Derivations of the statistical properties of reversed inverse regression (supplementary material 1)

Supplementary material for the paper entitled "Reversed inverse regression for the univariate linear calibration and its statistical properties using a new methodology" (IJMQE)

Pilsang Kang[1], Changhoi Koo and Hokyu Roh
Quality Management Center KEPCO NF
[1] pskang@knfc.co.kr

Note: "∑" denotes summation from i = 1 to n throughout this material.

A.1 Variances of slope and intercept

<Variance of slope>

$$\hat\beta = \sum(x_i - \bar x)(y_i - \bar y)/\sum(x_i - \bar x)^2 = \sum(x_i - \bar x)\, y_i/\sum(x_i - \bar x)^2$$

Let $A = \sum(x_i - \bar x)\, y_i$, $B = \sum(x_i - \bar x)^2$

$$\partial A/\partial x_i = (-1/n)\, y_1 + (-1/n)\, y_2 + \text{----} + (1 - 1/n)\, y_i + \text{---} + (-1/n)\, y_n$$
$$= (-1/n)\,(y_1 + y_2 + \text{---} + y_n) + y_i = (-1/n)\sum y_i + y_i$$
$$\partial B/\partial x_i = 2(x_1 - \bar x)(-1/n) + 2(x_2 - \bar x)(-1/n) + \text{----} + 2(x_i - \bar x)(1 - 1/n) + \text{---} + 2(x_n - \bar x)(-1/n)$$
$$= (-2/n)\sum(x_i - \bar x) + 2\,(x_i - \bar x) = 2\,(x_i - \bar x)$$
$$\partial\hat\beta/\partial x_i = \{(\partial A/\partial x_i)\, B - A\,(\partial B/\partial x_i)\}/B^2$$
$$= \{(-1/n)\,(\textstyle\sum y_i)\sum(x_i - \bar x)^2 + y_i\sum(x_i - \bar x)^2 - 2(x_i - \bar x)\sum(x_i - \bar x)\, y_i\}/\{\sum(x_i - \bar x)^2\}^2$$

$$\{\sigma(\hat\beta)\}^2 \approx \sum(\partial\hat\beta/\partial x_i)^2(\sigma_{xi})^2$$
$$= [(1/n)\,(\textstyle\sum y_i)^2\{\sum(x_i - \bar x)^2\}^2 + \sum y_i^2\{\sum(x_i - \bar x)^2\}^2 + 4\sum(x_i - \bar x)^2\{\sum(x_i - \bar x)\, y_i\}^2$$
$$+ (-2/n)\textstyle\sum y_i^2\{\sum(x_i - \bar x)^2\}^2 - 4\{\sum(x_i - \bar x)\, y_i\}^2\sum(x_i - \bar x)^2$$
$$+ (4/n)\textstyle\sum(x_i - \bar x)\sum y_i\sum(x_i - \bar x)^2\sum(x_i - \bar x)\, y_i]\sigma_x^2/\{\sum(x_i - \bar x)^2\}^4$$
$$= [\,(1/n - 2/n)(\textstyle\sum y_i)^2\{\sum(x_i - \bar x)^2\}^2 + \sum y_i^2\{\sum(x_i - \bar x)^2\}^2\,]\sigma_x^2/\{\sum(x_i - \bar x)^2\}^4$$
$$= (\textstyle\sum y_i^2 - n\bar y^2)\sigma_x^2/\{\sum(x_i - \bar x)^2\}^2$$
$$= [(\textstyle\sum y_i^2 - n\bar y^2)\sigma_x^2/\{\sum(x_i - \bar x)^2\}^2]\hat\beta^2\,[\{\sum(x_i - \bar x)^2\}^2/\{\sum(x_i - \bar x)\, y_i\}^2]$$
$$= [(\textstyle\sum y_i^2 - n\bar y^2)/\{\sum(x_i - \bar x)\, y_i\}^2]\sigma_x^2\hat\beta^2$$
$$= [(\textstyle\sum y_i^2 - n\bar y^2)/\{\sum(x_i - \bar x)\,(y_i - \bar y)\}^2]\sigma_x^2\hat\beta^2$$
$$= [S_{yy}/S_{xy}^2]\sigma_x^2\hat\beta^2$$

$$\therefore \mathrm{V}[\hat\beta] \approx [S_{yy}/S_{xy}^2]^*\sigma_x^2\beta^2 = [S_{yy}/S_{xy}^2]^*\sigma^2 \qquad \text{(A1)}$$

<Variance of intercept>

$$\bar y = \hat\alpha + \hat\beta\bar x \quad\rightarrow\quad \hat\alpha = \bar y - \hat\beta\bar x$$
$$\hat\alpha = \bar y - \hat\beta\bar x = \textstyle\sum y_i/n - \{\sum(x_i - \bar x)y_i/\sum(x_i - \bar x)^2\}(\sum x_i/n)$$

Let $C = \sum(x_i - \bar x)y_i/\sum(x_i - \bar x)^2$, $D = \sum x_i/n$

$$\sum(\partial C/\partial x_i)^2 D^2 \sigma_{xi}^2 = (\sum x_i/n)^2\sum(\partial C/\partial x_i)^2\sigma_{xi}^2$$
$$= (\sum x_i/n)^2 (S_{yy}/S_{xy}^2)\sigma_x^2\hat{\beta}^2$$
$$\sum C^2(\partial D/\partial x_i)^2 \sigma_{xi}^2 = C^2\sum (\partial D/\partial x_i)^2 \sigma_{xi}^2$$
$$= C^2(1/n)\sigma_x^2 = (1/n)\sigma_x^2\hat{\beta}^2$$
$$\sum 2CD(\partial C/\partial x_i)(\partial D/\partial x_i)\sigma_{xi}^2$$
$$= \sum 2[\{\sum (x_i - \bar{x})y_i/\sum(x_i - \bar{x})^2\}(\sum x_i/n)]\,[(-1/n)\sum y_i\sum(x_i - \bar{x})^2 + y_i\sum(x_i - \bar{x})^2$$
$$- 2(x_i - \bar{x})\{\sum (x_i - \bar{x})y_i\}]\sigma_{xi}^2(1/n)/\{\sum(x_i - \bar{x})^2\}^2$$
$$= (2/n)\,\bar{x}\hat{\beta}\sigma_{xi}^2\{-\sum y_i\sum(x_i-\bar{x})^2+\sum y_i\sum(x_i-\bar{x})^2-2\sum(x_i-\bar{x})y_i\sum(x_i-\bar{x})\}/\{\sum(x_i-\bar{x})^2\}^2 = 0$$

$$\{\sigma(\hat{\alpha})\}^2 \approx \sum\{(\partial C/\partial x_i)^2 D^2 + C^2(\partial D/\partial x_i)^2 + 2CD(\partial C/\partial x_i)(\partial D/\partial x_i)\}\sigma_{xi}^2$$
$$= \bar{x}^2(S_{yy}/S_{xy}^2)\sigma_x^2\hat{\beta}^2 + (1/n)\sigma_x^2\hat{\beta}^2$$
$$= [1/n + \bar{x}^2(S_{yy}/S_{xy}^2)]\sigma_x^2\hat{\beta}^2$$
$$= [1/n + \bar{x}^2(S_{yy}/S_{xy}^2)]^*[\hat{\beta}^2]^*\sigma_x^2$$
$$= [1/n + \bar{x}^2(S_{yy}/S_{xy})^2\}]^*\sigma^2$$

$$\therefore\ V[\hat{\alpha}] \approx [1/n + \bar{x}^2(S_{yy}/S_{xy}^2)]^*\sigma^2 \qquad \text{(A2)}$$

A.2 Expectations of slope and intercept

$K = (x_i - \bar{x})$

$\sigma_K^2 = \sum(\partial\hat{\beta}/\partial x_i)^2\sigma_{xi}^2$ $\because$ K is a linear function of variables x_i, (i = 1, 2, ---, n).

$\therefore$ $\mathrm{Var}[x_i - \bar{x}] = \{(n-1)/n\}\sigma_{xi}^2$

Accordingly, $\mathrm{E}[\sum(x_i-\bar{x})^2] = \sum(x_{i0}-\bar{x}_0)^2 + (n-1)\sigma_x^2$

$F = \sum(x_i - \bar{x})y_i/\sum(x_i - \bar{x})^2$

Let $A = \sum(x_i - \bar{x})y_i$, $B = 1/\sum(x_i - \bar{x})^2$, $F = A\cdot B$

$$\sigma_F^2 \approx \sum(\partial F/\partial x_i)^2\sigma_{xi}^2$$
$$\sigma_F^2 \approx (\partial F/\partial A)^2\sigma_A^2 + (\partial F/\partial B)^2\sigma_B^2 + 2(\partial F/\partial A)(\partial F/\partial B)\mathrm{cov}[A, B]$$

$\therefore$ $\mathrm{cov}[A, B] = \mathrm{cov}[\sum(x_i-\bar{x})y_i,\ 1/\sum(x_i-\bar{x})^2] \approx -[2/S_{xy}]\beta^2\sigma_x^2$

$F = \sum(x_i - \bar{x})y_i\sum(x_i - \bar{x})^2$

Let $A = \sum(x_i - \bar{x})y_i$, $B = \sum(x_i - \bar{x})^2$, $F = A\cdot B$

$$\sigma_F^2 \approx \sum(\partial F/\partial x_i)^2\sigma_{xi}^2$$
$$\sigma_F^2 \approx (\partial F/\partial A)^2\sigma_A^2 + (\partial F/\partial B)^2\sigma_B^2 + 2(\partial F/\partial A)(\partial F/\partial B)\mathrm{cov}[A, B]$$

$\therefore$ $\mathrm{cov}[A, B] = \mathrm{cov}[\sum(x_i-\bar{x})y_i,\ \sum(x_i-\bar{x})^2] \approx 2S_{xy}\sigma_x^2$

$F = \sum(x_i - \bar{x})^2\{1/\sum(x_i - \bar{x})^2\}$

Let $A = \sum(x_i - \bar{x})^2$, $B = 1/\sum(x_i - \bar{x})^2$, $F = A\cdot B$

$$\sigma_F^2 \approx \sum(\partial F/\partial x_i)^2\sigma_{xi}^2$$
$$\sigma_F^2 \approx (\partial F/\partial A)^2\sigma_A^2 + (\partial F/\partial B)^2\sigma_B^2 + 2(\partial F/\partial A)(\partial F/\partial B)\mathrm{cov}[A, B]$$

$$\therefore\ \text{cov}[A, B] = \text{cov}[\textstyle\sum(x_i-\bar{x})^2,\ 1/\sum(x_i-\bar{x})^2] \approx -[4/S_{xx}]\,\sigma_x^2$$

The expectation of $\sum(x_i-\bar{x})y_i/\sum(x_i-\bar{x})^2$ is derived as follows:

$$\beta_E = \text{E}[\textstyle\sum(x_i-\bar{x})\,y_i/\sum(x_i-\bar{x})^2]$$
$$= \text{E}[\textstyle\sum(x_i-\bar{x})y_i]\cdot\text{E}[1/\sum(x_i-\bar{x})^2] + \text{cov}[\sum(x_i-\bar{x})y_i,\ 1/\sum(x_i-\bar{x})^2]$$

$$\text{E}[1] = \text{E}[\textstyle\sum(x_i-\bar{x})^2/\sum(x_i-\bar{x})^2]$$
$$= \text{E}[\textstyle\sum(x_i-\bar{x})^2]\cdot\text{E}[1/\sum(x_i-\bar{x})^2] + \text{cov}[\sum(x_i-\bar{x})^2,\ 1/\sum(x_i-\bar{x})^2]$$
$$\text{E}[\textstyle\sum(x_i-\bar{x})^2] = \sum(x_{i0}-\bar{x}_0)^2 + (n-1)\sigma_x^2$$
$$\text{cov}[\textstyle\sum(x_i-\bar{x})y_i,\ 1/\sum(x_i-\bar{x})^2] \approx -[2/S_{xy}]^*\beta^2\sigma_x^2$$
$$\text{cov}[\textstyle\sum(x_i-\bar{x})^2,\ 1/\sum(x_i-\bar{x})^2] \approx -[4/S_{xx}]^*\sigma_x^2$$

$$\therefore\ \beta_E \approx \beta - \beta[1/S_{xx}]^*(n-3)\sigma_x^2 \qquad \text{(A3)}$$
$$\alpha_E \approx \alpha + \bar{x}_0\beta[1/S_{xx}]^*(n-3)\sigma_x^2$$

For comparison purposes, β_E is also derived using the Delta Method as follows:

$$\text{E}[A/B] \approx \text{E}[A]/\text{E}[B] + \text{V}[B]\text{E}[A]/\{\text{E}[B]\}^3 - \text{cov}[A, B]/\{\text{E}[B]\}^2 \text{ (By the Delta Method)}$$

{Parker et al. (2010) and Pham-Gia et al. (2006)}

$$\text{E}[\textstyle\sum(x_i-\bar{x})y_i/\sum(x_i-\bar{x})^2] \approx \text{E}[\sum(x_i-\bar{x})y_i]/\text{E}[\sum(x_i-\bar{x})^2]+\text{V}[\sum(x_i-\bar{x})^2]\text{E}[\sum(x_i-\bar{x})y_i]/\{\text{E}[\sum(x_i-\bar{x})^2]\}^3$$
$$- \text{cov}[\textstyle\sum(x_i-\bar{x})y_i,\ \sum(x_i-\bar{x})^2]/\{\text{E}[\sum(x_i-\bar{x})^2]\}^2$$
$$\approx [S_{yy}/\{S_{xx} + (n-1)\sigma_x^2\}]^* + [4S_{xx}S_{xy}\sigma_x^2/\{S_{xx} + (n-1)\sigma_x^2\}^3]^*$$
$$- [2S_{xy}\sigma_x^2/\{S_{xx} + (n-1)\sigma_x^2\}^2]^*$$

Based on the first-order Taylor approximation,

$$[S_{xy}/\{S_{xx} + (n-1)\sigma_x^2\}]^* \approx [S_{xy}/S_{xx}]^* - [S_{xy}/S_{xx}]^*[(n-1)\sigma_x^2/S_{xx}]^*$$
$$= \beta - \beta[1/S_{xx}]^*(n-1)\sigma_x^2$$
$$[4S_{xx}S_{xy}\sigma_x^2/\{S_{xx} + (n-1)\sigma_x^2\}^3]^* \approx [4S_{xy}\sigma_x^2/S_{xx}^2]^* - 3[4S_{xy}\sigma_x^2/S_{xx}^2]^*[(n-1)\sigma_x^2/S_{xx}]^*$$
$$\approx [4S_{xy}\sigma_x^2/S_{xx}^2]^*$$
$$[2S_{xy}\sigma_x^2/\{S_{xx} + (n-1)\sigma_x^2\}^2]^* \approx [2S_{xy}\sigma_x^2/S_{xx}^2]^* - 2[2S_{xy}\sigma_x^2/S_{xx}^2]^*[(n-1)\sigma_x^2/S_{xx}]^*$$
$$\approx [2S_{xy}\sigma_x^2/S_{xx}^2]^*$$

In the previous approximation calculations, $3[4S_{xy}\sigma_x^2/S_{xx}^2]^*[(n-1)\sigma_x^2/S_{xx}]^*$ and $2[2S_{xy}\sigma_x^2/S_{xx}^2]^*[(n-1)\sigma_x^2/S_{xx}]^*$ are error terms of orders higher than σ_x^2 and ignored.

$$\therefore\ \beta_E = \text{E}[\hat{\beta}] = \text{E}[\textstyle\sum(x_i-\bar{x})\,y_i/\sum(x_i-\bar{x})^2]$$
$$\approx \beta - \beta[1/S_{xx}]^*(n-1)\sigma_x^2 + [4S_{xy}\sigma_x^2/S_{xx}^2]^* - [2S_{xy}\sigma_x^2/S_{xx}^2]^*$$
$$= \beta - \beta[1/S_{xx}]^*(n-3)\sigma_x^2$$
$$\alpha_E = \text{E}[\hat{\alpha}] \approx \alpha + \bar{x}_0\,\beta[1/S_{xx}]^*(n-3)\sigma_x^2$$

A.3 Miscellaneous correlation coefficients, covariances and expectations

<Correlation coefficient $r(\hat{\beta}, \bar{x})$>

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\{\sigma(\hat{\alpha})\}^2 \approx (\partial\hat{\alpha}/\partial\hat{\beta})^2\{\sigma(\hat{\beta})\}^2 + (\partial\hat{\alpha}/\partial\bar{x})^2(\sigma\bar{x})^2 + 2(\partial\hat{\alpha}/\partial\hat{\beta})(\partial\hat{\alpha}/\partial\bar{x})\{\sigma(\hat{\beta})\}\{\sigma(\bar{x})\}r(\hat{\beta}, \bar{x})$$
$$\approx \bar{x}^2\{S_{yy}/S_{xy}^2\}\sigma_x^2\hat{\beta}^2 + (1/n)\,\sigma_x^2\hat{\beta}^2$$
$$+ 2\,(-\bar{x})\,(-\hat{\beta})\,[(S_{yy}/S_{xy}^2)\sigma_x^2\hat{\beta}^2]^{1/2}\,(\sigma_x^2/n)^{1/2}r(\hat{\beta}, \bar{x})$$
$$= [1/n + \bar{x}^2\,(S_{yy}/S_{xy}^2)]\sigma^2$$
$$+ 2\,(-\bar{x})\,(-\hat{\beta})\,[(S_{yy}/S_{xy}^2)\sigma_x^2\hat{\beta}^2]^{1/2}\,(\sigma_x^2/n)^{1/2}r(\hat{\beta}, \bar{x}) \qquad \text{(A4)}$$

From Eq. (A2) and (A4), correlation coefficient $r(\hat{\beta}, \bar{x})$ is obtained as follows:

$$r(\hat{\beta}, \bar{x}) \approx 0$$

<Correlation coefficient $r(\hat{\beta}, x_i)$>

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \sum y_i/n - \hat{\beta}(\sum x_i/n)$$

$$\{\sigma(\hat{\alpha})\}^2 \approx \sum(\partial\hat{\alpha}/\partial x_i)^2\sigma_{xi}^2 + (\partial\hat{\alpha}/\partial\hat{\beta})^2\{\sigma(\hat{\beta})\}^2 + 2\sum(\partial\hat{\alpha}/\partial x_i)(\partial\hat{\alpha}/\partial\hat{\beta})\{\sigma(x_i)\}\{\sigma(\hat{\beta})\}r(\hat{\beta}, x_i)$$
$$\approx [1/n + \bar{x}^2(S_{yy}/S_{xy}^2)]\sigma^2 + \sum 2(-\bar{x})(-\hat{\beta}/n)[(S_{yy}/S_{xy}^2)\,\hat{\beta}^2]^{1/2}\sigma_x^2 r(\hat{\beta}, x_i) \qquad \text{(A5)}$$

From Eq. (A2) and (A5), the following is obtained:

$$\sum r(\hat{\beta}, x_i) \approx 0$$

Similarly, $r(\hat{\beta}, x_i) \approx \{(y_i - \bar{y})S_{xx} - 2(x_i - \bar{x})S_{xy}\}/(S_{xx}S_{yy}^{1/2})$

<Expectation of $\sum\hat{\beta}(x_i - \bar{x})$>

$$\sum \mathrm{E}[\hat{\beta}x_i] = \sum\{\mathrm{cov}[\hat{\beta}, x_i] + \mathrm{E}[\hat{\beta}]\mathrm{E}[x_i]\} = \sum r(\hat{\beta}, x_i)\{\sigma(\hat{\beta})\}\sigma_{xi} + \sum\mathrm{E}[\hat{\beta}]\mathrm{E}[x_i]$$
$$= \sum\mathrm{E}[\hat{\beta}]\mathrm{E}[x_i] = \beta_E\sum x_{i0}$$
$$\mathrm{E}[\sum\hat{\beta}(x_i - \bar{x})] = \sum\mathrm{E}[\hat{\beta}(x_i - \bar{x})] = \sum\{\mathrm{E}[\hat{\beta}x_i] - \mathrm{E}[\hat{\beta}\bar{x}]\} = \sum\mathrm{E}[\hat{\beta}x_i] - \sum\mathrm{E}[\hat{\beta}]\mathrm{E}[\bar{x}]$$
$$= \beta_E\sum x_{i0} - \beta_E\sum\bar{x}_0 = \beta_E\sum x_{i0} - n\bar{x}_0\,\beta_E = 0$$
$$\therefore \sum\mathrm{E}[\hat{\beta}]\mathrm{E}[x_i - \bar{x}] = \beta_E\sum(x_{i0} - \bar{x}_0) = \beta_E\sum x_{i0} - n\,\bar{x}_0\,\beta_E = \mathrm{E}[\sum\hat{\beta}(x_i - \bar{x})]$$

<Covariance of $\hat{\beta}^2$ and $\sum(x_i - \bar{x})^2$>

Let $M = \hat{\beta}^2\sum(x_i - \bar{x})^2$, $A = \hat{\beta}^2$, $B = \sum(x_i - \bar{x})^2$, $M = A\cdot B$

$$\sigma_M^2 \approx \sum(\partial M/\partial x_i)^2\,\sigma_{xi}^2$$
$$= 4\hat{\beta}^2\{\sum(x_i - \bar{x})^2\}^2(S_{yy}/S_{xy}^2)\sigma_x^2\hat{\beta}^2 - 8\hat{\beta}^3\sigma_x^2\sum(x_i - \bar{x})y_i$$
$$+ 4\hat{\beta}^4\sigma_x^2\sum(x_i - \bar{x})^2$$
$$\sigma_M^2 \approx (\partial M/\partial A)^2\sigma_A^2 + (\partial M/\partial B)^2\sigma_B^2 + 2(\partial M/\partial A)(\partial M/\partial B)\,\mathrm{cov}[\hat{\beta}^2, \sum(x_i - \bar{x})^2]$$
$$\approx 4\hat{\beta}^2\{\sum(x_i - \bar{x})^2\}^2(S_{yy}/S_{xy}^2)\sigma_x^2\hat{\beta}^2 + 4\hat{\beta}^4\sigma_x^2\sum(x_i - \bar{x})^2$$
$$+ 2\hat{\beta}^2\{\sum(x_i - \bar{x})^2\}\,\mathrm{cov}[\hat{\beta}^2, \sum(x_i - \bar{x})^2]$$

$$\therefore \mathrm{cov}[\hat{\beta}^2, \sum(x_i - \bar{x})^2] \approx -4\hat{\beta}^2\sigma_x^2$$

<Covariance of $\hat{\beta}$ and $\sum(x_i - \bar{x})\,y_i$>

Let $P = \hat{\beta}\sum(x_i - \bar{x})\,y_i$, $A = \hat{\beta}$, $B = \sum(x_i - \bar{x})\,y_i$, $P = A\cdot B$

$$\sigma_p^2 \approx \sum (\partial P/\partial x_i)^2\sigma_{xi}^2$$
$$= S_{yy}\sigma_x^2\hat{\beta}^2 + \hat{\beta}^2\sigma_x^2S_{yy} + 2\hat{\beta}\sigma_x^2 (S_{xy}S_{yy}S_{xx} - 2S_{xy}^3)/S_{xx}^2$$
$$\sigma_p^2 \approx (\partial P/\partial A)^2\sigma_A^2 + (\partial P/\partial B)^2\sigma_B^2 + 2(\partial P/\partial A)(\partial P/\partial B)\,\mathrm{cov}[\hat{\beta}, \sum(x_i - \bar{x})y_i]$$
$$\approx S_{yy}\sigma_x^2\hat{\beta}^2 + \hat{\beta}^2\sigma_x^2S_{yy} + 2\hat{\beta}S_{xy}\mathrm{cov}[\hat{\beta}, \sum(x_i - \bar{x})y_i]$$
$$\therefore \mathrm{cov}[\hat{\beta}, \sum(x_i - \bar{x})y_i] \approx (S_{yy}S_{xx}/S_{xy}^2 - 2)\hat{\beta}^2\sigma_x^2$$

A.4 Expectation of mean squared error

$$SSE = \sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = \sum\{y_i - (\bar{y} - \hat{\beta}\bar{x}) - \hat{\beta}x_i\}^2 = \sum\{(y_i - \bar{y}) - (\hat{\beta}x_i - \hat{\beta}\bar{x})\}^2$$
$$= \sum(y_i - \bar{y})^2 - 2\sum(y_i - \bar{y})\hat{\beta}(x_i - \bar{x}) + \sum\hat{\beta}^2(x_i - \bar{x})^2 \quad [\text{Normally, } y_i - \bar{y} \neq \hat{\beta}(x_i - \bar{x})]$$

$$\therefore \mathrm{E}[SSE] = \mathrm{E}[\sum(y_i - \bar{y})^2 - 2\sum(y_i - \bar{y})\hat{\beta}(x_i - \bar{x}) + \sum\hat{\beta}^2(x_i - \bar{x})^2]$$
$$= \mathrm{E}[\sum(y_i - \bar{y})^2] - 2\mathrm{E}[\sum(y_i - \bar{y})\hat{\beta}(x_i - \bar{x})] + \mathrm{E}[\sum\hat{\beta}^2(x_i - \bar{x})^2]$$
$$= \mathrm{E}[\sum(y_i - \bar{y})^2] - 2\{\mathrm{E}[\sum\hat{\beta}(x_i - \bar{x})y_i] - \bar{y}\mathrm{E}[\hat{\beta}\sum(x_i - \bar{x})]\} + \mathrm{E}[\sum\hat{\beta}^2(x_i - \bar{x})^2]$$
$$= \mathrm{E}[\sum(y_i - \bar{y})^2] - 2\{\mathrm{E}[\hat{\beta}]\cdot\mathrm{E}[\sum(x_i - \bar{x})y_i] + \mathrm{cov}[\hat{\beta}, \sum(x_i - \bar{x})y_i] - \bar{y}\mathrm{E}[\hat{\beta}]\cdot\mathrm{E}[\sum(x_i - \bar{x})]\}$$
$$+ \mathrm{E}[\hat{\beta}^2]\cdot\mathrm{E}[\sum(x_i - \bar{x})^2] + \mathrm{cov}[\hat{\beta}^2, \sum(x_i - \bar{x})^2]$$
$$= \mathrm{E}[\sum(y_i - \bar{y})^2] - 2\{\mathrm{E}[\hat{\beta}]\cdot\mathrm{E}[\sum(x_i - \bar{x})y_i] + \mathrm{cov}[\hat{\beta}, \sum(x_i - \bar{x})y_i] - 0\}$$
$$+ \{\mathrm{V}[\hat{\beta}] + \mathrm{E}^2[\hat{\beta}]\}\{\sum\mathrm{V}[x_i - \bar{x}] + \sum\mathrm{E}^2[x_i - \bar{x}]\} + \mathrm{cov}[\hat{\beta}^2, \sum(x_i - \bar{x})^2]$$

To remove error terms of orders higher than σ_x^2, some approximate calculations for $\mathrm{E}[\hat{\beta}]$ and $\mathrm{E}^2[\hat{\beta}]$ are performed separately in advance:

$$\mathrm{E}^2[\hat{\beta}] \approx \{\beta - \beta[1/S_{xx}]^*(n-3)\sigma_x^2\}^2$$
$$\approx \beta^2 - 2\beta^2[1/S_{xx}]^*(n-3)\sigma_x^2$$

$$\{-2\mathrm{E}[\hat{\beta}]\cdot\mathrm{E}[\sum(x_i - \bar{x})y_i]\} + \mathrm{E}^2[\hat{\beta}]\cdot\{\sum\mathrm{V}[x_i - \bar{x}] + \sum\mathrm{E}^2[x_i - \bar{x}]\}$$
$$\approx -2\{\beta - \beta[1/S_{xx}]^*(n-1)\sigma_x^2 + [2/S_{xy}]^*\beta^2\sigma_x^2\}[S_{xy}]^*$$
$$+ \{\beta^2 - 2\beta^2[1/S_{xx}]^*(n-3)\sigma_x^2\}\cdot[(n-1)\sigma_x^2 + S_{xx}]^*$$
$$= -2\beta[S_{xy}]^* + 2[\beta/S_{xx}]^*(n-1)[S_{xy}]^*\sigma_x^2 - 4\beta^2\sigma_x^2 + (n-1)\beta^2\sigma_x^2 + \beta^2[S_{xx}]^*$$
$$- 2\beta^2[1/S_{xx}]^*(n-3)(n-1)\sigma_x^4 - 2[\beta^2/S_{xx}]^*(n-3)[S_{xx}]^*\sigma_x^2$$
$$\approx -2\beta[S_{xy}]^* + (n-1)\beta^2\sigma_x^2 + \beta^2[S_{xx}]^*$$

In the previous calculation, $2\beta^2[1/S_{xx}]^*(n-3)(n-1)\sigma_x^4$ was neglected.

$$\therefore \mathrm{E}[SSE] = \mathrm{E}[\sum(y_i - \bar{y})^2] - 2\mathrm{E}[\hat{\beta}]\cdot\mathrm{E}[\sum(x_i - \bar{x})y_i] - 2\mathrm{cov}[\hat{\beta}, \sum(x_i - \bar{x})y_i]$$
$$+ \mathrm{V}[\hat{\beta}]\cdot\{\sum\mathrm{V}[x_i - \bar{x}] + \sum\mathrm{E}^2[x_i - \bar{x}]\} + \mathrm{E}^2[\hat{\beta}]\cdot\{\sum\mathrm{V}[x_i - \bar{x}] + \sum\mathrm{E}^2[x_i - \bar{x}]\}$$
$$+ \mathrm{cov}[\hat{\beta}^2, \sum(x_i - \bar{x})^2]$$
$$\approx [S_{yy}]^* - 2[S_{yy}S_{xx}/S_{xy}^2]^*\beta^2\sigma_x^2 + 4\beta^2\sigma_x^2 + (n-1)[S_{yy}/S_{xy}^2]^*\sigma_x^4\beta^2$$
$$+ [S_{yy}/S_{xy}^2]^*\sigma_x^2\beta^2[S_{xx}]^* - 2\beta[S_{xy}]^* + (n-1)\beta^2\sigma_x^2$$
$$+ \beta^2[S_{xx}]^* - 4\beta^2\sigma_x^2$$

$$\approx \{(n-1) - [S_{yy}S_{xx}/S_{xy}^2]^*\}\sigma_x^2\beta^2 \qquad (\because\ [S_{yy}S_{xx}/S_{xy}^2]^* = 1)$$
$$= (n-2)\sigma^2$$

In the previous calculation, $(n - 1)[S_{yy}/S_{xy}^2]^*\sigma_x^4\beta^2$ is an error term of order higher than σ_x^2 and it was ignored.

$$MSE = \sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2/(n-2)$$
$$\therefore\ \mathrm{E}[MSE] \approx \sigma^2$$

A.5 Correlation coefficient between $\hat{\beta}$ and *SSE* (or *MSE*)

$$K = \hat{\beta}\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$
$$= (S_{xx}S_{yy}S_{xy} - S_{xy}^3)/S_{xx}^2$$

Let $A = \hat{\beta} = S_{xy}/S_{xx}$, $F = \sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = (S_{xx}S_{yy} - S_{xy}^2)/S_{xx}$

$$\sigma_K^2 \approx \sum(\partial K/\partial x_i)^2\sigma_{xi}^2$$
$$\sigma_K^2 \approx (\partial K/\partial A)^2\sigma_A^2 + (\partial K/\partial F)^2\sigma_F^2 + 2(\partial K/\partial A)(\partial K/\partial F)\sigma_A\sigma_F\, r(A, F)$$

$$\therefore\ r(A, F) = r(\hat{\beta}, \sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2) = r(\hat{\beta}, SSE)$$
$$\approx -\{(S_{xx}S_{yy}S_{xy} - S_{xy}^3)/S_{xx}\}\{S_{xx}/(S_{xx}S_{yy}^2S_{xy}^2 - S_{yy}S_{xy}^4)\}^{1/2}$$
$$= -\{1 - r^2(x, y)\}^{1/2} \approx 0 \qquad \text{(A6)}$$

where $r(x, y)$ is the correlation coefficient between x and y, i.e., $r(x, y) = S_{xy}/(S_{xx}S_{yy})^{1/2}$.

Note: The full derivation of Eq. (A6) is rather long. The hand-written full derivation in PDF format may be provided through an email by the corresponding author on request.

A.6 Variance of predicted *y* value

$$\hat{y} = \hat{\alpha} + \hat{\beta}x = \bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x$$
$$\hat{y} = \bar{y} - \{\sum(x_i - \bar{x})\, y_i/\sum(x_i - \bar{x})^2\}(\sum x_i/n) + \{\sum(x_i - \bar{x})\, y_i/\sum(x_i - \bar{x})^2\}x$$

$$\{\sigma(\hat{y})\}^2 \approx \sum(\partial\hat{y}/\partial x_i)^2\sigma_{xi}^2$$
$$= \{1/n + \bar{x}^2(S_{yy}/S_{xy}^2)\}\,\hat{\beta}^2\sigma_x^2 + x^2(S_{yy}/S_{xy}^2)\,\hat{\beta}^2\sigma_x^2$$
$$- 2x(\bar{x})(S_{yy}/S_{xy}^2)\hat{\beta}^2\sigma_x^2$$
$$= [1/n + (x - \bar{x})^2(S_{yy}/S_{xy}^2)]^*[\,\hat{\beta}^2]^*\sigma_x^2 \qquad \text{(A7)}$$

$$\therefore\ \mathrm{V}[\hat{y}] \approx [1/n + (x - \bar{x})^2(S_{yy}/S_{xy}^2)]^*\sigma^2$$

A.7 Correlation coefficient between slope and intercept

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

$$\{\sigma(\hat{y})\}^2 \approx (\partial\hat{y}/\partial\hat{\alpha})^2\{\sigma(\hat{\alpha})\}^2 + (\partial\hat{y}/\partial\hat{\beta})^2\{\sigma(\hat{\beta})\}^2 + 2(\partial\hat{y}/\partial\hat{\alpha})(\partial\hat{y}/\partial\hat{\beta})\{\sigma(\hat{\alpha})\}\{\sigma(\hat{\beta})\}r(\hat{\alpha}, \hat{\beta})$$
$$\approx \{1/n + (\bar{x})^2(S_{yy}/S_{xy}^2)\}\hat{\beta}^2\sigma_x^2 + x^2(S_{yy}/S_{xy}^2)\,\hat{\beta}^2\sigma_x^2$$
$$+ 2x\{(S_{yy}/S_{xy}^2)^{1/2}\hat{\beta}\sigma_x\}\{1/n + (\bar{x})^2(S_{yy}/S_{xy}^2)\}^{1/2}\hat{\beta}\sigma_x\, r(\hat{\alpha}, \hat{\beta}) \qquad \text{(A8)}$$

From Eq. (A7) and (A8), the correlation coefficient $r(\hat{\alpha},\ \hat{\beta})$ is as follows:

$$\therefore\ r(\hat{\alpha},\ \hat{\beta}) \approx -(\bar{x})(S_{yy})^{1/2}/(S_{xy})[1/n + (\bar{x})^2(S_{yy}/S_{xy}^{2})]^{1/2}$$