\beta}_1\sum a_i^2 + f_1\sum(a_i^2 + 2a_ib_i) + 2\tilde{h}\sum\big[(a_i + b_i + c_i)\tilde{x}_i^* + (a_i + b_i)\delta_i^*\big] + \\ &\quad + 2\Delta_1 h\sum\big[(a_i + b_i)\tilde{x}_i^* + a_i\delta_i^*\big] + 2\Delta_2 h\sum a_i\tilde{x}_i^* + \mathcal{O}_\mathsf{P}\big(\sigma^4\big).\end{aligned}$$

Here $\frac{d\mathcal{F}(\hat{\beta}_1)}{d\Delta_3\hat{\beta}_1}$ contains terms of order $\mathcal{O}_\mathsf{P}(\sigma)$, $\mathcal{O}_\mathsf{P}(\sigma^2)$, and $\mathcal{O}_\mathsf{P}(\sigma^3)$. Predictably, terms of the first and the second leading orders vanish. This follows from $\sum a_i\tilde{x}_i^* = 0$ and

$$2\tilde{h}\sum b_i\tilde{x}_i + 2\tilde{h}\sum a_i\delta_i^* + f_1\sum a_i^2 = 0,$$

which follows (4.4). Thus, $\frac{d\mathcal{F}(\hat{\beta}_1)}{d\Delta_3\hat{\beta}_1}$ becomes

$$\frac{d\mathcal{F}(\hat{\beta}_1)}{d\Delta_3\hat{\beta}_1} = 2f_2\Delta_1\hat{\beta}_1\sum a_i^2 + 2f_1\sum a_ib_i + 2f_0\sum(c_i\tilde{x}_i^* + b_i\delta_i^*) + 2f_1\Delta_1\hat{\beta}_1\sum(b_i\tilde{x}_i^* + a_i\delta_i^*).$$

Equating $\frac{d\mathcal{F}(\hat{\beta}_1)}{d\Delta_3\hat{\beta}_1}$ to zero, substituting the value of c_i, and solving for $\Delta_3\hat{\beta}_1$ yield

$$\begin{aligned}\Delta_3\hat{\beta}_1 = -\frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\Big(&r_2\Delta_1\hat{\beta}_1\sum a_i^2 + r_1\sum a_ib_i + \sum b_i\delta_i^* + \\ &+ r_1\Delta_1\hat{\beta}_1\sum(b_i\tilde{x}_i^* + a_i\delta_i^*) + \sum\tilde{x}_i^*\delta_i^*\Delta_2\hat{\beta}_1\Big),\end{aligned}$$

where we set $r_2 = f_2/f_0 = -(1-3\tilde{\beta}^2)/(1+\tilde{\beta}^2)^2$. Next, we substitute the values of a_i and b_i in the previous equation to get

$$\Delta_3\hat{\beta}_1 = -\frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\Big(\Delta_1\hat{\beta}_1\sum\big[(r_2\tilde{\beta}^2+2r_1\tilde{\beta}+1)(\delta_i^*)^2-2(r_2\tilde{\beta}+r_1)\delta_i^*\varepsilon_i^*+r_2(\varepsilon_i^*)^2\big]- \\ -r_2\|\tilde{\mathbf{x}}^*\|^2\Delta_1\hat{\beta}_1^3+3r_1\Big(\sum\delta_i^*\tilde{x}_i^*\Big)\Delta_1\hat{\beta}_1^2+r_1\|\tilde{\mathbf{x}}^*\|^2\Delta_1\hat{\beta}_1\Delta_2\hat{\beta}_1+2\Big(\sum\delta_i^*\tilde{x}_i^*\Big)\Delta_2\hat{\beta}_1\Big),$$

and further

$$\Delta_3\hat{\beta}_1 = -\frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\Big(\mathbf{G}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{A}\mathbf{h}^*-r_2\|\tilde{\mathbf{x}}^*\|^2(\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*\cdot\mathbf{G}^\top\mathbf{h}^*+ \\ +r_1\|\tilde{\mathbf{x}}^*\|^2\mathbf{G}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^*+3r_1\mathbf{E}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*+2\,\mathbf{E}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^*\Big),$$

where $\mathbf{E} = (\tilde{\mathbf{x}}^\top, \mathbf{0}^\top)^\top$ and $\mathbf{A} = \hat{\boldsymbol{\gamma}}_1 \otimes \mathbf{I}_n$ with

$$\hat{\boldsymbol{\gamma}}_1 = \begin{bmatrix} (r_2\tilde{\beta}^2+2r_1\tilde{\beta}+1) & -(r_2\tilde{\beta}+r_1) \\ -(r_2\tilde{\beta}+r_1) & r_2 \end{bmatrix}$$

(recall that $\Delta_1\hat{\beta}_1 = \mathbf{G}^\top\mathbf{h}^*$, $\Delta_2\hat{\beta}_1 = (\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^*$ and $\mathbf{B}_n = \tilde{\mathbf{x}}^* * (\tilde{\mathbf{x}}^*)^\top$). This completes the proof of the theorem. □

Proof of Theorem 3.3. To derive the MSE of $\hat{\beta}_1$, we need to compute $\mathsf{E}(\Delta_1\hat{\beta}_1\Delta_3\hat{\beta}_1)$, we start first with defining

$$\boldsymbol{\Gamma}_1 = \mathbf{A} - r_2\|\tilde{\mathbf{x}}^*\|^2\mathbf{G}\mathbf{G}^\top + 3r_1\mathbf{G}\mathbf{E}^\top = \boldsymbol{\gamma}_1\otimes\mathbf{B}_n + \hat{\boldsymbol{\gamma}}_1\otimes\mathbf{I}_n,$$
$$\boldsymbol{\Gamma}_2 = 2\,\mathbf{G}\mathbf{E}^\top + r_1\|\tilde{\mathbf{x}}^*\|^2\mathbf{G}\mathbf{G}^\top = \boldsymbol{\gamma}_2\otimes\mathbf{B}_n\,.$$

with

$$\boldsymbol{\gamma}_1 = \frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\begin{bmatrix} -r_2\tilde{\beta}^2-3r_1\tilde{\beta} & \frac{3}{2}r_1+r_2\tilde{\beta} \\ \frac{3}{2}r_1+r_2\tilde{\beta} & -r_2 \end{bmatrix}, \quad \boldsymbol{\gamma}_2 = \frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\begin{bmatrix} r_1\tilde{\beta}^2-2\tilde{\beta} & 1-r_1\tilde{\beta} \\ 1-r_1\tilde{\beta} & r_1 \end{bmatrix}. \tag{4.10}$$

Multiplying $\Delta_3\hat{\beta}_1$ by $\Delta_1\hat{\beta}_1$ yields

$$\Delta_1\hat{\beta}_1\Delta_3\hat{\beta}_1 = -\frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\Big((\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{A}\mathbf{h}^*-r_2\|\tilde{\mathbf{x}}^*\|^2((\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*)^2+ \\ +r_1\|\tilde{\mathbf{x}}^*\|^2(\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^*+3r_1(\mathbf{h}^*)^\top\mathbf{G}\mathbf{E}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*+ \\ +2(\mathbf{h}^*)^\top\mathbf{G}\mathbf{E}^\top\mathbf{h}^*\cdot(\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^*\Big).$$

Then, after simple algebra, we get

$$\Delta_3\hat{\beta}_1\Delta_1\hat{\beta}_1 = -\frac{1}{\|\tilde{\mathbf{x}}^*\|^2}\Big((\mathbf{h}^*)^\top\mathbf{G}\mathbf{G}^\top\mathbf{h}^*\cdot\mathbf{h}^*\boldsymbol{\Gamma}_1\mathbf{h}^*+(\mathbf{h}^*)^\top\mathbf{Q}\mathbf{h}^*\cdot\mathbf{h}^*\boldsymbol{\Gamma}_2\mathbf{h}^*\Big)\,. \tag{4.11}$$

Based on the previous argument, one gets

$$\mathsf{E}(\Delta_1\hat{\beta}_1\Delta_3\hat{\beta}_1) = -\frac{\sigma^4}{\|\tilde{\mathbf{x}}^*\|^2}\Big[\mathrm{tr}(\boldsymbol{\Gamma}_1\mathbf{N})\mathrm{tr}(\mathbf{G}\mathbf{G}^\top\mathbf{N})+\mathrm{tr}(\boldsymbol{\Gamma}_2\mathbf{N})\mathrm{tr}(\mathbf{Q}\mathbf{N})+ \\ +2\mathrm{tr}(\boldsymbol{\Gamma}_1\mathbf{N}\mathbf{G}\mathbf{G}^\top\mathbf{N})+2\mathrm{tr}(\boldsymbol{\Gamma}_2\mathbf{N}\mathbf{Q}\mathbf{N})\Big], \tag{4.12}$$

which simply becomes

$$\mathsf{E}(\Delta_1\hat{\beta}_1\Delta_3\hat{\beta}_1) = \frac{1+3\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^4}\sigma^4\,. \tag{4.13}$$

Also it is easy to verify that

$$\mathsf{E}((\Delta_2\hat{\beta})^2) = \frac{n+3\tilde{\beta}^2-1}{\|\tilde{\mathbf{x}}^*\|^4}\sigma^4\,. \tag{4.14}$$

Finally, if Eqs. (3.8), (4.14), and (4.13) are substituted into

$$\mathrm{MSE}(\hat{\beta}_1) = \mathsf{E}(\Delta_1\hat{\beta}_1^2) + \mathsf{E}(\Delta_2\hat{\beta}_1^2) + 2\,\mathsf{E}(\Delta_1\hat{\beta}_1\Delta_3\hat{\beta}_1) + \mathcal{O}\big(\sigma^6\big), \tag{4.15}$$

then the theorem will be established. □

Proof of Lemma 3.2. The proofs for all these identities go straightforward by using Theorems 3.1. Starting with (3.26), we notice that both $\mathbf{V}_n$ and $\mathbf{N}_n$ are idempotent matrices with $\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n) = n-2$, and $\boldsymbol{\gamma}^* \sim N\big(\mathbf{0}_n, (1+\tilde{\beta}^2)\sigma^2\mathbf{N}_n\big)$, then

$$\mathsf{E}\big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)^2\big] = \sigma^4(\tilde{\beta}^2+1)^2\big([\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)]^2 + 2\,\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)]\big) = n(n-2)(\tilde{\beta}^2+1)^2\sigma^4.$$

The second identity in (3.27) can be obtained after $\boldsymbol{\gamma}^*$ is expressed in terms of $\mathbf{h}^*$ (i. e., $\boldsymbol{\gamma}^* = (\check{\mathbf{a}}_1^\top \otimes \mathbf{I}_n)\mathbf{h}^*$). Therefore,

$$\mathsf{E}\big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*)\big) = \mathsf{E}\big((\mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*)\big).$$

After simple calculations

$$\mathrm{tr}(\boldsymbol{\Sigma}_1\mathbf{N}) = \mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n) = (\tilde{\beta}^2+1)(n-2),$$

and $\mathrm{tr}(\mathbf{Q}\mathbf{N}) = \frac{\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}$. Besides,

$$\mathrm{tr}(\boldsymbol{\Sigma}_1\mathbf{N}\mathbf{Q}\mathbf{N}) = \mathrm{tr}\big((\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathbf{a}\big)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n\mathbf{B}_n) + \mathrm{tr}\big((\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathbf{b}\big)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n) = 0.$$

This follows immediately from $\mathrm{tr}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathbf{b}) = 0$, $\mathrm{tr}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathbf{a}) = \frac{\tilde{\beta}(\tilde{\beta}^2+1)}{\|\tilde{\mathbf{x}}^*\|^4}$, and $\mathbf{V}_n\mathbf{N}_n\mathbf{B}_n = \mathbf{V}_n\mathbf{B}_n = \mathbf{0}_n$. As an immediate consequence,

$$\begin{aligned}\mathsf{E}\big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*)\big) &= \sigma^4(\tilde{\beta}^2+1)(\mathrm{tr}(\boldsymbol{\Sigma}_1\mathbf{N})\mathrm{tr}(\mathbf{Q}\mathbf{N}) + 2\,\mathrm{tr}(\boldsymbol{\Sigma}_1\mathbf{N}\mathbf{Q}\mathbf{N})) =\\ &= (n-2)\|\tilde{\mathbf{x}}^*\|^{-2}\tilde{\beta}(\tilde{\beta}^2+1)\sigma^4.\end{aligned}$$

Next, we compute (3.28). Again, since $\mathbf{V}_n\mathbf{N}_n\mathbf{B}_n = \mathbf{V}_n\mathbf{B}_n = \mathbf{0}_n$,

$$\mathsf{E}[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*)] = \sigma^4\mathrm{tr}\big[\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\big]^2\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)\mathrm{tr}(\mathbf{B}_n\mathbf{N}_n) = (n-2)\|\tilde{\mathbf{x}}^*\|^2(\tilde{\beta}^2+1)^2\sigma^4.$$

In the same way, (3.29) follows by expressing both of $\boldsymbol{\gamma}^*$ and $\boldsymbol{\delta}^*$ in terms of $\mathbf{h}^*$

$$\begin{aligned}\mathsf{E}\big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*)\big] &= \mathsf{E}\big[(\mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*)(\mathbf{h}^{*\top}(\mathbf{r}_1\otimes\mathbf{B}_n)\mathbf{h}^*)\big] =\\ &= \sigma^4\mathrm{tr}(\boldsymbol{\Sigma}_1\mathbf{N})\mathrm{tr}((\mathbf{r}_1\otimes\mathbf{B}_n)\mathbf{N}),\end{aligned}$$

where other terms equal zero because of $\mathbf{V}_n\mathbf{B}_n = \mathbf{0}_n$. Thus,

$$\begin{aligned}\mathsf{E}\big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*)\big] &= \sigma^4\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathrm{tr}(\mathbf{r}_1)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)\mathrm{tr}(\mathbf{B}_n\mathbf{N}_n) =\\ &= \mathsf{E}\big[(\mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*)(\mathbf{h}^{*\top}(\mathbf{r}_1\otimes\mathbf{B}_n)\mathbf{h}^*)\big] =\\ &= -\tilde{\beta}(\tilde{\beta}^2+1)(n-2)\|\tilde{\mathbf{x}}^*\|^2\sigma^4.\end{aligned}$$

(Here we used $\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top) = \tilde{\beta}^2+1$, $\mathrm{tr}(\mathbf{r}_1) = -\tilde{\beta}$, and $\mathbf{V}_n\mathbf{B}_n = \mathbf{0}_n$.

No, we derive (3.30), which becomes

$$\begin{aligned}\mathsf{E}\big(\mathbf{h}^{*\top}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1{}^\top)\otimes\mathbf{B}_n)\mathbf{h}^*\cdot\mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*\big) &= \sigma^4[\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1{}^\top)]^2\mathrm{tr}(\mathbf{B}_n)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n) =\\ &= (\tilde{\beta}^2+1)^2(n-2)\|\tilde{\mathbf{x}}^*\|^2\sigma^4.\end{aligned}$$

Lastly, we prove (3.31), which can be expressed as

$$\mathsf{E}\big(\mathbf{h}^{*\top}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1{}^\top)\otimes\mathbf{B}_n)\mathbf{h}^*\cdot\mathbf{h}^{*\top}\boldsymbol{\Sigma}_2\mathbf{h}^*\big) = \sigma^4\mathrm{tr}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1{}^\top))\mathrm{tr}(\mathbf{r}_1)\mathrm{tr}(\mathbf{B}_n)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n).$$

Thus,

$$\mathsf{E}\big(\mathbf{h}^{*\top}((\check{\mathbf{a}}_1\check{\mathbf{a}}_1{}^\top)\otimes\mathbf{B}_n)\mathbf{h}^*\cdot\mathbf{h}^{*\top}\boldsymbol{\Sigma}_2\mathbf{h}^*\big) = -\tilde{\beta}(\tilde{\beta}^2+1)(n-2)\|\tilde{\mathbf{x}}^*\|^2\sigma^4.$$

This completes the proof of the lemma. □

Proof of Lemma 3.3. Using (3.32), we write $\Delta_2\hat{\sigma}^2 = \frac{f_0}{n-2}\sum_{i=1}^n a_i^2$, where

$$\sum_{i=1}^n a_i^2 = \|\boldsymbol{\gamma}^*\|^2 - 2(\boldsymbol{\gamma}^{*\top}\tilde{\mathbf{x}}^*)\Delta_1\hat{\beta} + \|\tilde{\mathbf{x}}^*\|^2\Delta_1\hat{\beta}^2.$$

The identity $\|\tilde{\mathbf{x}}^*\|^2\Delta_1\hat{\beta} = \boldsymbol{\gamma}^{*\top}\tilde{\mathbf{x}}^*$ and $\mathbf{V}_n = \mathbf{I}_n - \frac{\mathbf{B}_n}{\|\tilde{\mathbf{x}}^*\|^2}$ leads to $\sum_{i=1}^n a_i^2 = \boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*$, and as such,

$$\Delta_2\hat{\sigma}^2 = \frac{f_0(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)}{n-2}.$$

The third order error term of $\hat{\sigma}^2$ is

$$\Delta_3\hat{\sigma}^2 = \frac{1}{n-2}\sum_i(f_1\Delta_1\hat{\beta}a_i^2 + 2f_0a_ib_i),$$

but

$$\sum_{i=1}^n a_ib_i = \Delta_1\hat{\beta}\sum_{i=1}^n a_i\boldsymbol{\delta}_i^* = \Delta_1\hat{\beta}(\Delta_1\hat{\beta}\tilde{\mathbf{x}}^{*T}\delta^* - \boldsymbol{\gamma}^{*\top}\boldsymbol{\delta}^*) = \Delta_1\hat{\beta}\left(\frac{\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\delta^*}{\|\tilde{\mathbf{x}}^*\|^2} - \boldsymbol{\gamma}^{*\top}\boldsymbol{\delta}^*\right).$$

Therefore, $\sum_i a_ib_i = -\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*$.

$$\Delta_3\hat{\sigma}^2 = \frac{1}{n-2}\sum_i(f_1\Delta_1\hat{\beta}a_i^2 + 2f_0a_ib_i) = \frac{\Delta_1\hat{\beta}}{n-2}\Big(f_1(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*) - 2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*\Big).$$

In the same approach, we get

$$\Delta_4\hat{\sigma}^2 = \frac{1}{n-2}\sum_i a_i^2(f_1\Delta_2\hat{\beta} + f_2\Delta_1\hat{\beta}^2) + 2f_1a_ib_i\Delta_1\hat{\beta} + 2f_0a_ic_i. \tag{4.16}$$

Using the fact $\sum_i a_i\tilde{x}_i^* = 0$ reduces this expression to

$$\begin{aligned}\Delta_4\hat{\sigma}^2 &= \frac{1}{n-2}\sum_i a_i^2(f_1\Delta_2\hat{\beta} + f_2\Delta_1\hat{\beta}^2) + 2f_1a_i\delta_i^*\Delta_1\hat{\beta}^2 + 2f_0a_i\delta_i^*\Delta_2\hat{\beta} = \\ &= \frac{1}{n-2}\sum_i(f_1a_i^2 + 2f_0a_i\delta_i^*)\Delta_2\hat{\beta} + (f_2a_i^2 + 2f_1a_i\delta_i^*)\Delta_1\hat{\beta}^2.\end{aligned}$$

From the previous expressions and $\Delta_1\hat{\beta}^2 = \frac{\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*}{\|\tilde{\mathbf{x}}^*\|^4}$ we get,

$$\Delta_4\hat{\sigma}^2 = \frac{1}{n-2}(f_1\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^* - 2f_0\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\Delta_2\hat{\beta} + (f_2\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^* - 2f_1\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\Delta_1\hat{\beta}^2.$$

This completes the proof of the lemma. □

Proof of Lemma 3.4. Equation (3.46) follows immediately from

$$\begin{aligned}&\mathsf{E}\big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)\|\boldsymbol{\delta}^*\|^2\big) = \mathsf{E}\big((\mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*)\cdot(\mathbf{h}^{*\top}\boldsymbol{\Sigma}_3\mathbf{h}^*)\big) = \\ &\quad= \sigma^4\big(\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathrm{tr}(\mathbf{s}_1\mathbf{s}_1^\top)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)\mathrm{tr}(\mathbf{N}_n) + 2\,\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top\mathbf{s}_1\mathbf{s}_1^\top)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)\big) = \\ &\quad= \sigma^4\Big((\tilde{\beta}^2+1)(n-2)(n-1) + 2\tilde{\beta}^2(n-2)\Big) = \\ &\quad= \sigma^4(n-2)\Big((\tilde{\beta}^2+1)(n-1) + 2\tilde{\beta}^2\Big).\end{aligned}$$

Next, we verify (3.47). Since $\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^* = \mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*$ and $\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^* = \mathbf{h}^{*\top}(\mathbf{s}_1\mathbf{s}_1^\top\otimes\mathbf{B}_n)\mathbf{h}^*$, we have

$$\begin{aligned}&\mathsf{E}\Big[(\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*)\Big] = \mathsf{E}\big[(\mathbf{h}^{*\top}\boldsymbol{\Sigma}_1\mathbf{h}^*)(\mathbf{h}^{*\top}(\mathbf{s}_1\mathbf{s}_1^\top\otimes\mathbf{B}_n)\mathbf{h}^*)\big] = \\ &\quad= \sigma^4\big(\mathrm{tr}(\boldsymbol{\Sigma}_1\mathbf{N})\mathrm{tr}((\mathbf{s}_1\mathbf{s}_1^\top\otimes\mathbf{B}_n)\mathbf{N}) + 2\,\mathrm{tr}(\boldsymbol{\Sigma}_1(\mathbf{s}_1\mathbf{s}_1^\top\otimes\mathbf{B}_n)\mathbf{N})\big) =\end{aligned}$$

$$= \sigma^4\big(\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)\mathrm{tr}(\mathbf{B}_n\mathbf{N}_n) + 2\,\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top\mathbf{s}_1\mathbf{s}_1^\top)\mathrm{tr}(\mathbf{V}_n\mathbf{B}_n\mathbf{N}_n)\big) =$$
$$= (n-2)\|\tilde{\mathbf{x}}^*\|^2(\tilde{\beta}^2+1)\sigma^4.$$

Now, we verify (3.48). Since $\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^* = \mathbf{h}^{*\top}(\mathbf{r}_1\otimes\mathbf{V}_n)\mathbf{h}^*$ and $\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^* = = \mathbf{h}^{*\top}(\mathbf{r}_1\otimes\mathbf{B}_n)\mathbf{h}^*$, we have

$$\mathsf{E}\Big[(\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}^{*\top}\mathbf{B}_n\boldsymbol{\delta}^*)\Big] = \mathsf{E}\big[(\mathbf{h}^{*\top}\boldsymbol{\Sigma}_2\mathbf{h}^*)(\mathbf{h}^{*\top}(\mathbf{r}_1\otimes\mathbf{B}_n)\mathbf{h}^*)\big] =$$
$$= \sigma^4(\mathrm{tr}(\boldsymbol{\Sigma}_2\mathbf{N})\mathrm{tr}((\mathbf{r}_1\otimes\mathbf{B}_n)\mathbf{N}) + 2\,\mathrm{tr}(\boldsymbol{\Sigma}_2(\mathbf{r}_1\otimes\mathbf{B}_n)\mathbf{N})) =$$
$$= \sigma^4\big(\mathrm{tr}(\mathbf{r}_1^2)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)\mathrm{tr}(\mathbf{B}_n\mathbf{N}_n) + 2\,\mathrm{tr}(\mathbf{r}_1^2)\mathrm{tr}(\mathbf{V}_n\mathbf{B}_n\mathbf{N}_n)\big) =$$
$$= (\tilde{\beta}^2+.5)(n-2)\|\tilde{\mathbf{x}}^*\|^2\sigma^4.$$

Next, we verify (3.49). Note that

$$(\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*) = \mathbf{h}^{*\top}(\mathbf{r}_1\otimes\mathbf{V}_n)\mathbf{h}^*\cdot\mathbf{h}^{*\top}(\mathbf{a}\otimes\mathbf{B}_n+\mathbf{b}\otimes\mathbf{I}_n)\mathbf{h}^*.$$

Thus

$$\mathsf{E}\Big((\boldsymbol{\delta}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\mathbf{h}^{*\top}\mathbf{Q}\mathbf{h}^*)\Big) = \mathrm{tr}(\mathbf{r}_1)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)(\mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{B}_n) + \mathrm{tr}(\mathbf{b})\mathrm{tr}(\mathbf{N}_n)) +$$
$$+ 2\,\mathrm{tr}(\mathbf{r}_1\mathbf{b})\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)$$
$$= \frac{(n-2)(1-\tilde{\beta}^2)\sigma^4}{\|\tilde{\mathbf{x}}^*\|^2},$$

which follows from $\mathrm{tr}(\mathbf{r}_1\mathbf{b}) = \frac{1}{2\|\tilde{\mathbf{x}}^*\|^2}$, $\mathrm{tr}(\mathbf{r}_1) = -\tilde{\beta}$, $\mathrm{tr}(\mathbf{a}) = \frac{\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^4}$, $\mathrm{tr}(\mathbf{b}) = 0$, and $\mathbf{V}_n\mathbf{B}_n = \mathbf{0}_n$ (see (3.11)–(3.15)).

Lastly, we prove (3.50). Since

$$\mathsf{E}\Big((\boldsymbol{\gamma}^{*\top}\mathbf{V}_n\boldsymbol{\gamma}^*)(\boldsymbol{\delta}^{*\top}\mathbf{B}_n\boldsymbol{\gamma}^*)\Big) = \sigma^4\big(\mathrm{tr}(\check{\mathbf{a}}_1\check{\mathbf{a}}_1^\top)\mathrm{tr}(\mathbf{r}_1)\mathrm{tr}(\mathbf{V}_n\mathbf{N}_n)\mathrm{tr}(\mathbf{B}_n)\big) =$$
$$= -\sigma^4(n-2)\|\tilde{\mathbf{x}}^*\|^2\tilde{\beta}(\tilde{\beta}^2+1).$$

This completes the proof of the lemma. □

The author would like to thank the reviewers and the editor for providing many suggestions that resulted in improving this paper.

References

1. A. Al-Sharadqah, *A new perspective in functional EIV linear model: Part I*, Comm. Statist. Theory Methods, **47** (2017), no. 14, 7039–7062.
2. A. Al-Sharadqah, N. Chernov, *Statistical analysis of curve fitting methods in Errors-In-Variables models*, Theory Probab. Math. Statist., **84** (2011), 4–17.
3. A. Al-Sharadqah, N. Chernov, Q. Huang, *Errors-In-Variables regression and the problem of moments*, Brazilian Journal of Probability and Statistics, **84** (2013), 401–415.
4. Y. Amemiya, W. A. Fuller, *Estimation for the nonlinear functional relationship*, Annals Statist., **16** (1988), 147–160.
5. T. W. Anderson, *Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics*, J. R. Statist. Soc. B, **38** (1976), 1–36.
6. T. W. Anderson, T. Sawa, *Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions*, Econometrica, **41** (1973), 683–714.
7. T. W. Anderson, T. Sawa, *Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient*, J. R. Statist. Soc. B, **44** (1982), 52–62.
8. C.-L. Cheng, J. W. Van Ness, *Statistical Regression with Measurement Error*, Arnold, London, 1999.
9. C. L. Cheng, A. Kukush, *Non-existence of the first moment of the adjusted least squares estimator in multivariate errors-in-variables model*, Metrika, **64** (2006), 41–46.
10. N. Chernov, *Circular and linear regression: Fitting circles and lines by least squares*, CRC Monographs on Statistics & Applied Probability, vol. 117, Chapman & Hall, 2010.

11. N. Chernov, *Fitting circles to scattered data: parameter estimates have no moments*, Metrika, **73** (2011), 373–384.
12. N. Chernov, C. Lesort, *Statistical efficiency of curve fitting algorithms*, Comp. Stat. Data Anal., **47** (2004), 713–728.
13. L. J. Gleser, *Functional, structural and ultrastructural errors-in-variables models*, Proc. Bus. Econ. Statist. Sect. Am. Statist. Ass., 1983, 57–66.
14. S. van Huffel, ed., *Total Least Squares and Errors-in-Variables Modeling*, Kluwer, Dordrecht, 2002.
15. K. Kanatani, *Statistical optimization for geometric computation: theory and practice*, Elsevier, Amsterdam, 1996.
16. K. Kanatani, *Cramer-Rao lower bounds for curve fitting*, Graph. Mod. Image Process., **60** (1998), pp. 93–99.
17. K. Kanatani, *For geometric inference from images, what kind of statistical model is necessary?*, Syst. Comp. Japan, **35** (2004), 1–9.
18. A. Kukush, E.-O. Maschke, *The efficiency of adjusted least squares in the linear functional relationship*, J. Multivar. Anal., **87** (2003), 261–274.
19. J. R. Magnus, H. Neudecker, *The commutation matrix: some properties and applications*, Annals Statist., **7** (1979), 381–394.
20. K. M. Wolter, W. A. Fuller, *Estimation of Nonlinear Errors-in-Variables Models*, Annals Statist., **10** (1982), 539–548.
21. E. Zelniker, V. Clarkson, *A statistical analysis of the Delogne-Kåsa method for fitting circles*, Digital Signal Proc., **16** (2006), 498–522.

Department of Mathematics and Interdisciplinary Research Institute for the Sciences, California State University-Northridge, Northridge, CA 91330-8313, USA
E-mail address: ali.alsharadqah@csun.edu

Received 3.05.2017

ПОРІВНЯЛЬНЕ ДОСЛІДЖЕННЯ ДВОХ НЕЩОДАВНО РОЗРОБЛЕНИХ ОЦІНОК ДЛЯ КОЕФІЦІЄНТА НАХИЛУ У ФУНКЦІОНАЛЬНІЙ ЛІНІЙНІЙ МОДЕЛІ З ПОХИБКАМИ У ЗМІННИХ

А. А. АЛЬ-ШАРАДКАХ

Анотація. Нещодавно в [1] були розроблені дві оцінки для коефіцієнта нахилу лінії у функціональній моделі з похибками у змінних. Вони обидві є незміщеними до порядку σ^4, де σ — стандартне відхилення похибки. Одна з оцінок була побудована як функція оцінки максимальної вірогідності (ОМВ). Тому її названо покращеною ОМВ. Другу оцінку побудовано за допомогою зовсім іншого підходу. Хоча обидві оцінки є незміщеними до порядку σ^4, остання оцінка є набагато точнішою, ніж покращена ОМВ. Тут ми вивчаємо ці дві оцінки більш строго і показуємо, чому одна оцінка перевершує іншу.

СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ ДВУХ НЕДАВНО РАЗРАБОТАННЫХ ОЦЕНОК ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА В ФУНКЦИОНАЛЬНОЙ ЛИНЕЙНОЙ МОДЕЛИ С ОШИБКАМИ В ПЕРЕМЕННЫХ

А. А. АЛЬ-ШАРАДКАХ

Аннотация. Недавно в [1] были разработаны две оценки для коэффициента наклона линии в функциональной модели с ошибками в переменных. Они обе являются несмещёнными до порядка σ^4, где σ — стандартное отклонение ошибки. Одна из оценок была построена как функция оценки максимального правдоподобия (ОМП). Поэтому её назвали улучшенной ОМП. Вторая оценка была построена с помощью совсем другого подхода. Хотя обе оценки являются несмещёнными до порядка σ^4, последняя оценка — намного точнее, чем улучшенная ОМП. Здесь мы изучаем эти две оценки более строго и показываем, почему одна оценка превосходит другую.