Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

A new perspective in functional EIV linear model: Part I

Ali Al-Sharadqah

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2017, VOL. 46, NO. 14, 7039–7062
http://dx.doi.org/10.1080/03610926.2016.1143009

Taylor & Francis
Taylor & Francis Group

A new perspective in functional EIV linear model: Part I

Ali Al-Sharadqah

Department of Mathematics, East Carolina University, Greenville, NC, USA

ABSTRACT

Simple linear regression in the functional errors-in-variables (EIV) model is revisited from a different perspective, where the problem is addressed by using the small-sigma model instead of large sample theory. A general analysis is developed to study the slope's estimator that minimizes a family of objective functions, of which the least-squares fit and the maximum likelihood estimator are minimizers of such special functions. General formulas for the higher-order terms of the bias, the variance, and the mean square error are derived. Accordingly, two efficient estimators are proposed after implementing the pre- and the post-bias elimination techniques. Numerical tests confirm the superiority of the proposed estimators over others.

ARTICLE HISTORY
Received 2 February 2015
Accepted 11 January 2016

KEYWORDS
Bias elimination; Computer vision; Errors-in-variables models; Maximum likelihood estimator; Mean squared error; Simple linear regression; Small-noise model.

MATHEMATICS SUBJECT CLASSIFICATION
62J0; 62H12

1. Introduction

Suppose that two variables, y and x, are linearly related through the relationship $y = \alpha + \beta x$, where α is the y-intercept and β is the slope of the line. In the classical regression, the regressor x is considered error free, while the dependent variable y is contaminated by some error. If n experimental observations, say $\mathbf{m}_i = (x_i, y_i)^T$, $i = 1, \ldots, n$, were recorded, then the "maximum likelihood estimator" (MLE) of (α, β) is equivalent to minimizing the objective function

$$\mathcal{F}_0(\alpha, \beta) = \sum_{i=1}^{n} d_i^2, \quad d_i = y_i - \alpha - \beta x_i. \tag{1}$$

Its minimum is attained at $\hat{\alpha}_0 = \bar{y} - \beta\bar{x}$ and $\hat{\beta}_0 = s_{xy}/s_{xx}$. These estimators $(\hat{\alpha}_0, \hat{\beta}_0)$ are known as *the classical least-squares fit* (LS) or *ordinary least-squares regression*. Here we used the standard statistical notations for the sample means; that is, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the sample means of $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, \ldots, y_n)^T$. Also, we used $s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$, and $s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ as the components of the so-called "scatter matrix." Moreover, we will use the notations $x_i^* = x_i - \bar{x}$ and $y_i^* = y_i - \bar{y}$ as the "centered" coordinates of x_i and y_i.

The classical linear regression was published by Legendre in 1805 and Gauss in 1809. The estimators $\hat{\alpha}_0$ and $\hat{\beta}_0$ have excellent statistical properties. They are optimal in all senses. However, in the case where some regressors have been measured with errors, the standard assumption leads to inconsistent estimates. For instance, the biases of the least-squares estimators

CONTACT Ali Al-Sharadqah alsharadqaha@ecu.edu Department of Mathematics, East Carolina University, Greenville, NC 27858, USA.

persist even in very large samples, and thus the above estimators $(\hat{\alpha}_0, \hat{\beta}_0)$ lose their appealing features.

This "new" model is known as measurement error (ME) model or errors-in-variables (EIV) model, which dates back to the 1870s when Adcock (Adcock, 1877, 1878) derived the formulas for the slope and the intercept estimates. However, all of his calculations were based on geometric rather than statistical considerations. Statisticians have been studying this problem since 1901 (Pearson, 1901). Intensive research focused on a line fitting in the twentieth century because of its applications in economics, sciences, image processing, and computer vision (Cheng and Van Ness, 1999; van Huffel, 2002). See Gillard (2006) for a nice overview and (Jung, 2007; Soderstrom, 2007; Amiri-Simkooeii and Jazaeri, 2012) for other recent publications.

The EIV regression problem is quite different from and more difficult than the classical regression. In the EIV linear model, the observable measurement (x, y) is regarded as a perturbation of a true point $(\tilde{x}, \tilde{y})$ that lies on the true line, $\tilde{y} = \tilde{\alpha} + \tilde{\beta}\tilde{x}$. But the true point $(\tilde{x}, \tilde{y})$ is unobservable and the observed point (x, y) is regarded as the true point $(\tilde{x}, \tilde{y})$ contaminated by some noisy errors. That is,

$$x_i = \tilde{x}_i + \delta_i, \qquad y_i = \tilde{y}_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ represent the true values of intercept α and slope β. The noisy vector $(\delta_i, \varepsilon_i)$, $i = 1, \ldots, n$ is assumed to be i.i.d. normally distributed random vectors with zero mean (0,0) and variance–covariance matrix $\Sigma = \text{diag}(\sigma_\delta^2, \sigma_\varepsilon^2)$. Here, we will assume that the ratio $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ is known. For simplicity, we write $\sigma_\delta^2 = \sigma^2$ and $\sigma_\varepsilon^2 = \lambda\sigma^2$. Also, through our discussion, we often use some vector notations. We will express Equation (2) 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 $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)^T$ and $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^T$.

There are also two standard statistical assumptions about the nature of true points $(\tilde{x}_i, \tilde{y}_i)$: If true points are unknown but *fixed* then the model is known in the literature as a *functional model*. In this case, the true points are regarded as nuisance or latent variables. The second model is called *structural model*, which regards the true points as random variables (i.e., realizations of some random numbers); see (Fuller, 1987; Gillard, 2006) for more details. The primary concern in this article is only in the functional model.

The MLEs, say $(\hat{\alpha}_1, \hat{\beta}_1)$, of (α, β) in the functional EIV have been known in the literature since long time (Deming, 1943). They have the formal expression

$$\hat{\beta}_1 = \frac{s_{yy} - \lambda s_{xx} + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda s_{xy}^2}}{2s_{xy}}, \qquad \hat{\alpha}_1 = \bar{y} - \hat{\beta}_1\bar{x}. \tag{3}$$

However, only in 1976 were explicit formulas derived for the pdfs of the MLE $\hat{\alpha}_1$ and $\hat{\beta}_1$; (see Anderson, 1976; Anderson and Sawa, 1982). It turns out that they are not normal and do not belong to any standard family of probability distributions. Those formulas are overly complicated and involve double-infinite series, and it was promptly noted (Anderson, 1976) that they are not very useful for practical purposes. Moreover, the estimates $\hat{\alpha}_1$ and $\hat{\beta}_1$ do not have finite moments (Anderson, 1976). As a result, they have infinite variances and infinite mean squared errors, though such an erratic behavior is barely seen in practice (Anderson, 1976; Anderson and Sawa, 1982; Chernov, 2010; Al-Sharadqah and Chernov, 2011). Another difficulty in studying the simple linear regression in the functional EIV model stems from dealing with large sample problem $n \to \infty$. In this case, an infinite number of latent variables and the problem of inconsistency (Cheng and Van Ness, 1999) immediately appear.

This article, however, is tailored for image processing applications, where the number of observed points (pixels) is limited, but the noise is small. Therefore, we will study estimators whenever $\sigma \to 0$, which is known as the *small-sigma model*. This regime is not new; for instance, Anderson and Sawa (Anderson, 1976; Anderson and Sawa, 1982) investigated the asymptotic properties of the MLE $\hat{\alpha}$ and $\hat{\beta}$ assuming that n is fixed and $\sigma \to 0$. They treated σ as a small parameter and employed the Taylor expansion (up to σ^4) to derive approximations for those distributions. However, statisticians have completely focused only on the large sample theory as the main tool to study the statistical properties of estimator. Therefore, they completely ignored this asymptotic regime despite its wide range of applications.

The small-sigma model has a great impact on many research topics in image processing, signal processing, computer vision, and many other research topics (Kanatani, 1993; Chernov, 2010). 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. In the analog of consistency of an estimator, we call an estimator *geometrically consistent* if it returns the true values of parameters if all points are observed with no noise (i.e., the dataset is noiseless). Informally, $\lim_{\sigma \to 0} \hat{\boldsymbol{\theta}}(\mathbf{m}_1, \ldots \mathbf{m}_n) = \tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ is the true value of the parameter vector.

1.1. Goals and outlines

Our goal is to develop a simplified error analysis that nonetheless allows us to effectively develop new estimators for the line slope with excellent properties. Our ultimate goal is to generalize our approach in this article to more complex situations, such as multivariate linear regression, polynomial regression, or other non linear regression problems. Thus, studying simple linear regression in depth is a must.

In this article, we will study the EIV linear regression in a new perspective. We will adopt the small-sigma regime to study the statistical properties of any estimator in a general setting. We will also study the minimizers of a family of objective functions parameterized by a weight function, say $g(\beta)$, and as such, a family of estimators can be obtained. That is, we will consider in this article a general class of objective functions:

$$\mathcal{F}(\alpha, \beta) = g(\beta) \sum_{i=1}^{n} d_i^2, \quad d_i = y_i - \alpha - \mathbf{x}_i \tag{4}$$

where g is a smooth function that depends only upon β.

The LS fit $\hat{\beta}_0 = \frac{s_{xy}}{s_{xx}}$ minimizes the objective function $\mathcal{F}_0$ in (1) , which is a special case of (4) with $g(\beta) = 1 = g_0(\beta)$ (say) . Another special case is the MLE that comes if we consider $g(\beta) = \frac{1}{\beta^2+\lambda} = g_1(\beta)$ (say). That is, in the functional linear EIV model, the MLE of the slope β and the intercept α, with the aforementioned assumptions, minimize

$$\mathcal{F}_1(\alpha, \beta) = (\beta^2 + \lambda)^{-1} \sum_{i=1}^{n} d_i^2. \tag{5}$$

Then after simple algebraic manipulations, the MLE (Cheng and Van Ness, 1999) is one of the roots of the quadratic equation

$$s_{xy}\beta^2 - (s_{yy} - \lambda s_{xx})\beta - \lambda s_{xy} = 0. \tag{6}$$

Now one can easily show that the MLE of β is the expression given in (3).

In the literature, researchers usually studied an estimator that either minimizes a specific objective function, such as the MLE, or solves certain equation, such as the method of the moments estimators (MMEs). As a result, this limits the number of choices to obtain other good estimators, especially when the sample size is small. However, our treatment for the general family of the objective functions as in (4) gives us more flexibility to propose estimators with excellent properties.That is, having an unbiased estimator up to certain orders (as $\mathcal{O}(\sigma^4)$ or even $\mathcal{O}(\sigma^4/n^2)$) with the smallest mean square error (MSE) (up to order $\mathcal{O}(\sigma^6)$) is our ultimate goal.

Furthermore, instead of approximating the probability distribution of an estimator (such as the MLE that has infinite moment!), the Taylor expansion will be employed here to approximate the general form of the estimators themselves. Consequently, general formulas for the mean, the variance, and the MSE of estimators will be presented after very lengthy calculations. It turns out that these formulas are very useful. For instance, the formula for the second-order error bias helps us to propose the first efficient estimator with zero second-order bias. This can be done by appropriately choosing its associated function $g_2(\beta)$ (as will be shown shortly). We will call this process of correction the bias procedure *pre-bias-elimination technique* and we will denote the resulting estimator by $\hat{\beta}_2$.

On the other hand, if one substitutes $g_1(\beta) = (\beta^2 + \lambda)^{-1}$ in the general formula of the second-order bias, then the second-order bias of the MLE $\hat{\beta}_1$ will be obtained. This estimator can be corrected to obtain a more accurate estimator. That is, subtracting an unbiased estimator of the MLE's bias from the MLE itself gives *second-order unbiased estimator*. Therefore, we call this adjustment *post-bias elimination*.

Our numerical experiments for these two corrections show the superiority of the pre-bias correction over the MLE and its post-bias correction. This motivates us enough to go one step further and derive general formulas for the higher-order bias (up to order $\mathcal{O}(\sigma^4/n^2)$) and the MSE (up to order $\mathcal{O}(\sigma^6)$). Smaller MSE gives better estimator. Based on this principle and the formulas for higher-order terms of the MSE, we will rigorously demonstrate why our choice for $g_2(\beta)$ gives an estimator $\hat{\beta}_2$ with zero second-order bias and has the smallest MSE among all other estimators.

Moreover, with the aid of those formulas, we will show that although the proposed estimator $\hat{\beta}_2$ has zero second-order bias, its higher-order bias of magnitude σ^4/n^2 persists. Hence, an adjustment is necessary to correct this term, and as such, a more accurate estimator can be obtained if this bias is removed. We denote this adjustment by $\check{\beta}_2$. We finally conclude that the pre-bias elimination technique—if its is followed by post-bias elimination step—yields more accurate estimators. Their excellent performances appear in all cases and for any sample size, and this can be greatly seen when the same size is relatively small.

The structure of this article is outlined as follows: In Section 2, some previous results are summarized and our new approach is proposed, followed by our first proposed estimator for the slope. In Section 3, general formulas for higher-order terms of the MSE and the bias are derived. It also discusses another new estimator and a comparison between several estimators. Section 4 is devoted to an experimental validation of our error analysis scheme. We probe it on several test cases to demonstrate its superiority. These numerical simulations confirm the superiority of our proposed estimators. They also validate our approach; that is, an estimator produced by adopting pre-bias elimination procedure (with or without post-bias elimination) will always be better than the MLE and its adjustment. Section 4 concludes our findings . The Appendix provides technical proofs.

2. General perspective

In the small-sigma model, Kanatani (1998) derived a general *Cramér–Rao* (CR) lower bound for arbitrary curves for *any unbiased estimators*. In geometric fitting problems, all estimators, however, are biased. This makes the natural CR lower bound not practical. In the early 2000s, Chernov and Lesort (2004) realized that Kanatani's formula does not work for any practical estimator in curve-fitting problems. To overcome this situation, Chernov and Lesort (2004) employed first-order analysis for any geometrically consistent estimators.

They have shown that Kanatani's formula works for all geometrically consistent estimators, *up to the leading order*. Thus, they called it the *Kanatani–Cramér–Rao (KCR) lower bound*. From that time on, the KCR has been used as a measure of efficiency for any meaningful estimator. In the course of linear regression, the KCR lower bound means that the first leading term of the "approximative" covariance matrix $\mathbf{V}$ for any geometrically consistent estimators of (α, β) has a natural bound. The KCR lower bound is given by

$$\mathbf{V} \geq \sigma^2 \mathbf{V}_{\min}, \qquad \mathbf{V}_{\min} = \frac{\tilde{\beta}^2 + \lambda}{s_{\tilde{x}\tilde{x}}} \begin{bmatrix} \overline{\tilde{x}\tilde{x}} & -\bar{\tilde{x}} \\ -\bar{\tilde{x}} & 1 \end{bmatrix}. \tag{7}$$

Here $\mathbf{A} \geq \mathbf{B}$ means $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix.

Chernov and Lesort (2004) considered a general curve fitting. They proved that an estimator $\hat{\boldsymbol{\theta}}$ is efficient if and only if it minimizes the weighted objective function $\mathcal{F}(\boldsymbol{\theta}) = \sum w_i d_i^2$. The weights w_i's must be proportional to the square of gradient of d_i with respect to $\mathbf{m}_i$, i.e., $w_i = \frac{a(\boldsymbol{\theta})}{\|\nabla_{\tilde{\mathbf{m}}_i} d_i\|^2}$ and $a(\boldsymbol{\theta})$ in an arbitrary function on $\boldsymbol{\theta}$.

In linear regression, $d_i = y_i - \alpha - \beta x_i$ and $w_i = \frac{a(\alpha,\beta)}{\beta^2+\lambda}$. Note here, if $a(\alpha, \beta) = 1$, the objective function associated with the MLE, i.e., $\mathcal{F}_1$ (5), is obtained, while if $a(\alpha, \beta) = (\beta^2 + \lambda)$, the objective function associated with LS estimator, i.e., $\mathcal{F}_0$ (1), is obtained. This demonstrates that both of the MLE and LS are efficient in the KCR sense.

Note that both of the denominator $\beta^2 + \lambda$ and the function $a(\alpha, \beta)$ depend only upon α and β but not on the observations; thus one can define the weight w_i as $w_i = g(\alpha, \beta)$, where $g(\alpha, \beta)$ is an arbitrary smooth function of α and β. The weight function g will play a key role to obtain estimators for (α, β) that are unbiased and their variances, up to leading term, attain their minimal values as given in Equation (7). To keep our analysis simple, we will assume at this point that the weight function g depends upon β only.

Using the step-by-step minimization technique, we eliminate α by first differentiating $\mathcal{F}$ with respect to α and equating its derivative to zero to get $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. Substitute its value in $\mathcal{F}(\alpha, \beta)$ to obtain

$$\mathcal{F}(\beta) = g(\beta) \sum d_i^{*2}, \tag{8}$$

We emphasize here that $\mathcal{F}(\beta)$ is a function of β only and $d_i^* = y_i^* - \beta x_i^*$. Now suppose that $\hat{\beta}$ is the slope's estimator that minimizes (8) and let $\hat{\beta}_{\mathrm{Q}} = \tilde{\beta} + \Delta_1\hat{\beta} + \Delta_2\hat{\beta}$ be its quadratic approximation obtained by expanding the $\hat{\beta}$ up to the second-order term. That is, $\hat{\beta} = \hat{\beta}_{\mathrm{Q}} + \mathcal{O}_{\mathrm{P}}(\sigma^3)$, where $\Delta_1\hat{\beta}$ and $\Delta_2\hat{\beta}$ are the first- and the second-order errors, respectively. More precisely ,

$$\Delta_1\hat{\beta} = \sum \beta_{x_i}\delta_i + \sum \beta_{y_i}\varepsilon_i, \qquad \Delta_2\hat{\beta} = \tfrac{1}{2}\Bigg[\sum_{i,j} \beta_{x_i x_j}\delta_i\delta_j + \sum_{i,j} \beta_{x_i y_j}\delta_i\varepsilon_j + \sum_{i,j} \beta_{y_i y_j}\varepsilon_i\varepsilon_j\Bigg]. \tag{9}$$

From now on, we denote the partial derivative of $\hat{\beta}$ with respect to x_i by $\hat{\beta}_{x_i} = \frac{\partial \hat{\beta}}{\partial x_i}$ and its second partial derivatives with respect to x_i and y_j by $\hat{\beta}_{x_i y_j} = \frac{\partial^2 \hat{\beta}}{\partial x_i \partial y_j}$. Also, we will use the notations β_{x_i} and $\beta_{x_i y_j}$ as their true values, i.e., when those expressions are evaluated at the true values, $\tilde{\beta}$ and $(\tilde{x}_k, \tilde{y}_k)$ for all $k = 1, \ldots, n$. Accordingly,

$$\mathrm{Var}(\hat{\beta}) = \mathbb{E}\big(\Delta_1 \hat{\beta}^2\big) + \mathcal{O}(\sigma^4) = \sigma^2 \sum_{i=1}^{n} \left(\beta_{x_i}^2 + \lambda \beta_{y_i}^2\right) + \mathcal{O}(\sigma^4) \tag{10}$$

and

$$\mathrm{Bias}(\hat{\beta}) = \mathbb{E}(\Delta_2 \hat{\beta}) = \tfrac{\sigma^2}{2}\left[\sum_i \beta_{x_i x_i} + \lambda \beta_{y_i y_i}\right] \tag{11}$$

are the first leading terms for each of the variance and the bias of $\hat{\beta}$.

It turns out that the explicit formulas for the expressions of bias and the variance can be expressed in terms of $\tilde{\beta}$, $g(\tilde{\beta})$, and $\|\tilde{\mathbf{x}}^*\|$. That is, since $\hat{\beta}$ minimizes Equation (8), it satisfies the "first derivative test," say $\mathcal{F}_\beta(\hat{\beta}) = 0$. Differentiating this equation with respect to x_i and y_i gives us

$$\mathcal{F}_{\beta\beta} \hat{\beta}_{x_i} + \mathcal{F}_{\beta x_i} = 0, \quad \mathcal{F}_{\beta\beta} \hat{\beta}_{y_i} + \mathcal{F}_{\beta y_i} = 0. \tag{12}$$

When the true values are substituted in Equation (12), we get

$$\beta_{x_i} = -\tilde{\mathcal{F}}_{\beta x_i} / \tilde{\mathcal{F}}_{\beta\beta}, \quad \beta_{y_i} = -\tilde{\mathcal{F}}_{\beta y_i} / \tilde{\mathcal{F}}_{\beta\beta}, \tag{13}$$

where $\tilde{\mathcal{F}}_{\beta\beta}$ and $\tilde{\mathcal{F}}_{\beta x_i}$ are $\mathcal{F}_{\beta\beta}$ and $\mathcal{F}_{\beta x_i}$, respectively, evaluated at the true point $(\tilde{x}_i, \tilde{y}_i)$. These second-order partial derivatives of $\mathcal{F}$ can be derived by first differentiating Equation (8) directly with respect to β, i.e.,

$$\mathcal{F}_\beta(\beta) = g' \sum d_i^{*2} - 2g \sum d_i^* x_i^*. \tag{14}$$

Then, differentiating $\mathcal{F}_\beta$ with respect to β, x_i, and y_i gives us

$$\mathcal{F}_{\beta\beta} = g'' \sum d_i^{*2} - 4g' \sum d_i^* x_i^* + 2g \sum [x_i^*]^2, \tag{15}$$

$$\mathcal{F}_{\beta x_i} = -2\beta g' d_i^* - 2g y_i^* + 4\beta g x_i^*, \tag{16}$$

$$\mathcal{F}_{\beta y_i} = 2g' d_i^* - 2g x_i^*. \tag{17}$$

Note here that at the true points $\tilde{d}_i^* = \tilde{y}_i^* - \tilde{\beta}\tilde{x}_i^* = 0$ for every i; hence $\tilde{\mathcal{F}}_\beta = 0$, which is consistent with the equation $\mathcal{F}_\beta(\tilde{\beta}) = 0$. Moreover, if we use the notation $\sum[\tilde{x}_i^*]^2 = s_{xx} = nS$, where $S \sim \mathcal{O}(1)$ represents the "spread" or "scatter" of the true x-coordinates $\tilde{x}_1, \ldots, \tilde{x}_n$, then the true values of the partial derivatives in Equations (15)–(17) are

$$\tilde{\mathcal{F}}_{\beta\beta} = 2\tilde{g} nS, \quad \tilde{\mathcal{F}}_{\beta x_i} = 2\tilde{\beta}\tilde{g}\tilde{x}_i^*, \quad \tilde{\mathcal{F}}_{\beta y_i} = -2\tilde{g}\tilde{x}_i^*, \tag{18}$$

where $\tilde{g} = g(\tilde{\beta})$. Now, according to Equation (13), we obtain

$$\beta_{x_i} = -\frac{\tilde{\beta}\tilde{x}_i^*}{nS}, \quad \beta_{y_i} = \frac{\tilde{x}_i^*}{nS}. \tag{19}$$

The main term of the variance is

$$\mathrm{Var}(\hat{\beta}) = \left[\sum \beta_{x_i}^2 + \lambda \sum \beta_{y_i}^2\right]\sigma^2 = \frac{(\tilde{\beta}^2 + \lambda)\sigma^2}{nS} = \mathrm{MSE}(\Delta_1 \hat{\beta}), \tag{20}$$

which is the most important term in the expression of the MSE and it represents the leading term of the variance. Since it coincides the KCR bound, and it does not depend on g, then any estimator minimizing $\mathcal{F}$ is optimal in the KCR sense. However, estimators such as MLE and LS behave differently in practice. To analytically compare between estimators, we must track the second most important term of the MSE. Therefore, we go one step further and study the main term of the bias.

To compute the second-order bias, the second-order partial derivatives of $\mathcal{F}$ must be evaluated. Differentiating the first equation in Equation (12) with respect to x_j, the second equation in Equation (12) with respect to y_j, and the first equation in Equation (12) with respect to y_j, respectively, gives us the following:

$$0 = \mathcal{F}_{\beta\beta}\hat{\beta}_{x_ix_j} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}_{x_i}\hat{\beta}_{x_j} + \mathcal{F}_{\beta\beta x_i}\hat{\beta}_{x_j} + \mathcal{F}_{\beta\beta x_j}\hat{\beta}_{x_i} + \mathcal{F}_{\beta x_ix_j}, \tag{21}$$

$$0 = \mathcal{F}_{\beta\beta}\hat{\beta}_{x_iy_j} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}_{x_i}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta x_i}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta y_j}\hat{\beta}_{x_i} + \mathcal{F}_{\beta x_iy_j}, \tag{22}$$

$$0 = \mathcal{F}_{\beta\beta}\hat{\beta}_{y_iy_j} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}_{y_i}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta y_i}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta y_j}\hat{\beta}_{y_i} + \mathcal{F}_{\beta y_iy_j}. \tag{23}$$

Consequently, their true values are

$$\beta_{x_ix_j} = -\frac{\tilde{\mathcal{F}}_{\beta\beta\beta}\beta_{x_i}\beta_{x_j} + \tilde{\mathcal{F}}_{\beta\beta x_i}\beta_{x_j} + \tilde{\mathcal{F}}_{\beta\beta x_j}\beta_{x_i} + \tilde{\mathcal{F}}_{\beta x_ix_j}}{\tilde{\mathcal{F}}_{\beta\beta}}, \tag{24}$$

$$\beta_{x_iy_j} = -\frac{\tilde{\mathcal{F}}_{\beta\beta\beta}\beta_{x_i}\beta_{y_j} + \tilde{\mathcal{F}}_{\beta\beta x_i}\beta_{y_j} + \tilde{\mathcal{F}}_{\beta\beta y_j}\beta_{x_i} + \tilde{\mathcal{F}}_{\beta x_iy_j}}{\tilde{\mathcal{F}}_{\beta\beta}}, \tag{25}$$

$$\beta_{y_iy_j} = -\frac{\tilde{\mathcal{F}}_{\beta\beta\beta}\beta_{y_i}\beta_{y_j} + \tilde{\mathcal{F}}_{\beta\beta y_i}\beta_{y_j} + \tilde{\mathcal{F}}_{\beta\beta y_j}\beta_{y_i} + \tilde{\mathcal{F}}_{\beta y_iy_j}}{\tilde{\mathcal{F}}_{\beta\beta}}. \tag{26}$$

The true values of these partial derivatives are summarized in the following lemma.

Lemma 2.1. *Denotes the Kronecker symbol with δ_{ij} and define $\hat{\delta}_{ij} = \delta_{ij} - \frac{1}{n}$. Then, we have*

$$\tilde{\mathcal{F}}_{\beta\beta\beta} = 6\tilde{g}'nS,\quad \tilde{\mathcal{F}}_{\beta\beta x_i} = 4(\tilde{\beta}\tilde{g}' + \tilde{g})\tilde{x}_i^*,\quad \tilde{\mathcal{F}}_{\beta\beta y_i} = -4\tilde{g}'\tilde{x}_i^*,$$
$$\tilde{\mathcal{F}}_{\beta x_ix_j} = [2\tilde{\beta}^2\tilde{g}' + 4\tilde{\beta}\tilde{g}]\hat{\delta}_{ij},\quad \tilde{\mathcal{F}}_{\beta x_iy_j} = -[2\beta\tilde{g}' + 2\tilde{g}]\hat{\delta}_{ij},\quad \tilde{\mathcal{F}}_{\beta y_iy_j} = 2\tilde{g}'\hat{\delta}_{ij}. \tag{27}$$

The proof of the deferred lemma is moved to the Appendix. Now we can compute derivatives according to Equations (24)–(26):

$$\beta_{x_ix_j} = \frac{-(\tilde{\beta}^2\tilde{g}' + 2\tilde{\beta}\tilde{g})\hat{\delta}_{ij} + (\tilde{\beta}^2\tilde{g}' + 4\tilde{\beta}\tilde{g})\frac{\tilde{x}_i^*\tilde{x}_j^*}{nS}}{\tilde{g}nS}. \tag{28}$$

$$\beta_{x_iy_j} = \frac{(\tilde{\beta}\tilde{g}' + \tilde{g})\hat{\delta}_{ij} - (\tilde{\beta}\tilde{g}' + 2\tilde{g})\frac{\tilde{x}_i^*\tilde{x}_j^*}{nS}}{\tilde{g}nS}. \tag{29}$$

$$\beta_{y_iy_j} = \frac{-\tilde{g}'\hat{\delta}_{ij} + \tilde{g}'\frac{\tilde{x}_i^*\tilde{x}_j^*}{nS}}{\tilde{g}nS}. \tag{30}$$

Let us define $\kappa = (\beta^2 + \lambda)g$, i.e., $\kappa' = (\beta^2 + \lambda)g' + 2\beta g$. The function κ and its derivative depends on g and they will play a key role in the sequel analysis. Now substituting Equations (28) and (30) in Equation (11) reduces the second-order bias to

$$\text{Bias}\,(\hat{\beta}) = \frac{-n\tilde{\kappa}' + 2(\tilde{\beta}^2 + \lambda)\tilde{g}' + 6\tilde{\beta}\tilde{g}}{2\tilde{g}nS}\sigma^2 = \frac{-\tilde{\kappa}'\sigma^2}{2\tilde{g}S} + \frac{(\tilde{\kappa}' + \tilde{\beta}\tilde{g})\sigma^2}{\tilde{g}nS} + \mathcal{O}(\sigma^4). \tag{31}$$

It is clear that the second-order bias can be decomposed into two components. The most important part is of the order of magnitude σ^2, which represents the *essential bias*. That is,

$$\text{Bias}_{\text{ess}}(\hat{\beta}) = \frac{-\tilde{\kappa}'\sigma^2}{2\tilde{g}S} = -\frac{(\tilde{\beta}^2+\lambda)\tilde{g}' + 2\beta\tilde{g}}{2\tilde{g}S}\sigma^2. \tag{32}$$

The second part of Equation (31), i.e., $\frac{(\tilde{\kappa}'+\tilde{\beta}\tilde{g})\sigma^2}{\tilde{g}nS}$, is of the order σ^2/n and we call it the *non essential bias*.

We can now eliminate the essential bias by solving the separable first-order differential equation $(\tilde{\beta}^2+\lambda)\tilde{g}' = -2\tilde{\beta}\tilde{g}$, from which $\frac{d}{d\beta}\ln\tilde{g}(\beta) = -\frac{d}{d\beta}\ln(\tilde{\beta}^2+\lambda)$; therefore $g(\beta) = \frac{C}{\beta^2+\lambda}$, where C is an arbitrary constant. This constant obviously does not affect the minimum of the function (8); hence we can set $C=1$. This leads to $g_1(\beta) = (\beta^2+\lambda)^{-1}$, which is the weight associated with the MLE (cf., Equation (5)). Since $\tilde{g}_1' = -2\tilde{\beta}(\tilde{\beta}^2+\lambda)^{-2}$, then $\tilde{\kappa}'=0$. After simple calcluation, we obtain $\text{Bias}(\hat{\beta}_1) = \frac{\sigma^2\tilde{\beta}}{nS} = \frac{\sigma^2\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}$. This means that the MLE has only the non essential bias.

Post-Bias Elimination. It is standard in statistics to use a bias-correction technique to get an unbiased estimator. Accordingly, since the bias of the quadratic approximation of $\hat{\beta}_1$ is $\frac{\sigma^2\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}$, then one can verify that *the adjusted MLE* (AMLE) defined by $\check{\beta}_1 = (1-\frac{\hat{\sigma}_1^2}{\|\mathbf{x}^*\|^2})\hat{\beta}_1$ has zero second-order bias, where $\hat{\sigma}_1^2$ is an estimate of σ^2. Here

$$\hat{\sigma}_1^2 = (n-2)^{-1}\mathcal{F}_1(\hat{\beta}_1) = \frac{1}{(n-2)(\hat{\beta}_1^2+\lambda)}\sum_{i=1}^{n}(y_i^* - \hat{\beta}_1 x_i^*)^2.$$

This can be easily verified if one proves that $\mathbb{E}(\hat{\sigma}_1^2) = \sigma^2 + \mathcal{O}(\sigma^4)$; hence $\mathbb{E}(\check{\beta}_1 - \tilde{\beta}) = \mathcal{O}(\sigma^4)$. This bias-correction technique used here is regarded as post-bias elimination.

Pre-Bias Elimination. Another unbiased estimator can be proposed but in a different approach. The idea here is equating Equation (31) with zero. Then, one tries to find a weight function g that solves the new differential equation. In this case, the function $g(\beta)$ must satisfy the differential equation $(n-2)\kappa' + 2\beta g = 0$. In terms of g, it can be written as

$$(n-2)(\beta^2+\lambda)g' = -2(n-3)\beta g. \tag{33}$$

Separating the variables and solving the resulting differential equation give $\frac{d}{d\beta}\ln g(\beta) = -\frac{n-3}{n-2}\frac{d}{d\beta}\ln(\beta^2+\lambda)$. Therefore,

$$g(\beta) = C(\beta^2+\lambda)^{-\frac{n-3}{n-2}}, \tag{34}$$

where $n \geq 3$ and C is an irrelevant factor. Hence, we can set $C=1$. Accordingly, the weight function $g_2(\beta) = (\beta^2+\lambda)^{-\frac{n-3}{n-2}}$ is associated with the estimator that has zero second-order bias. Note that whenever $n=3$, g_2 equals 1, which turns to be the problem of classical least squares. This is the particular case when the least-squares estimator has zero second-order bias. The next theorem summarizes our first contribution in this article.

Theorem 2.1. *Up to an irrelevant scalar factor, the fit (8) has 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 Equation (8) has zero second-order bias if and only if* $g = g_2$ *(up to an irrelevant scalar factor).*

Now we turn our attention to how we compute $\hat{\beta}_2$. Since $\hat{\beta}_2$ minimizes

$$\mathcal{F}_2(\beta) = (\beta^2+\lambda)^{-\frac{n-3}{n-2}}\sum d_i^{*2}, \tag{35}$$

then it is the critical that solves $\frac{\partial \mathcal{F}_2}{\partial \beta} = 0$. The latter turns to be the cubic equation

$$s_{xx}\beta^3 + (n-4)s_{xy}\beta^2 - [(n-3)s_{yy} - \lambda(n-2)s_{xx}]\beta - \lambda(n-2)s_{xy} = 0. \tag{36}$$

It might seem a hopeless problem. However, recall that the MLE $\hat{\beta}_1$ is the solution of the quadratic equation given in Equation (6), and as such, Equation (36) can be regarded as a "correction" of Equation (6), for finite-sample size. For relatively large n, Equation (36) reduces to Equation (6). Now, Equation (36) can be solved either numerically and choose the solution of in Equation (3) as its initial guess; or using the mathematical expression of the solution of the cubic equation and selecting the root that minimizes the objective function.

Theorem 2.1 proposes a new estimator $\hat{\beta}_2$ that is optimal in two aspects: $\hat{\beta}_2$ attains the KCR lower bound and has zero second-order bias. It also sheds more light about the linear EIV models by distinguishing the MLE given in Equation (3) from the classical least squares that minimizes Equation (8).

It is a common fact that the MLE $\hat{\beta}_1$ is more accurate than the least-squares fit $\hat{\beta}_0$. This is what Anderson and others explained in great detail in 1976–1984 (Anderson, 1984). This can also be demonstrated as a special case of our general conclusion. To compare the biases of the two classical estimators, note that

$$\text{Bias}\,(\hat{\beta}_0) = \frac{-\tilde{\beta}(n-3)}{nS}\,\sigma^2 + \mathcal{O}(\sigma^4), \qquad \text{Bias}\,(\hat{\beta}_1) = \frac{\tilde{\beta}}{nS}\,\sigma^2 + \mathcal{O}(\sigma^4). \tag{37}$$

Remark 2.1. For $n = 2$, both estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, just find the unique line passing through the two observed points; hence they give identical estimates, and for this reason, their biases coincide. The above formulas show that both biases are equal to $\frac{\tilde{\beta}}{2S}\sigma^2 + \mathcal{O}(\sigma^4)$.

Remark 2.2. For $n = 3$, *the classical least squares* $(\hat{\alpha}_0, \beta_0)$ *has zero second-order bias and is better than the MLE (3)*, which is a weird exceptional case. For $n > 3$, the classical estimate is biased toward smaller values of β, and the bias is heavy. It does not get smaller as n increases. On the other hand, the MLE given in Equation (3) is biased toward larger values of $\tilde{\beta}$, but the bias gets smaller as n increases.

3. Higher-order analysis

In Section 2, a general form of the second-order bias has been derived. Interestingly enough, the MLE $\hat{\beta}_1$ only has the non essential second-order bias, while the proposed estimator $\hat{\beta}_2$ has zero second-order bias. However, we still do not know how they behave in practice and what are their MSEs. Indeed, since the leading term of the MSE does not depend on $g(\beta)$ and it attains the KCR lower bound, the higher-order error analysis for the MSE must be derived. In this section, we will derive the general expression of the MSE for any estimator minimizing $\mathcal{F}$ in Equation (4). Then, we will analytically compare between the three estimators, the LS $\hat{\beta}_0$ and the MLE $\hat{\beta}_1$ and the proposed estimator $\hat{\beta}_2$.

Using the Taylor expansion of $\hat{\beta} = \tilde{\beta} + \Delta_1\hat{\beta} + \cdots$, one can decompose the MSE into

$$\text{MSE}(\hat{\beta}) = \text{MSE}(\Delta_1\hat{\beta}) + \text{MSE}(\Delta_2\hat{\beta}) + 2\,\mathbb{E}(\Delta_1\hat{\beta}\Delta_3\hat{\beta}) + \mathcal{O}(\sigma^6). \tag{38}$$

At this point, recall that $\text{MSE}(\Delta_1\hat{\beta}) = \frac{(\tilde{\beta}^2+\lambda)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2}$. The other terms must be carefully handled by tracing all terms with orders of magnitude σ^4, σ^4/n, and σ^4/n^2, while other less important terms are discarded. We start with $\text{MSE}(\Delta_2\hat{\beta}) = \mathbb{E}(\Delta_2\hat{\beta}^2)$. i.e.,

$$\mathbb{E}(\Delta_2\hat{\beta}^2) = [\text{Bias}(\Delta_2\hat{\beta})]^2 + \text{Var}(\Delta_2\hat{\beta})), \tag{39}$$

where

$$\mathrm{Var}(\Delta_2\hat{\beta}) = \frac{\sigma^4}{2}\sum_{i,j}\left(\beta_{x_ix_j}^2 + 2\beta_{x_iy_j}^2 + \beta_{y_iy_j}^2\right). \tag{40}$$

The variance of $\Delta_2\hat{\beta}$ consists of terms with different magnitudes: σ^4/n and σ^4/n^2. Also, its components are simply

$$\sum_{i,j}^{n}\beta_{x_ix_j}^2 = \frac{[\tilde{\gamma}']^2}{\tilde{g}^2nS^2} - \frac{2\tilde{\beta}^2(\tilde{g}'\tilde{\gamma}' + 2\tilde{\beta}\tilde{g}\tilde{g}' + 2\tilde{g}^2)}{\tilde{g}^2n^2S^2}, \qquad \sum_{i,j}^{n}\beta_{y_iy_j}^2 = \frac{[\tilde{g}']^2}{\tilde{g}^2nS^2}\left(1-\frac{2}{n}\right), \tag{41}$$

and

$$\sum_{i,j}^{n}\beta_{x_iy_j}^2 = \frac{[\beta\tilde{g}' + \tilde{g}]^2}{\tilde{g}^2nS^2} + \frac{\tilde{g}^2 - 2(\tilde{\beta}\tilde{g}' + \tilde{g})^2}{\tilde{g}^2n^2S^2}, \tag{42}$$

where $\gamma = \beta^2 g$ and $\kappa = (\beta^2 + \lambda)g$. Substitute Equations (41) and (42) into Equation (40) and use the identities $\tilde{\gamma}' = \tilde{\beta}^2\tilde{g}' + 2\tilde{\beta}\tilde{g}$ and $\tilde{\kappa}' = \tilde{\gamma}' + \lambda\tilde{g}'$ to get

$$\mathrm{Var}(\Delta_2\hat{\beta}) = \frac{\sigma^4}{2}\left(\frac{[\tilde{\kappa}']^2 + 2\lambda\tilde{g}^2}{n\tilde{g}^2S^2} + \frac{-2[\tilde{\kappa}']^2 + 2(2\tilde{\beta}^2 - \lambda)\tilde{g}^2}{n^2\tilde{g}^2S^2}\right), \tag{43}$$

and, as such,

$$\mathrm{Var}(\Delta_2\hat{\beta}) = \frac{\sigma^4}{nS^2}\left(\lambda + \frac{2\tilde{\beta}^2 - \lambda}{n}\right) + \frac{\sigma^4(n-2)[\tilde{\kappa}']^2}{2n^2\tilde{g}^2S^2}. \tag{44}$$

Squaring the bias of $\Delta_2\hat{\beta}$ (cf., Equation (31)) gives

$$[\mathbb{E}(\Delta_2\hat{\beta})]^2 = \frac{\sigma^4}{4\tilde{g}^2n^2S^2}(n^2[\tilde{\kappa}']^2 - 4n\tilde{\kappa}'(\tilde{\kappa}' + \tilde{\beta}\tilde{g}) + 4[\tilde{\kappa}' + \tilde{\beta}\tilde{g}]^2). \tag{45}$$

Combine the expressions (44) and (45) to get the MSE of $\Delta_2\hat{\beta}$. That is,

$$\mathrm{MSE}(\Delta_2\hat{\beta}) = \frac{\sigma^4}{nS^2}\left(\lambda + \frac{2\tilde{\beta}^2 - \lambda}{n}\right) + \frac{\sigma^4}{4\tilde{g}^2nS^2}\left((n-2)[\tilde{\kappa}']^2 - 4\tilde{\kappa}'\tilde{\beta}\tilde{g} + \frac{4(2\tilde{\beta}\tilde{g}\tilde{\kappa}' + \tilde{\beta}^2\tilde{g}^2)}{n}\right). \tag{46}$$

For simplicity, define the function $\tau(\beta)$ such that its derivative τ' satisfies $\tau' = 0$, where

$$\tau' = (n-2)(\beta^2 + \lambda)g' + 2(n-3)\beta g, \quad \text{i.e.,} \quad \tau' = (n-2)\kappa' - 2\beta g, \tag{47}$$

then $\mathrm{MSE}(\Delta_2\hat{\beta})$ can be written as

$$\mathrm{MSE}(\Delta_2\hat{\beta}) = \frac{\sigma^4}{nS^2}\left(\lambda + \frac{2\tilde{\beta}^2 - \lambda}{n}\right) + \frac{\sigma^4}{4\tilde{g}^2nS^2}\left(\tilde{\kappa}'(\tilde{\tau}' - 2\tilde{\beta}\tilde{g}) + \frac{4(2\tilde{\beta}\tilde{g}\tilde{\kappa}' + \tilde{\beta}^2\tilde{g}^2)}{n}\right). \tag{48}$$

It is worth mentioning here that the first term of the $\mathrm{MSE}(\Delta_2\hat{\beta})$ is free of g, while the second term is the product of κ' and $(\tau' - 2\beta g)$. The definitions of κ' and τ' have a meaningful explanation here. For instance, solving $\kappa' = 0$ represents equating the second-order essential bias with zero, and as such solving this ordinary differential equation (ODE) gives us $g_1(\beta)$. On the other hand, the solution of the ODE $\tau' = 0$ (cf., Equation (33)) is $g_2(\beta)$.

Remark 3.1. For $n > 4$,

$$\mathrm{MSE}(\Delta_2\hat{\beta}_0) > \mathrm{MSE}(\Delta_2\hat{\beta}_1) > \mathrm{MSE}(\Delta_2\hat{\beta}_2).$$

Remark 3.1 demonstrates that the quadratic approximation of $\hat{\beta}_2$ performs better than the quadratic approximation of $\hat{\beta}_1$, which, in turn, outperforms $\hat{\beta}_0$. However, it does not show that MSE($\hat{\beta}_2$) is less than MSE($\hat{\beta}_1$), because other terms of order σ^4 must be taken in account in the expression of the MSE of $\hat{\beta}$ before we assess the performance of such estimators. These terms come from $2\,\mathbb{E}(\Delta_1\hat{\beta}\Delta_3\hat{\beta})$. Since

$$\Delta_3\hat{\beta} = \tfrac{1}{6}\left[\sum_{i,j,k} \beta_{x_ix_jx_k}\delta_i\delta_j\delta_k + 3\beta_{x_ix_jy_k}\delta_i\delta_j\varepsilon_k + 3\beta_{x_iy_jy_k}\delta_i\varepsilon_j\varepsilon_k + \beta_{y_iy_jy_k}\varepsilon_i\varepsilon_j\varepsilon_k\right],$$

then

$$\mathbb{E}(\Delta_1\hat{\beta}\Delta_3\hat{\beta}) = \mathbb{E}\left[\frac{1}{6}\sum_{i,j,k,l} \beta_{x_ix_jx_k}\beta_{x_l}\delta_i\delta_j\delta_k\delta_l + 3\beta_{x_ix_jy_k}\beta_{y_l}\delta_i\delta_j\varepsilon_k\varepsilon_l \right.$$
$$\left. + \ 3\beta_{x_iy_jy_k}\beta_{x_l}\delta_i\delta_l\varepsilon_j\varepsilon_k + \beta_{y_iy_jy_k}\beta_{y_l}\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l\right].$$

Recall that δ_i and ε_i are i.i.d. random variables, then $\mathbb{E}(\Delta_1\hat{\beta}\Delta_3\hat{\beta})$

$$= \tfrac{1}{6}\left[\sum_{i,j,k,l} (\beta_{x_ix_jx_k}\beta_{x_l} + \lambda^2\beta_{y_iy_jy_k}\beta_{y_l})\mathbb{E}[\delta_i\delta_j\delta_k\delta_l] + 3\beta_{x_ix_jy_k}\beta_{y_l}\mathbb{E}[\delta_i\delta_j\varepsilon_k\varepsilon_l]\right.$$
$$\left. + 3\beta_{x_iy_jy_k}\beta_{x_l}\mathbb{E}[\delta_i\delta_l\varepsilon_j\varepsilon_k]\right]$$
$$= \tfrac{\sigma^4}{6}\left[\sum_{i=1}^{n} 3(\beta_{x_ix_ix_i}\beta_{x_i} + \lambda^2\beta_{y_iy_iy_i}\beta_{y_i}) + \sum_{i\neq j} 3(\beta_{x_ix_ix_j}\beta_{x_j} + \lambda^2\beta_{y_iy_iy_j}\beta_{y_j})\right.$$
$$\left. + \sum_{i,j} 3\lambda(\beta_{x_ix_iy_j}\beta_{y_j} + \beta_{x_iy_jy_j}\beta_{x_i})\right],$$

from which

$$\mathbb{E}(\Delta_1\hat{\beta}\Delta_3\hat{\beta}) = \tfrac{\sigma^4}{2}\sum_{i,j=1}^{n} \zeta_{i,j}, \tag{49}$$

where

$$\zeta_{i,j} = \beta_{x_ix_ix_j}\beta_{x_j} + \lambda^2\beta_{y_iy_iy_j}\beta_{y_j} + \lambda\beta_{x_ix_iy_j}\beta_{y_j} + \lambda\beta_{x_iy_jy_j}\beta_{x_i}. \tag{50}$$

The next lemma helps us compute each term of $\zeta_{i,j}$. Its full proof is quite lengthy so we only provide the derivation of $\mathcal{O}(\sigma^4)$ terms in the Appendix.

Lemma 3.1. *Define $\gamma = \beta^2 g$, then*

$$-S\tilde{\mathcal{F}}_{\beta\beta}\sum_{i,j=1}^{n} \beta_{x_iy_jy_j}\beta_{x_i} = 2\left(1-\frac{2}{n}\right)\left[\tilde{\beta}^2\tilde{g}'' + 2\tilde{\beta}\tilde{g}' - \frac{\tilde{\beta}^2(\tilde{g}')^2}{\tilde{g}}\right]. \tag{51}$$

$$-S\tilde{\mathcal{F}}_{\beta\beta}\sum_{i,j=1}^{n} \beta_{y_iy_iy_j}\beta_{y_j} = 2\left(1-\frac{2}{n}\right)\left[\tilde{g}'' - \frac{(\tilde{g}')^2}{\tilde{g}}\right]. \tag{52}$$

$$-S\tilde{\mathcal{F}}_{\beta\beta}\sum_{i,j=1}^{n} \beta_{x_ix_iy_j}\beta_{y_j} = 2\left(1-\frac{2}{n}\right)\left[(\tilde{\gamma}'' - \frac{\tilde{\gamma}'\tilde{g}'}{\tilde{g}}\right] - \frac{4}{n}g. \tag{53}$$

Moreover,

$$-S\tilde{\mathcal{F}}_{\beta\beta}\sum_{i,j=1}^{n}\beta_{x_ix_ix_j}\beta_{x_j}=2\left(1-\frac{2}{n}\right)\left[\tilde{\beta}^2\tilde{\gamma}''+2\tilde{\beta}\tilde{\gamma}'-\frac{\tilde{\beta}^2\tilde{\gamma}'\tilde{g}'}{\tilde{g}}\right]-\frac{12\tilde{\beta}^2}{n}\tilde{g}. \quad (54)$$

Now, we are in a position to find Equation (49). Recall that $\kappa=(\beta^2+\lambda)g=\gamma+\lambda g$. Then

$$\kappa'=\gamma'+\lambda g'=(\beta^2+\lambda)g'+2\beta g.$$

Then, summing up all terms in Lemma 3.1 gives us

$$-S\tilde{\mathcal{F}}_{\beta\beta}\sum_{i,j=1}^{n}\zeta_{i,j}=2\left(1-\frac{2}{n}\right)\left[(\tilde{\beta}^2+\lambda)\tilde{\kappa}''+\left(2\tilde{\beta}-\frac{(\tilde{\beta}^2+\lambda)\tilde{g}'}{\tilde{g}}\right)\tilde{\kappa}'\right]$$
$$-\frac{4(3\tilde{\beta}^2+\lambda)\tilde{g}}{n}+\mathcal{O}(\sigma^6). \quad (55)$$

Consequently, up to order $\mathcal{O}(\sigma^6)$, the expectation

$$\mathbb{E}(\Delta_3\hat{\beta}\Delta_1\hat{\beta})=-\frac{\sigma^4}{4\tilde{g}nS^2}\left\{2\left(1-\frac{2}{n}\right)\left[(\tilde{\beta}^2+\lambda)\tilde{\kappa}''+\left(2\tilde{\beta}-\frac{(\tilde{\beta}^2+\lambda)\tilde{g}'}{\tilde{g}}\right)\tilde{\kappa}'\right]\right.$$
$$\left.-\frac{4(3\tilde{\beta}^2+\lambda)\tilde{g}}{n}\right\}. \quad (56)$$

Finally, substitute Equation (48) and Equation (56) with Equation (38) to get the general form of the MSE for any estimator that minimizes (4).

$$\mathrm{MSE}(\hat{\beta})=\frac{(\tilde{\beta}^2+\lambda)\sigma^2}{nS}+\frac{\sigma^4}{nS^2}\left(\lambda+\frac{2\tilde{\beta}^2-\lambda}{n}\right)+\frac{\sigma^4}{4\tilde{g}^2nS^2}$$
$$\times\left[(\tilde{\tau}'-2\tilde{\beta}\tilde{g})\tilde{\kappa}'+4\left(1-\frac{2}{n}\right)\left(-(\tilde{\beta}^2+\lambda)\tilde{g}\tilde{\kappa}''-\left(2\tilde{\beta}\tilde{g}-(\tilde{\beta}^2+\lambda)\tilde{g}'\right)\tilde{\kappa}'\right)\right.$$
$$\left.+\frac{8\tilde{\beta}\tilde{g}\tilde{\kappa}'+4(7\tilde{\beta}^2+2\lambda)\tilde{g}^2}{n}\right]+\mathcal{O}(\sigma^6).$$

In an attempt to compare the new estimator $\hat{\beta}_2$ and the MLE $\hat{\beta}_1$ and $\hat{\beta}_0$, we use the derived formulas to decompose their $O(\sigma^4)$ MSEs.

As shown in Table 1, *while* $\mathbb{E}(\Delta_1\hat{\beta}_1\Delta_3\hat{\beta}_1)$ *is always positive,* $\mathbb{E}(\Delta_1\hat{\beta}_2\Delta_3\hat{\beta}_2)$ *turns out to be exactly zero*. This means that the MSE of $\hat{\beta}_2$ depends only on the MSEs of the first- and the second-order errors (i.e., $\Delta_1\hat{\beta}_2$ and $\Delta_2\hat{\beta}_2$). Based on this observation and Remark 3.1, we conclude that $\hat{\beta}_2$ is superior to $\hat{\beta}_1$ for all values of β and $n>4$. Consequently, since the weight function $g_2(\beta)$ is the solution for the ODE $\tau'=0$ (up to scalar), one would think of looking for another weight function, say g_3, which also solves $\tau'=0$ such that its corresponding estimator, say $\hat{\beta}_3$, minimizes $\mathcal{F}_3=g_3(\beta)\sum_i(y_i^*-\beta x_i^*)^2$ and satisfies $\mathbb{E}(\Delta_1\hat{\beta}_3\Delta_3\hat{\beta}_3)=0$ while $\mathrm{MSE}(\Delta_2\hat{\beta}_3)$ attains its minimal value.

The weight function g_3 should incorporate α and $g_2(\beta)$ together. Let us define $g_3(\alpha,\beta)=c(\alpha)g_2(\beta)$. Then, it is easy to show that g_3 is also a solution for $\tau'=0$ (see Equation (47)). Substituting $g_3(\alpha,\beta)$ in Equation (48) gives the same expression g_2. This proves that $c(\alpha)$ is irrelevant and can be set to 1, and as such, $\hat{\beta}_3=\hat{\beta}_2$. Hence no possible further reduction in

Table 1. Mean squared error (and its components) for estimators: least-squares estimator $\hat{\beta}_0$, the MLE $\hat{\beta}_1$, and the new proposed estimator $\hat{\beta}_2$.

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

MSE. Accordingly, we conclude that

$$\mathrm{MSE}(\hat{\beta}_2)=\mathrm{Var}(\Delta_1\hat{\beta}_2)+\mathrm{Var}(\Delta_2\hat{\beta}_2)+\mathcal{O}(\sigma^6)$$

is smaller than the MSE of any other estimator of β. This shows that $\hat{\beta}_2$ is the best estimator among all other estimators: it has zero bias up to order σ^4/n^2 and has minimal variance, i.e.,

$$\mathrm{MSE}(\hat{\beta}_2)=\frac{(\lambda+\tilde{\beta}^2)\sigma^2}{\|\tilde{\mathbf{x}}^*\|^2}+\frac{\sigma^4}{nS^2}\left(\lambda+\frac{2\tilde{\beta}^2-\lambda}{n}\right)+\frac{2\sigma^4\tilde{\beta}^2}{n^2(n-2)S^2}.$$

3.1. Higher-order bias

The numerical experiments (presented shortly) confirm that $\hat{\beta}_2$ outperforms not only the MLE, but also the AMLE, $\check{\beta}_1$, which was obtained using the post-bias elimination technique. Then, one might wonder what happened to the higher-order terms of the bias for these estimators. Here, we will derive a general expression for the fourth-order bias with a hope to get another estimator that outperforms all other estimators, including $\hat{\beta}_2$, by applying post-bias elimination step to the fourth-order bias. However, including all terms of order σ^4 might be a difficult task; hence terms of order of magnitude σ^4 will be considered here while terms of order σ^4/n, or less, will be discarded.

With the aid of the Taylor expansion, $\hat{\beta}$ can be expressed as $\hat{\beta}=\tilde{\beta}+\Delta_1\hat{\beta}+\Delta_2\hat{\beta}+\Delta_3\hat{\beta}+\Delta_4\hat{\beta}+\mathcal{O}(\sigma^5)$. Therefore,

$$\mathrm{Bias}(\hat{\beta})=\mathbb{E}(\Delta_2\hat{\beta})+\mathbb{E}(\Delta_4\hat{\beta})+\cdots.$$

The first term is given in Equation (31) so we only need to find $\mathbb{E}(\Delta_4\hat{\beta})$. After long derivations (see the Appendix), we get

$$\mathbb{E}(\Delta_4\hat{\beta})=\frac{\sigma^4\tilde{\kappa}'}{4\tilde{g}^2S^2}\left[2\tilde{\kappa}''-\frac{3\tilde{g}'\tilde{\kappa}'}{\tilde{g}}\right]+\mathcal{O}(\sigma^4/n), \tag{57}$$

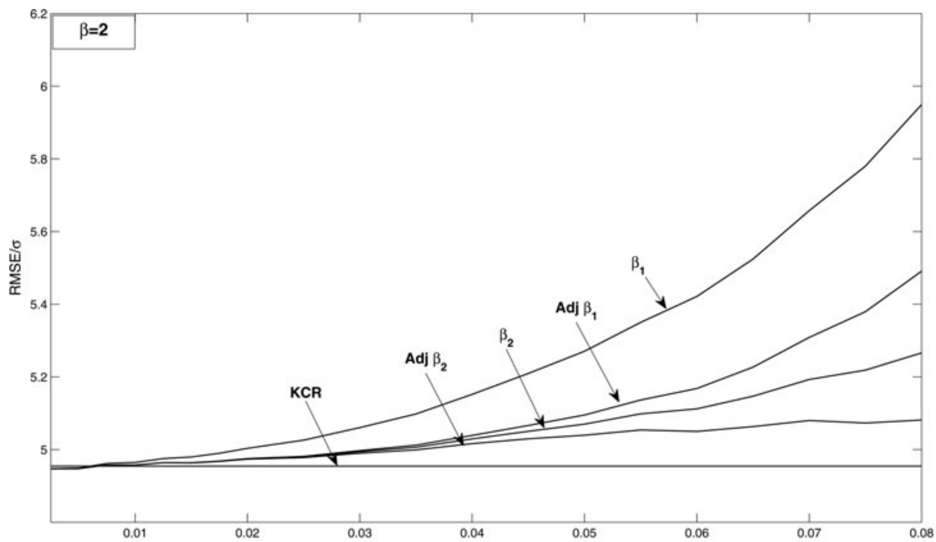
where $S=\|\tilde{\mathbf{x}}^*\|^2/n$. We call this part of the bias "*the fourth-order essential bias*." For the MLE $\hat{\beta}_1$, $\tilde{\kappa}'=0$, and this term is zero. Also, $\hat{\beta}_2$ has zero fourth-order essential bias. In fact, after simple algebra, one gets

$$\mathbb{E}(\Delta_4\hat{\beta}_2)=\frac{2\left(\lambda(n-2)+(2n-5)\tilde{\beta}^2\right)\tilde{\beta}}{S^2(n-2)^3(\tilde{\beta}^2+\lambda)}\sigma^4+\mathcal{O}(\sigma^4/n^2). \tag{58}$$

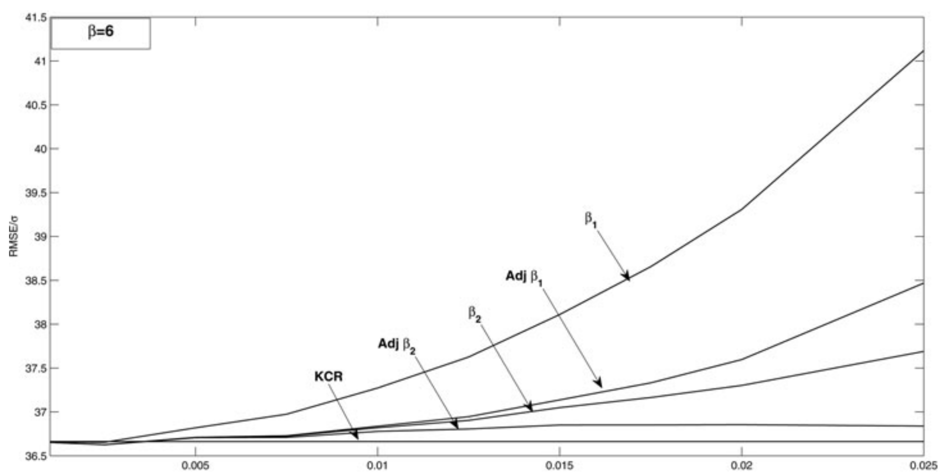
It is interesting to determine a general form of $g(\beta)$ such that the estimator that minimizes $\mathcal{F}(\beta)$ also has zero fourth-order essential bias. This exactly represents solving the ODE $\frac{d(\kappa'(\beta))^2}{d\beta} = \frac{3g'(\kappa'(\beta))^2}{g(\beta)}$, or equivalently, $\frac{d(\kappa'(\beta))^2}{(\kappa'(\beta))^2} = \frac{3g'(\beta)d\beta}{g}$. Thus, $\kappa'(\beta) = C[g(\beta)]^{3/2}$ for some constant C. This is exactly a Bernoulli linear differential equation, and as such, its solution is

$$g(\beta) = \frac{4\lambda^2}{C\left(C\beta^2 - 4D\lambda\beta\sqrt{\beta^2+\lambda}\right) + 4\lambda^2 D^2(\beta^2+\lambda)}, \tag{59}$$

where C and D are constants. Note that $g_1(\beta) = (\beta^2+\lambda)^{-1}$ is a special case of Equation (59) that appears whenever $C = 0$ and $D = 1$, and it is the only member of this family associated



(a) $\tilde{\beta} = 2,\ n = 10,\ \lambda = 1.$



(b) $\tilde{\beta} = 6,\ n = 10,\ \lambda = 1.$

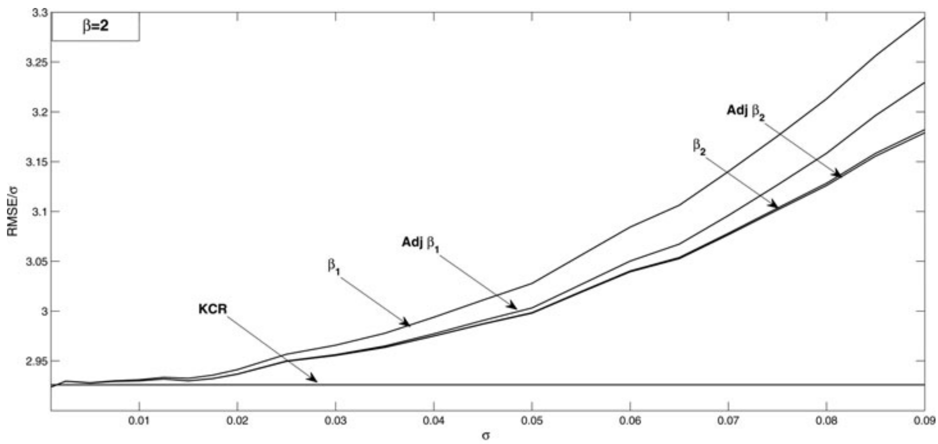
Figure 1. NRMSE$(\hat{\beta})$ for MLE, AMLE (Adj. β_1), and the proposed estimators $\hat{\beta}_2$ and Adj. β_2 versus σ. The horizontal line represents the KCR divided by σ.

with an estimator that has a very nice property: *zero essential bias of the second and fourth orders*.

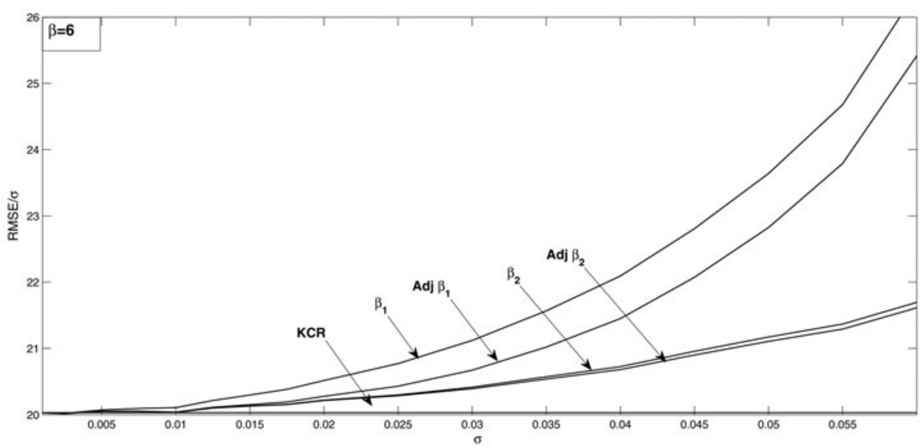
Adjusted $\hat{\beta}_2$. Although $\hat{\beta}_2$ has zero $\mathcal{O}(\sigma^2)$ bias, it has a non zero bias of order of magnitude σ^4/n^2. Therefore, this estimator could be adjusted by subtracting the unbiased estimator from its $\mathcal{O}(\sigma^4/n^2)$ bias. Then, the new estimator will have zero bias up to order σ^6. Its formula depends on $\|\tilde{\mathbf{x}}^*\|^2 = S/n$ and σ^4, so if replaced by $\|\mathbf{x}^*\|^2$ and $\hat{\sigma}_2^4 = (n-2)^{-2}[\mathcal{F}_2(\hat{\beta}_2)]^2$, then it is easy to verify that the new estimator

$$\check{\beta}_2 = \left[1 - \frac{2n^2\left(n-2+(2n-5)\hat{\beta}_2^2\right)\hat{\sigma}_2^4}{\|\mathbf{x}^*\|^4(n-2)^3(1+\hat{\beta}_2^2)}\right]\hat{\beta}_2 \tag{60}$$

is an unbiased estimator of β, up to order σ^6. This adjustment improves the accuracy of the estimator for small n.



(a) $\tilde{\beta} = 2$, $n = 40$, $\lambda = 2$.



(b) $\tilde{\beta} = 6$, $n = 40$, $\lambda = 2$.

Figure 2. NRMSE($\hat{\beta}$) for MLE, AMLE (Adj. β_1), and the proposed estimators $\hat{\beta}_2$ and Adj. β_2 versus σ. The horizontal line represents the KCR divided by σ.

4. Numerical experiments and conclusion

To demonstrate our findings, we turn our attention to some numerical experiments. In our experiments, we considered the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ and their adjustments: AMLE ($\check{\beta}_1$) and $\check{\beta}_2$. Since the MLE is invariant under translations, the value of α is irrelevant. The sample size n was set to 10 (see Figure 1) and 40 (see Figure 2). For β, we tested two values: $\beta = 2$ and $\beta = 6$. For each case, we positioned n-equally spaced true points on the line (spanning an interval of length $L = 1$; i.e., the distance between the first and last true point was one). For a given noise level σ, $N = 10^6$ samples of size n were simulated, and then the *normalized-root-mean-squared-error* (NRMSE) of $\hat{\beta}$ was estimated for each fit by

$$\text{NRMSE}(\hat{\beta}) = \sqrt{\sum_{i=1}^{N} \frac{(\hat{\beta} - \tilde{\beta})^2}{N\sigma^2}}. \tag{61}$$

We plotted the NRMSE of $\hat{\beta}$ against the noise level σ. The noise level σ was varied from 0 up to the point, at which the values of NRMSE($\hat{\beta}$) became large.

As a general observation, as σ is approaching 0, the NRMSE approaches the KCR (represented by the horizontal line). However, the higher-order terms of the MSE play a key role in the performances of the estimators as σ increases. This numerically demonstrates our theoretical results. As seen in Figures 1 and 2, $\hat{\beta}_2$ and its adjustment outperform the AMLE $\check{\beta}_1$, which performs better than the MLE for each value of σ, n, and β. The adjustment of $\hat{\beta}_2$ always outperforms $\hat{\beta}_2$, especially whenever n is relatively small. For relatively large sample size, the improvement gained by this adjustment is small as being of order σ^4/n^2.

Moreover, the proposed estimators are more robust than the MLE and its adjustment. Some other numerical experiments show that when the noise level σ becomes large, the MLE and its adjustment start returning very unsatisfactory results while the proposed estimators return very reasonable estimates.

5. Conclusion

The first contribution in this article is applying the error analysis to a general class of objective functions, of which the MLE of the slope is a special case. Accordingly, we proposed the first novel estimator, $\hat{\beta}_2$, which minimizes the objective function $\mathcal{F}_2 = g_2(\beta)(s_{yy} - 2\beta s_{xy} + s_{xx}\beta^2)$. The weight function $g_2(\beta)$ was chosen such that the minimum of its associated objective function $\mathcal{F}_2(\beta)$ has zero bias (up to order σ^4). This estimator came as a result of using the pre-bias elimination technique that developed here. The solution was computed by solving the cubic equation (36). Moreover, another estimator was proposed by applying the post-bias elimination technique to the MLE $\hat{\beta}_1$. The idea in the post-bias elimination technique is simple. We subtracted the unbiased estimator of its second-order bias from $\hat{\beta}_1$. We called this estimator AMLE.

We also derived general formulas for the higher-order bias and the MSE for the slope's estimators up to orders σ^4 and σ^4/n^2, respectively. Consequently, we applied the post-bias elimination technique to the proposed estimator $\hat{\beta}_2$, and as such, another unbiased estimator up (to order σ^6) is obtained. We called this new estimator adjusted $\hat{\beta}_2$ and we denoted it by $\check{\beta}_2$. We also compared three estimators (least-squares fit $\hat{\beta}_0$, MLE ($\hat{\beta}_1$), and $\hat{\beta}_2$) based on the bias and the MSE criteria. We showed why $\hat{\beta}_2$ outperforms $\hat{\beta}_1$, which also outperforms $\hat{\beta}_0$. We validated our findings through a series of numerical experiments where four estimators were tested: MLE $\hat{\beta}_1$, AMLE $\check{\beta}_1$, $\hat{\beta}_2$, and its adjustment $\check{\beta}_2$. Our proposed estimators are superior to other estimators for any sample size, especially when the sample size is small.

Acknowledgments

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

References

Adcock, R.J. (1877). Note on the method of least squares. *Anal. London* 4:183–184.
Adcock, R.J. (1878). A problem in least squares. *Anal. London* 5:53–54.
Al-Sharadqah, A., Chernov, N. (2011). Statistical analysis of curve fitting methods in errors-in-variables models. *Theory Probab. Math. Stat.* 84:4–17.
Amiri-Simkooeii, A., Jazaeri, S. (2012). Weighted total least squares formulated by standard least squares theory. *J. Geodetic Sci.* 2(2):113–124.
Anderson, T.W. (1976). Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics. *J. R. Stat. Soc. B* 38:1–36.
Anderson, T.W. (1984). Estimating linear statistical relationships. *Ann. Stat.* 12:1–45.
Anderson, T.W., Sawa, T. (1982). Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient. *J. R. Stat. Soc. B* 44:52–62.
Cheng, C.-L., Van Ness, J.W. (1999). *Statistical Regression with Measurement Error*. London: Arnold.
Chernov, N. (2010). *Circular and Linear Regression: Fitting Circles and Lines by Least Squares*, volume 117. Boca Raton: Chapman & Hall/CRC.
Chernov, N., Lesort, C. (2004). Statistical efficiency of curve fitting algorithms. *Comp. Stat. Data Anal.* 47:713–728.
Deming, W.E. (1943). *Statistical Adjustment of Data*. New York, NY: Wiley.
Fuller, W.A. (1987). *Measurement Error Models*. New York: L. Wiley & Son.
Gillard, J. (2006). An historical overview of linear regression with errors in both variables. Technical report.
Jung, K. (2007). Least trimmed squares estimator in the errors-in-variables model. *J. Appl. Stat.* 34(3):331–338.
Kanatani, K. (1993). *Geometric Computation for Machine Vision*. Oxford: Oxford University Press.
Kanatani, K. (1996). *Statistical Optimization for Geometric Computation: Theory and Practice*. Amsterdam, Netherlands: Elsevier Science.
Kanatani, K. (1998). Cramer-Rao lower bounds for curve fitting. *Graph. Mod. Image Process.* 60:93–99.
Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philos. Mag.* 2:559–572.
van Huffel, S. Ed. (2002). *Total Least Squares and Errors-in-Variables Modeling*. Dordrecht: Kluwer.
Soderstrom, T. (2007). Errors-in-variables methods in system identification. *Automatica* 43(6):939–958.

Appendix

Proof of Lemma 2.1 Differentiating Equation (15) with respect to β, x_i, and y_i gives

$$\mathcal{F}_{\beta\beta\beta} = g_{\beta\beta\beta}\sum d_i^{*2} - 6g_{\beta\beta}\sum d_i^* x_i^* + 6g_\beta \sum [x_i^*]^2, \tag{A.1}$$

$$\mathcal{F}_{\beta\beta x_i} = -2\beta g_{\beta\beta} d_i^* - 4g_\beta y_i^* + 8\beta g_\beta x_i^* + 4g x_i^*, \tag{A.2}$$

$$\mathcal{F}_{\beta\beta y_i} = 2g_{\beta\beta} d_i^* - 4g_\beta x_i^*. \tag{A.3}$$

Similarly, one can obtain

$$\mathcal{F}_{\beta x_i x_j} = \left[2\beta^2 g_\beta + 4\beta g\right]\hat{\delta}_{ij}, \tag{A.4}$$

$$\mathcal{F}_{\beta x_i y_j} = -\left[2\beta g_\beta + 2g\right]\hat{\delta}_{ij}, \tag{A.5}$$

and

$$\mathcal{F}_{\beta y_i y_j} = 2g_\beta \hat{\delta}_{ij}. \tag{A.6}$$

If these derivatives are evaluated at the true values $\tilde{\beta}$ and $(\tilde{x}_i, \tilde{y}_i)$ for each $i = 1, \ldots, n$, Lemma 2.1 will be established. □

Proof of Lemma 3.1 Before proceeding, we present the following lemma that gives the fourth partial derivatives of $\mathcal{F}$ with respect to x, y, and β evaluated at the true data $(\tilde{\mathbf{x}}_i, \tilde{\mathbf{y}}_i)'$s. □

Lemma A.1. *Define $\hat{\delta}_{ij} = \delta_{ij} - \frac{1}{n}$ and $\gamma(\beta) = \beta^2 g(\beta)$. At the true values $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$, we have the following:*

$$\begin{aligned} &\tilde{\mathcal{F}}_{\beta\beta\beta\beta} = 12\tilde{g}'' nS, \quad \tilde{\mathcal{F}}_{\beta\beta\beta x_i} = 6(\beta\tilde{g}'' + 2\tilde{g}')\tilde{x}_i^*, \quad \tilde{\mathcal{F}}_{\beta\beta\beta y_i} = -6\tilde{g}''\tilde{x}_i^*, \\ &\tilde{\mathcal{F}}_{\beta\beta x_i x_j} = 2\tilde{\gamma}''\hat{\delta}_{ij} \quad \tilde{\mathcal{F}}_{\beta\beta y_i y_j} = 2\tilde{g}''\hat{\delta}_{ij}, \quad \tilde{\mathcal{F}}_{\beta\beta x_i y_j} = -2(\tilde{\beta}\tilde{g}'' + 2\tilde{g}')\hat{\delta}_{ij}, \end{aligned} \tag{A.7}$$

where $\tilde{\gamma}'(\tilde{\beta}) = \tilde{\beta}^2\tilde{g}' + 2\tilde{\beta}\tilde{g}$ and $\tilde{\gamma}''(\tilde{\beta}) = \tilde{\beta}^2\tilde{g}'' + 4\tilde{\beta}\tilde{g}' + 2\tilde{g}$. Besides, for all i, j, k:

$$\tilde{\mathcal{F}}_{\beta x_i x_j x_k} = \tilde{\mathcal{F}}_{\beta x_i x_j y_k} = \tilde{\mathcal{F}}_{\beta y_i y_j y_k} = \tilde{\mathcal{F}}_{\beta x_i y_j y_k} = 0.$$

Proof. First, we differentiate Equation (A.1) with respect to β, x_i, y_i, respectively, to get

$$\mathcal{F}_{\beta\beta\beta\beta} = g_{\beta\beta\beta\beta}\sum d_i^{*2} - 8g_{\beta\beta\beta}\sum d_i^* x_i^* + 12g_{\beta\beta}\sum [x_i^*]^2, \tag{A.8}$$

$$\mathcal{F}_{\beta\beta\beta x_i} = -2\beta g_{\beta\beta\beta} d_i^* - 6g_{\beta\beta}(y_i^* - 2\beta x_i^*) + 12g_\beta x_i^*, \tag{A.9}$$

$$\mathcal{F}_{\beta\beta\beta y_i} = 2g_{\beta\beta\beta}(y_i^* - \beta x_i^*) - 6g_{\beta\beta}x_i^*. \tag{A.10}$$

Note that we used the fact $\sum_{i=1}^n x_i^* = \sum_{i=1}^n y_i^* = 0$. To get $\mathcal{F}_{\beta\beta x_i x_j}$ and $\mathcal{F}_{\beta\beta x_i y_j}$, we differentiate Equation (A.2)

$$\mathcal{F}_{\beta\beta x_i x_j} = 2(\beta^2 g_{\beta\beta} + 4\beta g_\beta + 2g)\hat{\delta}_{ij}, \tag{A.11}$$

$$\mathcal{F}_{\beta\beta x_i y_j} = -2(\beta g_{\beta\beta} + 2g_\beta)\hat{\delta}_{ij}. \tag{A.12}$$

Finally, differentiating $\mathcal{F}_{\beta y_i y_j}$ in Equation (A.6) with respect to β gives us

$$\mathcal{F}_{\beta\beta y_i y_j} = 2g_{\beta\beta}\hat{\delta}_{ij}. \tag{A.13}$$

If the true values of all (x_i, y_i)'s are substituted into Equations (A.8)–(A.13), the results will be established. Finally, both derivatives of $\mathcal{F}_{\beta x_i x_j}$ with respect to x_k and y_k equal zero, i.e., $\mathcal{F}_{\beta x_i x_j x_k} = 0$, $\mathcal{F}_{\beta x_i x_j y_k} = 0$. In the same analog, the other derivatives, such as $\mathcal{F}_{\beta y_i y_j y_k}$ and $\mathcal{F}_{\beta x_i y_j x_k}$, are all zeroes. This completes the proof of the lemma. □

Now, let us denote the total derivatives of $\mathcal{F}_{\beta\beta}$ and $\mathcal{F}_{\beta\beta\beta}$ with respect to x_i by

$$\rho_{x_i}^{(1)} = \mathcal{F}_{\beta\beta x_i} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}_{x_i} \quad \text{and } \rho_{x_i}^{(2)} = \mathcal{F}_{\beta\beta\beta x_i} + \mathcal{F}_{\beta\beta\beta\beta}\hat{\beta}_{x_i}, \tag{A.14}$$

respectively. At the true values, $\rho_{x_i}^{(1)}$ and $\rho_{x_i}^{(2)}$ reduce to

$$\tilde{\rho}_{x_i}^{(1)} = (-2\tilde{\beta}\tilde{g}' + 4\tilde{g})\tilde{x}_i^* \quad \text{and } \tilde{\rho}_{x_i}^{(2)} = 6(-\tilde{\beta}\tilde{g}'' + 2\tilde{g}')\tilde{x}_i^*, \tag{A.15}$$

respectively. Similarly, denote by $\xi_{y_i}^{(1)}$ and $\xi_{y_i}^{(2)}$ the total derivatives of $\mathcal{F}_{\beta\beta}$ and $\mathcal{F}_{\beta\beta\beta}$ (respectively) with respect to y_i. That is,

$$\xi_{y_i}^{(1)} = \mathcal{F}_{\beta\beta y_i} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}_{y_i} \quad \text{and} \quad \xi_{y_i}^{(2)} = \mathcal{F}_{\beta\beta\beta y_i} + \mathcal{F}_{\beta\beta\beta\beta}\hat{\beta}_{y_i}. \tag{A.16}$$

At the true values, they take the form

$$\tilde{\xi}^{(1)}_{y_i} = 2\tilde{g}'\tilde{x}^*_i \quad \text{and } \tilde{\xi}^{(2)}_{y_i} = 6\tilde{g}''\tilde{x}^*_i. \tag{A.17}$$

These quantities play a key role in the sequel analysis. Substitute $j = i$ in Equation (21) to get

$$\mathcal{F}_{\beta\beta}\beta_{x_ix_i} + \beta^2_{x_i}\mathcal{F}_{\beta\beta\beta} + 2\beta_{x_i}\mathcal{F}_{\beta\beta x_i} + \mathcal{F}_{\beta x_ix_i} = 0. \tag{A.18}$$

Differentiating Equation (A.18) with respect to x_j gives us

$$\mathcal{F}_{\beta\beta}\hat{\beta}_{x_ix_ix_j} + \rho^{(1)}_{x_j}\hat{\beta}_{x_ix_i} + \rho^{(2)}_{x_j}\hat{\beta}^2_{x_i} + 2\rho^{(1)}_{x_i}\hat{\beta}_{x_ix_j} + 2(\mathcal{F}_{\beta\beta x_ix_j} + \mathcal{F}_{\beta\beta x_i\beta}\beta_{x_j})\hat{\beta}_{x_i} + \mathcal{F}_{\beta x_ix_i\beta}\hat{\beta}_{x_j} = 0, \tag{A.19}$$

from which $\beta_{x_ix_ix_j}$ can be evaluated. If we solve Equation (A.19) for $\beta_{x_ix_ix_j}$ after we evaluated each term at the true points, then we will get $\beta_{x_ix_ix_j} \sim \mathcal{O}(n^{-2})$. Since $\beta_{x_j} = \frac{-\tilde{\beta}\tilde{x}^*_j}{nS} \sim \mathcal{O}(n^{-1})$, $\sum_i \beta_{x_ix_ix_i}\beta_{x_i} \sim \mathcal{O}(n^{-2})$. Therefore, all terms in $\sum_{i,j}\beta_{x_ix_ix_j}\beta_{x_j}$ with order $\mathcal{O}(n^{-1})$ are

$$\sum_{i\neq j}^{n}\beta_{x_ix_ix_j}\beta_{x_j} = \frac{-1}{\tilde{\mathcal{F}}_{\beta\beta}}\sum_{i\neq j}^{n}(\tilde{\rho}^{(1)}_{x_j}\beta_{x_ix_i} + \tilde{\mathcal{F}}_{\beta x_ix_i\beta}\beta_{x_j})\beta_{x_j} + \mathcal{O}(n^{-2}).$$

From Equation (A.15), $\tilde{\rho}^{(1)}_{x_j} = (-2\tilde{\beta}\tilde{g}' + 4\tilde{g})\tilde{x}^*_j$, then up to the leading term, one gets

$$\sum_{i\neq j}^{n}\beta_{x_ix_ix_j}\beta_{x_j} = \frac{-2\tilde{\beta}}{n^2S^2\tilde{\mathcal{F}}_{\beta\beta}}\sum_{i\neq j}^{n}\left((-\tilde{\beta}\tilde{g}' + 2\tilde{g})\frac{\tilde{\gamma}'}{\tilde{g}} + \tilde{\beta}\tilde{\gamma}''\right)[\tilde{x}^*_j]^2,$$

and further,

$$\sum_{i\neq j}^{n}\beta_{x_ix_ix_j}\beta_{x_j} = \frac{-2n(n-1)S\tilde{\beta}}{n^2S^2\tilde{\mathcal{F}}_{\beta\beta}}\left((-\tilde{\beta}\tilde{g}' + 2\tilde{g})\frac{\tilde{\gamma}'}{\tilde{g}} + \tilde{\beta}\tilde{\gamma}''\right) + \mathcal{O}(n^{-2}).$$

Here, we used $\beta_{x_ix_i} = -\frac{\tilde{\gamma}'}{\tilde{g}nS} + \mathcal{O}(n^{-2})$ and $\sum_{i=1}^{n}[\tilde{x}^*_j]^2 = nS$. Moreover, since $\sum_{i\neq j}[\tilde{x}^*_i]^2 = n(n-1)S$ and $\tilde{\mathcal{F}}_{\beta\beta} = 2\tilde{g}nS$, one has

$$\sum_{i\neq j}^{n}\beta_{x_ix_ix_j}\beta_{x_j} = -\frac{\tilde{\beta}^2\tilde{\gamma}'' + \tilde{\beta}(-\tilde{\beta}\tilde{g}' + 2\tilde{g})\frac{\tilde{\gamma}'}{\tilde{g}}}{\tilde{g}nS^2} + \mathcal{O}(n^{-2}).$$

Next, we find $\sum_{i\neq j}^{n}\beta_{y_iy_iy_j}\beta_{y_j}$. Substitute $j = i$ in Equation (23) to get

$$\mathcal{F}_{\beta\beta}\hat{\beta}_{y_iy_i} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}^2_{y_i} + 2\mathcal{F}_{\beta\beta y_i}\hat{\beta}_{y_i} + \mathcal{F}_{\beta y_iy_i} = 0. \tag{A.20}$$

Differentiate (A.20) with respect to y_j to get

$$\mathcal{F}_{\beta\beta}\hat{\beta}_{y_iy_iy_j} + \xi^{(1)}_{y_j}\hat{\beta}_{y_iy_i} + \xi^{(2)}_{y_j}\hat{\beta}^2_{y_i} + 2\xi^{(1)}_{y_i}\hat{\beta}_{y_iy_j} + 2\hat{\beta}_{y_i}(\mathcal{F}_{\beta\beta\beta y_i}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta y_iy_j}) + \mathcal{F}_{\beta\beta y_iy_i}\hat{\beta}_{y_j} = 0. \tag{A.21}$$

The desired results will be obtained if we follow the same procedures implemented above. This gives us

$$\sum_{i\neq j}^{n}\beta_{y_iy_iy_j}\beta_{y_j} = \frac{-1}{\tilde{\mathcal{F}}_{\beta\beta}}\sum_{i\neq j}^{n}(\tilde{\xi}^{(1)}_{y_j}\beta_{y_iy_i} + \tilde{\mathcal{F}}_{\beta y_iy_i\beta}\beta_{y_j})\beta_{y_j} + \mathcal{O}(n^{-2}).$$

From Equations (A.17) and (30), we obtain $\tilde{\xi}^{(1)}_{y_j} = 2\tilde{g}'\tilde{x}^*_j$ and $\beta_{y_iy_i} = \frac{-\tilde{g}'}{\tilde{g}nS} + \mathcal{O}(n^{-2})$. Since $\beta_{y_j} = \frac{\tilde{x}^*_j}{nS}$, then

$$\sum_{i\neq j}^{n} \beta_{y_iy_iy_j}\beta_{y_j} = \frac{-2n(n-1)S(\tilde{g}'' - (\tilde{g}')^2/\tilde{g})}{n^2S^2\tilde{\mathcal{F}}_{\beta\beta}} + \mathcal{O}(n^{-2}),$$

and, as such,

$$\sum_{i\neq j}^{n} \beta_{y_iy_iy_j}\beta_{y_j} = -\frac{\tilde{g}'' - (\tilde{g}')^2/\tilde{g}}{\tilde{g}nS^2} + \mathcal{O}(n^{-2}).$$

Next we derive (53), by first differentiating (A.18) with respect to y_j in order to get $\beta_{x_ix_iy_j}$. That is,

$$\mathcal{F}_{\beta\beta}\hat{\beta}_{x_ix_iy_j} + \xi^{(1)}_{y_j}\hat{\beta}_{x_ix_i} + 2\rho^{(1)}_{x_i}\hat{\beta}_{x_iy_j} + \hat{\beta}^2_{x_i}\xi^{(2)}_{y_j} + 2\hat{\beta}_{x_i}(\mathcal{F}_{\beta\beta\beta x_i}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta x_iy_j}) + \mathcal{F}_{\beta\beta x_ix_i}\hat{\beta}_{y_j} = 0, \tag{A.22}$$

from which we obtain

$$\sum_{i\neq j} \beta_{x_ix_iy_j}\beta_{y_j} = \frac{-1}{\tilde{\mathcal{F}}_{\beta\beta}}\sum_{i\neq j}^{n}\left(\tilde{\xi}^{(1)}_{y_j}\beta_{x_ix_i} + \tilde{\mathcal{F}}_{\beta x_ix_i\beta}\beta_{y_j}\right)\beta_{y_j} = \frac{-2n(n-1)S(\tilde{\gamma}'' - \tilde{\gamma}'\tilde{g}'/\tilde{g})}{n^2S^2\tilde{\mathcal{F}}_{\beta\beta}} + \mathcal{O}(n^{-2})$$

and further

$$\sum_{i\neq j} \beta_{x_ix_iy_j}\beta_{y_j} = -\frac{\tilde{\gamma}'' - \tilde{\gamma}'\tilde{g}'/\tilde{g}}{\tilde{g}nS^2} + \mathcal{O}(n^{-2}).$$

Finally, to get Equation (54), we differentiate (22) with respect to y_j to get

$$\begin{aligned}&\mathcal{F}_{\beta\beta}\hat{\beta}_{x_iy_jy_j} + 2\xi^{(1)}_{y_j}\hat{\beta}_{x_iy_j} + \rho^{(1)}_{x_i}\hat{\beta}_{y_jy_j} + \xi^{(2)}_{y_j}\hat{\beta}_{x_i}\hat{\beta}_{y_j} + (\mathcal{F}_{\beta\beta x_iy_j} + \mathcal{F}_{\beta\beta\beta x_i}\hat{\beta}_{y_j})\hat{\beta}_{y_j}\\ &\quad + (\mathcal{F}_{\beta\beta y_jy_j} + \mathcal{F}_{\beta\beta\beta y_j}\hat{\beta}_{y_j})\hat{\beta}_{x_i} + \mathcal{F}_{\beta\beta x_iy_j}\hat{\beta}_{y_j} = 0.\end{aligned} \tag{A.23}$$

With the aid of this equation, we get

$$\sum_{i\neq j}^{n} \beta_{x_iy_jy_j}\beta_{x_i} = \frac{-1}{\tilde{\mathcal{F}}_{\beta\beta}}\sum_{i\neq j}^{n}(\tilde{\rho}^{(1)}_{x_i}\beta_{y_jy_j} + \tilde{\mathcal{F}}_{\beta\beta y_jy_j}\beta_{x_i})\beta_{x_i} = -\frac{\tilde{\beta}^2\tilde{g}'' + \tilde{\beta}\tilde{g}'(2\tilde{g} - \tilde{\beta}\tilde{g}')/\tilde{g}}{\tilde{g}nS^2} + \mathcal{O}(n^{-2}).$$

This completes the derivation of the most important terms in each of the expressions given in the lemma. □

Derivation of Equation (57). First, we present the following lemma without proof. This lemma summarizes the fifth derivatives of $\mathcal{F}$.

Lemma A.2. *Define* $\hat{\delta}_{ij} = \delta_{ij} - \frac{1}{n}$ *and* $\gamma(\beta) = \beta^2 g(\beta)$. *At the true values* $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \dots, n$, *we have the following:*

$$\tilde{\mathcal{F}}_{\beta\beta\beta\beta\beta} = 20\tilde{g}^{(3)}nS, \quad \tilde{\mathcal{F}}_{\beta\beta\beta\beta y_i} = -8\tilde{g}^{(3)}\tilde{x}^*_i, \quad \tilde{\mathcal{F}}_{\beta\beta\beta\beta x_i} = 8\tilde{g}^{(3)}\tilde{x}^*_i + 24\tilde{g}''\tilde{x}^*_i,$$
$$\tilde{\mathcal{F}}_{\beta\beta\beta x_ixj} = 2\tilde{\gamma}^{(3)}\hat{\delta}_{ij}, \quad \tilde{\mathcal{F}}_{\beta\beta\beta x_iyj} = (-2\tilde{\beta}\tilde{g}^{(3)} - 6\tilde{g}'')\hat{\delta}_{ij}, \quad \tilde{\mathcal{F}}_{\beta\beta\beta y_iyj} = 2\tilde{g}^{(3)}\hat{\delta}_{ij}.$$

Now, since δ_i and ε_i are normal random variables with mean 0 and variance σ^2 for each $i = 1, \dots, n$, the expected values of $\delta_i\delta_j\delta_k\varepsilon_l$ and $\varepsilon_i\varepsilon_j\varepsilon_k\delta_l$ are zero. Hence,

$$\begin{aligned}
&\mathbb{E}(\Delta_4\hat{\beta})\\
&= \tfrac{1}{24}\left[\sum_{i,j,k,l}\beta_{x_ix_jx_kx_l}\mathbb{E}(\delta_i\delta_j\delta_k\delta_l) + 6\beta_{x_ix_jy_ky_l}\mathbb{E}(\delta_i\delta_j\varepsilon_k\varepsilon_l) + \beta_{y_iy_jy_ky_l}\mathbb{E}(\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l)\right]\\
&= \tfrac{1}{24}\left[\sum_{i,j,k,l}(\beta_{x_ix_jx_kx_l}\mathbb{E}(\delta_i\delta_j\delta_k\delta_l) + \beta_{y_iy_jy_ky_l}\mathbb{E}(\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l)) + 6\sum_{i,j,k,l}\beta_{x_ix_jy_ky_l}\mathbb{E}(\delta_i\delta_j)\mathbb{E}(\varepsilon_k\varepsilon_l)\right].
\end{aligned}$$

Moreover, the expected values of terms such as $\delta_i\delta_j\delta_k\delta_l$ and $\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l$ are non zero only if each pair of indices are equal, for instance, when $i = j$ and $k = l$, or $i = k$ and $j = l$ or $i = l$ and $j = k$. Also, $\mathbb{E}(\delta_i\delta_j\varepsilon_k\varepsilon_l) = 1$ only if $i = j$ and $k = l$ and 0, otherwise. Thus,

$$\begin{aligned}
\mathbb{E}(\Delta_4\hat{\beta}) &= \tfrac{1}{24}\Bigg[\sum_{i=1}(\beta_{x_ix_ix_ix_i} + \lambda^2\beta_{y_iy_iy_jy_i})\mathbb{E}(\delta_i^4) + \sum_{i\neq j}3(\beta_{x_ix_ix_jx_j} + \lambda^2\beta_{y_iy_iy_jy_j})\mathbb{E}(\delta_i^2\delta_j^2)\\
&\quad + 6\sum_{i,j=1}\beta_{x_ix_iy_jy_j}\mathbb{E}(\delta_i^2)\mathbb{E}(\varepsilon_j^2)\Bigg]\\
&= \tfrac{1}{24}\Bigg[\sum_{i=1}(\beta_{x_ix_ix_ix_i} + \lambda^2\beta_{y_iy_iy_jy_i})\mathbb{E}(\delta_i^4) + \sum_{i\neq j}3(\beta_{x_ix_ix_jx_j} + \lambda^2\beta_{y_iy_iy_jy_j})\mathbb{E}(\delta_i^2\delta_j^2)\\
&\quad + 6\sum_{i,j=1}\lambda\beta_{x_ix_iy_jy_j}\mathbb{E}(\delta_i^2)\mathbb{E}(\delta_j^2)\Bigg].
\end{aligned}$$

The simple relations, such as $\mathbb{E}(\delta_i^4) = 3\sigma^4$ and $\mathbb{E}(\delta_i^2\delta_j^2) = \sigma^4$ when $i \neq j$, and $\mathbb{E}(\delta_i^2)\mathbb{E}(\varepsilon_j^2) = \sigma^4$ for all i, j, follow from our statistical assumptions. Thus, if we define

$$\omega_{ij} = \beta_{x_ix_ix_jx_j} + \lambda^2\beta_{y_iy_iy_jy_j} + 2\lambda\beta_{x_ix_iy_jy_j}, \tag{A.24}$$

then

$$\mathbb{E}(\Delta_4\hat{\beta}) = \tfrac{\sigma^4}{8}\sum_{i,j=1}^{n}\omega_{ij}. \tag{A.25}$$

Including all terms of order σ^4 in $\mathbb{E}(\Delta_4\beta)$ is a very difficult task, and it results in very lengthy formulas that might not be informative. Therefore, we will consider only terms of typical magnitude σ^4 and will discard other less important terms, such as, σ^4/n, σ^4/n^2, though the computations involved are still somehow awkward.

Lemma A.3. *All summations*

$$\sum_{i=1}\beta_{x_ix_iy_iy_i},\quad \sum_{i=1}\beta_{x_ix_ix_ix_i},\ \textit{and}\ \sum_{i=1}\beta_{y_iy_iy_iy_i}$$

are of order $\mathcal{O}(n^{-1})$. Moreover,

$$\sum_{i\neq j}\beta_{x_ix_ix_jx_j} = \frac{\tilde{\gamma}'}{\tilde{g}^2S^2}\left(2\tilde{\gamma}'' - \frac{3\tilde{\gamma}'\tilde{g}'}{\tilde{g}}\right), \tag{A.26}$$

$$\sum_{i\neq j}\beta_{y_iy_iy_jy_j} = \frac{\tilde{g}'}{\tilde{g}^2S^2}\left(2\tilde{g}'' - \frac{3(\tilde{g}')^2}{\tilde{g}}\right), \tag{A.27}$$

$$\sum_{i\neq j}^{n} \beta_{x_i x_i y_j y_j} = \frac{1}{\tilde{g}^2 S^2}\left(\tilde{\gamma}''\tilde{g}' + \tilde{\gamma}'\tilde{g}'' - \frac{3(\tilde{g}')^2\tilde{\gamma}'}{\tilde{g}}\right). \tag{A.28}$$

Proof of Lemma A.3 First, we differentiate Equation (A.19) with respect to x_j, and we get

$$\begin{aligned}&\mathcal{F}_{\beta\beta}\hat{\beta}_{x_i x_i x_j x_j} + 2\rho^{(1)}_{x_j}\hat{\beta}_{x_i x_i x_j} + \rho^{(1)}_{x_j,x_j}\hat{\beta}_{x_i x_i} + 2\rho^{(2)}_{x_j}\hat{\beta}_{x_i}\hat{\beta}_{x_i x_j} + \rho^{(2)}_{x_j x_j}\hat{\beta}^2_{x_i} + 2\rho^{(1)}_{x_i,x_j}\hat{\beta}_{x_i x_j} + 2\rho^{(1)}_{x_i}\hat{\beta}_{x_i x_j x_j}\\ &+2(\mathcal{F}_{\beta\beta x_i x_j} + \mathcal{F}_{\beta\beta\beta x_i}\hat{\beta}_{x_j})\hat{\beta}_{x_i x_j} + 2(\mathcal{F}_{\beta\beta\beta x_i x_j}\hat{\beta}_{x_j} + (\mathcal{F}_{\beta\beta\beta\beta x_i}\hat{\beta}_{x_j} + \mathcal{F}_{\beta\beta\beta x_i x_j})\hat{\beta}_{x_j}\\ &+\mathcal{F}_{\beta\beta\beta x_i}\hat{\beta}_{x_j x_j})\hat{\beta}_{x_i} + \mathcal{F}_{\beta x_i x_i \beta\beta}\hat{\beta}^2_{x_j} + \mathcal{F}_{\beta x_i x_i \beta}\hat{\beta}_{x_j x_j} = 0,\end{aligned} \tag{A.29}$$

where

$$\rho^{(1)}_{x_i x_j} = \mathcal{F}_{\beta\beta x_i x_j} + \mathcal{F}_{\beta\beta\beta x_i}\hat{\beta}_{x_j} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}_{x_i x_j} + \rho^{(2)}_{x_j}\hat{\beta}_{x_i}, \tag{A.30}$$

$$\rho^{(2)}_{x_i x_j} = \mathcal{F}_{\beta\beta\beta x_i x_j} + \mathcal{F}_{\beta\beta\beta\beta x_i}\hat{\beta}_{x_j} + \mathcal{F}_{\beta\beta\beta\beta}\hat{\beta}_{x_i x_j} + (\mathcal{F}_{\beta\beta\beta\beta\beta}\hat{\beta}_{x_j} + \mathcal{F}_{\beta\beta\beta\beta x_j})\hat{\beta}_{x_i} \tag{A.31}$$

represent the total derivatives of $\rho^{(1)}_{x_i}$ and $\rho^{(2)}_{x_i}$ with respect to x_j, respectively. When evaluated at the true values of the observations, Equation (A.29) gives us $\beta_{x_i x_i x_j x_j}$. A careful look at each term in Equation (A.29) shows that the order of magnitude of each term is of at most $\frac{1}{n}$, since $\frac{1}{\tilde{\mathcal{F}}_{\beta\beta}} \sim \mathcal{O}(n^{-1})$. This shows that $\sum_i \beta_{x_i x_i x_i x_i} \sim \mathcal{O}(n^{-1})$.

On the other hand, $\sum_{i\neq j}\beta_{x_i x_i x_j x_j} \sim \mathcal{O}(1)$. More precisely, only $\tilde{\rho}^{(1)}_{x_j x_j}\beta_{x_i x_i}$ and $\tilde{\mathcal{F}}_{\beta x_i x_i \beta}\beta_{x_j x_j}$ have the highest order of magnitude among all other terms in Equation (A.29). They are of order $\mathcal{O}(1)$, and as such

$$\sum_{i\neq j}\beta_{x_i x_i x_j x_j} = -\frac{1}{\tilde{\mathcal{F}}_{\beta\beta}}\sum_{i\neq j}\tilde{\rho}^{(1)}_{x_j x_j}\beta_{x_i x_i} + \tilde{\mathcal{F}}_{\beta x_i x_i \beta}\beta_{x_j x_j} + \mathcal{O}(n^{-1}).$$

From Lemma (A.1) and Equation (28), Equation (A.30) takes the form

$$\tilde{\rho}^{(1)}_{x_j x_j} = 2\tilde{\gamma}'' - \frac{6\tilde{g}'\tilde{\gamma}'}{\tilde{g}} + \mathcal{O}(n^{-1}) \tag{A.32}$$

when evaluated at the true values $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$. Here $\beta_{x_i x_i} = -\frac{\tilde{\gamma}'}{\tilde{g}nS} + \mathcal{O}(n^{-2})$ and $\tilde{\mathcal{F}}_{\beta\beta\beta} = 6\tilde{g}'nS$. For this reason,

$$\sum_{i\neq j}\beta_{x_i x_i x_j x_j} = \frac{2\tilde{\gamma}'(n-1)\left(2\tilde{\gamma}'' - 3\frac{\tilde{\gamma}'\tilde{g}'}{\tilde{g}}\right)}{\tilde{g}S\tilde{\mathcal{F}}_{\beta\beta}} + \mathcal{O}(n^{-1}). \tag{A.33}$$

Substituting $\tilde{\mathcal{F}}_{\beta\beta} = 2\tilde{g}nS$ in Equation (A.33) and ignoring -1 in the numerator of Equation (A.33) establish Equation (A.26).

Next we compute $\beta_{y_i y_i y_j y_j}$. If Equation (A.21) is differentiated with respect to y_j, we get an equation that is similar in the structure to (A.29), and as such, it is easy to verify that $\sum_i \beta_{y_i y_i y_i y_i} \sim \mathcal{O}(n^{-1})$ and

$$\sum_{i\neq j}\beta_{y_i y_i y_j y_j} = -\frac{1}{\tilde{\mathcal{F}}_{\beta\beta}}\sum_{i\neq j}\tilde{\xi}^{(1)}_{y_j y_j}\beta_{y_i y_i} + \tilde{\mathcal{F}}_{\beta y_i y_i \beta}\beta_{y_j y_j},$$

where

$$\xi^{(1)}_{y_i y_j} = \mathcal{F}_{\beta\beta y_i y_j} + \mathcal{F}_{\beta\beta\beta y_i}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta\beta}\hat{\beta}_{y_i y_j} + (\mathcal{F}_{\beta\beta\beta\beta}\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta\beta y_j})\hat{\beta}_{y_i} \tag{A.34}$$

is the total derivative of $\xi^{(1)}_{y_i}$ with respect to y_j. Since, up to the leading term, $\beta_{y_iy_i} = \frac{-\tilde{g}'}{\tilde{g}nS}$ and $\tilde{\mathcal{F}}_{\beta\beta y_iy_i} = 2\tilde{g}''$ and $\tilde{\mathcal{F}}_{\beta\beta\beta} = 6\tilde{g}'nS$, then we have

$$\tilde{\xi}^{(1)}_{y_jy_j} = 2\tilde{g}'' - \frac{6(\tilde{g}')^2}{\tilde{g}} + \mathcal{O}(n^{-1}), \tag{A.35}$$

and as such

$$\sum_{i\neq j} \beta_{y_iy_iy_jy_j} = \frac{2(n-1)\tilde{g}'\left(2\tilde{g}'' - 3\frac{(\tilde{g}')^2}{\tilde{g}}\right)}{\tilde{g}S\tilde{\mathcal{F}}_{\beta\beta}}. \tag{A.36}$$

Substituting $\tilde{\mathcal{F}}_{\beta\beta} = 2\tilde{g}nS$ in Equation (A.36) and ignoring -1 in the numerator establish Equation (A.27).

Next we verify the last identity in Equation (A.28), which takes more efforts. To get $\beta_{x_ix_iy_jy_j}$, we differentiate Equation (A.22) with respect to y_j. This gives us

$$\begin{aligned}
&\mathcal{F}_{\beta\beta}\hat{\beta}_{x_ix_iy_jy_j} + 2\xi^{(1)}_{y_j}\hat{\beta}_{x_ix_iy_j} + \xi^{(1)}_{y_jy_j}\hat{\beta}_{x_ix_i} + 2\rho^{(1)}_{x_iy_j}\hat{\beta}_{x_iy_j} + 2\rho^{(1)}_{x_i}\hat{\beta}_{x_iy_jy_j} + \xi^{(2)}_{y_jy_j}\hat{\beta}^2_{x_i}\\
&\quad + 2\xi^{(2)}_{y_j}\hat{\beta}_{x_i}\hat{\beta}_{x_iy_j} + 2(\mathcal{F}_{\beta\beta x_iy_j} + \mathcal{F}_{\beta\beta\beta x_i}\hat{\beta}_{y_j})\hat{\beta}_{x_iy_j} + 2(\mathcal{F}_{\beta\beta\beta x_iy_j}\hat{\beta}_{y_j} + (\mathcal{F}_{\beta\beta\beta\beta x_i}\hat{\beta}_{y_j}\\
&\quad + \mathcal{F}_{\beta\beta\beta x_iy_j})\hat{\beta}_{y_j} + \mathcal{F}_{\beta\beta\beta x_i}\hat{\beta}_{y_jy_j})\hat{\beta}_{x_i} + \mathcal{F}_{\beta x_ix_i\beta\beta}\hat{\beta}^2_{y_j} + \mathcal{F}_{\beta x_ix_i\beta}\hat{\beta}_{y_jy_j} = 0.
\end{aligned} \tag{A.37}$$

Equation (A.37) involves $\beta_{x_iy_jy_j}$, which is given in Equation (A.23). Again it should be obvious that $\sum_i \beta_{x_ix_iy_iy_i} \sim \mathcal{O}(n^{-1})$, while

$$\sum_{i\neq j} \beta_{x_ix_iy_jy_j} = \frac{-1}{\tilde{\mathcal{F}}_{\beta\beta}} \sum_{i\neq j}^{n} \left(\tilde{\xi}^{(1)}_{y_jy_j}\beta_{x_ix_i} + \tilde{\mathcal{F}}_{\beta x_ix_i\beta}\beta_{y_jy_j}\right) + \mathcal{O}(n^{-1}). \tag{A.38}$$

Using Equation (A.35) and $\beta_{x_ix_i} = -\frac{\tilde{\gamma}'}{\tilde{g}nS} + \mathcal{O}(n^{-2})$ and $\tilde{\mathcal{F}}_{\beta\beta\beta} = 6\tilde{g}'nS$ reduces Equation (A.38) to

$$\sum_{i\neq j} \beta_{x_ix_iy_jy_j} = \frac{2(n-1)}{\tilde{g}S\tilde{\mathcal{F}}_{\beta\beta}}\left(\tilde{\gamma}''\tilde{g}' + \tilde{\gamma}'\tilde{g}'' - \frac{3(\tilde{g}')^2\tilde{\gamma}'}{\tilde{g}}\right) + \mathcal{O}(n^{-1}). \tag{A.39}$$

If we substitute $\tilde{\mathcal{F}}_{\beta\beta} = 2\tilde{g}nS$ in Equation (A.39) and ignore -1 in the numerator, we will get Equation (A.28). This completes the proof of the Lemma. □

Lemma A.3 shows that the most important terms in Equation (A.25) are represented by $\sum_{i\neq j}\omega_{ij}$ which is of order $\mathcal{O}(1)$, while the terms in $\sum_{i=1}^{n}\omega_{i,i}$ are of order $\mathcal{O}(\frac{1}{n})$, and as such, they are less important and will be discarded in our analysis. Therefore, if one substitutes Equations (A.26)–(A.28) into Equation (A.25), then up to order $\mathcal{O}(\frac{1}{n})$

$$\begin{aligned}
\sum_{i\neq j}^{n}\omega_{ij} &= \frac{1}{\tilde{g}^2S^2}\left[2\tilde{\gamma}''\tilde{\gamma}' + 2\lambda^2\tilde{g}''\tilde{g}' + 2\lambda\tilde{g}'\gamma'' + 2\lambda\tilde{g}''\gamma' - \frac{3\tilde{g}'}{\tilde{g}}\left((\tilde{\gamma}')^2 + 2\lambda\tilde{g}'\tilde{\gamma}' + \lambda^2(\tilde{g}')^2\right)\right]\\
&= \frac{1}{\tilde{g}^2S^2}\left[2\tilde{\kappa}'\tilde{\kappa}'' - \frac{3\tilde{g}'(\kappa')^2}{\tilde{g}}\right],
\end{aligned}$$

where $\kappa(\beta) = (\beta^2 + \lambda)g(\beta)$ (and as such $\kappa'(\beta) = (\beta^2 + \lambda)g' + 2\beta g = \gamma' + \lambda g'$). For this reason

$$\sum_{i \neq j}^{n} \omega_{ij} = \frac{1}{\tilde{g}^2 S^2} \left[2\tilde{\kappa}'\tilde{\kappa}'' - \frac{3\tilde{g}'(\kappa')^2}{\tilde{g}} \right] + \mathcal{O}(n^{-1}). \tag{A.40}$$

This completes the derivation of Equation (57).