"A note on the robust interpretation of regression coefficients"

by Wong M.Y. and Cox D.R. Test, Vol. 7, pp. 289-294.

1. For the multiple regression of $y_i = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \varepsilon_i$ where $\varepsilon \sim N(0, \sigma^2)$, the least squares estimates of β_1 and β_2 are

$$\hat{\beta}_1 = \frac{S_{02}\, S_{1y} - S_{11}\, S_{2y}}{S_{20}\, S_{02} - S_{11}^2}$$

and

$$\hat{\beta}_2 = \frac{S_{20}\, S_{2y} - S_{11}\, S_{1y}}{S_{20}\, S_{02} - S_{11}^2}$$

respectively, where

$$S_{rs} = \frac{1}{n} \sum_{i=1}^{n} (x_{1i} - \bar{x_1})^r \, (x_{2i} - \bar{x_2})^s \,,$$

$$S_{1y} = \frac{1}{n} \sum_{i=1}^{n} (x_{1i} - \bar{x_1})\, (y_i - \bar{y})$$

and

$$S_{2y} = \frac{1}{n} \sum_{i=1}^{n} (x_{2i} - \bar{x_2})\, (y_i - \bar{y})\,.$$

Under the "true" model, i.e. $\log y_i = \gamma_0 + \gamma_1 x_{1_i} + \gamma_2 x_{2_i} + \eta_i$ where $\eta \sim N(0, \tau^2)$,

$$\begin{aligned} E(\hat{\beta}_1) &= \frac{S_{02}}{S_{20}\, S_{02} - S_{11}^2}\, E(S_{1y}) - \frac{S_{11}}{S_{20}\, S_{02} - S_{11}^2}\, E(S_{2y}) \\ &\approx e^{\tau^2/2}\, e^{\gamma_0 + \gamma_1 \bar{x}_1 + \gamma_2 \bar{x}_2} \Big[\gamma_1 + \frac{1}{2\,(S_{20}\, S_{02} - S_{11}^2)} \Big\{\gamma_1^2\, (S_{02}\, S_{30} - S_{11}\, S_{21}) + \\ &\quad 2\, \gamma_1\, \gamma_2\, (S_{02}\, S_{21} - S_{11}\, S_{12}) + \gamma_2^2\, (S_{02}\, S_{12} - S_{11}\, S_{03})\Big\}\Big] \end{aligned}$$

and

$$\begin{aligned} E(\hat{\beta}_2) &= \frac{S_{20}}{S_{20}\, S_{02} - S_{11}^2}\, E(S_{2y}) - \frac{S_{11}}{S_{20}\, S_{02} - S_{11}^2}\, E(S_{1y}) \\ &\approx e^{\tau^2/2}\, e^{\gamma_0 + \gamma_1 \bar{x}_1 + \gamma_2 \bar{x}_2} \Big[\gamma_2 + \frac{1}{2\,(S_{20}\, S_{02} - S_{11}^2)} \Big\{\gamma_1^2\, (S_{20}\, S_{21} - S_{11}\, S_{30}) + \\ &\quad 2\, \gamma_1\, \gamma_2\, (S_{20}\, S_{12} - S_{11}\, S_{21}) + \gamma_2^2\, (S_{20}\, S_{03} - S_{11}\, S_{12})\Big\}\Big]\,. \end{aligned}$$

Therefore,

$$
\begin{aligned}
E\left(\frac{\hat\beta_1}{\hat\beta_2}\right) &= \frac{E(\hat\beta_1)}{E(\hat\beta_2)} + O\left(\frac{1}{n}\right) \\
&\approx \frac{\gamma_1}{\gamma_2} + \frac{\gamma_2}{2\,(S_{20}\,S_{02}\,-\,S_{11}^2)}\Big\{(S_{11}\,S_{30}\,-\,S_{20}\,S_{21})\left(\frac{\gamma_1}{\gamma_2}\right)^3 + (S_{02}\,S_{30}\,+ \\
&\quad S_{11}\,S_{21}\,-\,2\,S_{20}\,S_{12})\left(\frac{\gamma_1}{\gamma_2}\right)^2 - (S_{20}\,S_{03}\,+\,S_{11}\,S_{12}\,-\,2\,S_{02}\,S_{21})\left(\frac{\gamma_1}{\gamma_2}\right) + \\
&\quad (S_{02}\,S_{12}\,-\,S_{11}\,S_{03})\Big\} + O\left(\frac{1}{n}\right).
\end{aligned}
$$

The estimated ratio of standardized regression coefficients can be obtained correspondingly.

2. Denote $\hat\beta_1$ and $\hat\beta_2$ to be the least squares estimates of β_1 and β_2 for the model of $y_i = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \varepsilon_i$. The variances of $\hat\beta_1$ and $\hat\beta_2$ and their covariance under the "true" model of $\log y_i = \gamma_0 + \gamma_1 x_{1_i} + \gamma_2 x_{2_i} + \eta_i$ are obtained by Taylor expansion. They are shown as follows:

$$
\begin{aligned}
Var(\hat\beta_1) &\approx \frac{(e^{2\tau^2} - e^{\tau^2})\,e^{2(\gamma_0+\gamma_1\bar x_1+\gamma_2\bar x_2)}}{n\,(S_{20}\,S_{02}\,-\,S_{11}^2)^2}\Big\{S_{02}\,(S_{20}\,S_{02}\,-\,S_{11}^2)\,+\,2\,\gamma_1\,(S_{02}^2\,S_{30}\,+ \\
&\quad S_{11}^2 S_{12}\,-\,2\,S_{11}\,S_{02}\,S_{21}) + 2\,\gamma_2\,(S_{02}^2\,S_{21}\,+\,S_{11}^2\,S_{03}\,-\,2\,S_{11}\,S_{02}\,S_{12}) \\
&\quad + 2\,\gamma_1^2\,(S_{02}^2\,S_{40}\,+\,S_{11}^2\,S_{22}\,-\,2\,S_{11}\,S_{02}\,S_{31})\,+\,4\,\gamma_1\,\gamma_2\,(S_{02}^2\,S_{31}\,+ \\
&\quad S_{11}^2\,S_{13}\,-\,2\,S_{11}\,S_{02}\,S_{22})\,+\,2\,\gamma_2^2\,(S_{02}^2\,S_{22}\,+\,S_{11}^2 S_{04}\,-\,2\,S_{11}\,S_{02}\,S_{13})\Big\},
\end{aligned}
$$

$$
\begin{aligned}
Var(\hat\beta_2) &\approx \frac{(e^{2\tau^2} - e^{\tau^2})\,e^{2(\gamma_0+\gamma_1\bar x_1+\gamma_2\bar x_2)}}{n\,(S_{20}\,S_{02}\,-\,S_{11}^2)^2}\Big\{S_{20}\,(S_{20}\,S_{02}\,-\,S_{11}^2)\,+\,2\,\gamma_1\,(S_{11}^2\,S_{30}\,+ \\
&\quad S_{20}^2 S_{12}\,-\,2\,S_{20}\,S_{11}\,S_{21}) + 2\,\gamma_2\,(S_{11}^2\,S_{21}\,+\,S_{20}^2\,S_{03}\,-\,2\,S_{20}\,S_{11}\,S_{12}) \\
&\quad + 2\,\gamma_1^2\,(S_{11}^2\,S_{40}\,+\,S_{20}^2\,S_{22}\,-\,2\,S_{20}\,S_{11}\,S_{31})\,+\,4\,\gamma_1\,\gamma_2\,(S_{11}^2\,S_{31}\,+ \\
&\quad S_{20}^2\,S_{13}\,-\,2\,S_{20}\,S_{11}\,S_{22})\,+\,2\,\gamma_2^2\,(S_{11}^2\,S_{22}\,+\,S_{20}^2 S_{04}\,-\,2\,S_{20}\,S_{11}\,S_{13})\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
Cov(\hat\beta_1, \hat\beta_2) &\approx -\frac{(e^{2\tau^2} - e^{\tau^2})\,e^{2(\gamma_0+\gamma_1\bar x_1+\gamma_2\bar x_2)}}{n\,(S_{20}\,S_{02}\,-\,S_{11}^2)^2}\Big\{S_{11}\,(S_{20}\,S_{02}\,-\,S_{11}^2)\,+ \\
&\quad 2\,\gamma_1\,[S_{11}\,(S_{02}\,S_{30}\,+\,S_{20} S_{12})\,-\,S_{21}\,(S_{20}\,S_{02}\,+S_{11}^2)]\,+
\end{aligned}
$$

$$2\,\gamma_2\,[S_{11}\,(S_{02}\,S_{21}\;+\;S_{20}\,S_{03})\;-\;S_{12}\,(S_{20}\,S_{02}\;+\;S_{11})]$$
$$+\,2\;\gamma_1^2\,[S_{11}\,(S_{02}\,S_{40}\;+\;S_{20}\,S_{22})\;-\;S_{31}\,(S_{20}\,S_{02}\;+S_{11}^2)]\;+$$
$$4\,\gamma_1\,\gamma_2\,[S_{11}\,(S_{02}\,S_{31}\;+\;S_{20}\,S_{13})\;-\;S_{22}\,(S_{20}\,S_{02}\;+S_{11}^2)]\;+$$
$$2\;\gamma_2^2\,[S_{11}\,(S_{02}\,S_{22}\;+\;S_{20}S_{04})\;-\;S_{13}\,(S_{20}\,S_{02}\;+S_{11}^2)]\Big\}\,,$$

where τ^2 is the error variance in the "true" model.

The approximate formula for $Var(\hat\beta_1/\hat\beta_2)$ obtained by expanding in a Taylor series is equal to

$$Var\left[\frac{\hat\beta_1}{\hat\beta_2}\right]\;=\;\left(\frac{E(\hat\beta_1)}{E(\hat\beta_2)}\right)^2\;\left(\frac{Var(\hat\beta_1)}{E(\hat\beta_1)^2}\;+\;\frac{Var(\hat\beta_2)}{E(\hat\beta_2)^2}\;-\;\frac{2\,Cov(\hat\beta_1,\hat\beta_2}{E(\hat\beta_1)\,E(\hat\beta_2)}\right)\,,$$

where $E(\hat\beta_1)$ and $E(\hat\beta_2)$ are the expectations of $\hat\beta_1$ and $\hat\beta_2$ under the true model. The variance of the estimated ratio of regression coefficients β_1 and β_2 under the "true" model is approximately equal to

$$\frac{e^{\tau^2}\;-\;1}{n\,\gamma_2^4\,(S_{20}\,S_{02}\;-\;S_{11}^2)}\,(S_{20}\,\gamma_1^2\;+\;2\,S_{11}\,\gamma_1\,\gamma_2\;+\;S_{02}\,\gamma_2^2)\;+\;\frac{e^{\tau^2}\;-\;1}{n\,\gamma_2^4\,(S_{20}\,S_{02}\;-\;S_{11}^2)^2}$$
$$\Big[\gamma_1^3\,\{S_{30}\,(S_{20}\,S_{02}\;+\;S_{11}^2)\;-\;4\,S_{21}\,S_{20}\,S_{11}\;+\;2\,S_{12}\,S_{20}^2\}\;+\;\gamma_1^2\,\gamma_2\,\{4\,S_{30}\,S_{11}\,S_{02}\;-$$
$$S_{21}\,(S_{20}\,S_{02}\;+\;5\,S_{11}^2)\;+\;2\,S_{03}\,S_{20}^2\}\;+\;\gamma_1\,\gamma_2^2\,\{2\,S_{30}\,S_{02}^2\;-\;S_{12}\,(S_{20}\,S_{02}\;+\;5\,S_{11}^2)\;+$$
$$4\,S_{03}\,S_{20}\,S_{11}\}\;+\;\gamma_2^3\,\{2\,S_{21}\,S_{02}^2\;-\;4\,S_{12}\,S_{11}\,S_{02}\;+\;S_{03}\,(S_{20}\,S_{02}\;+\;S_{11}^2)\}\Big]\;+$$
$$\frac{e^{\tau^2}\;-\;1}{12\,n\,\gamma_2^4\,(S_{20}\,S_{02}\;-\;S_{11}^2)^2}\,\Big[\gamma_1^4\,\{4\,S_{40}\,(S_{20}\,S_{02}\;+\;5\,S_{11}^2)\;-\;48\,S_{31}\,S_{20}\,S_{11}\;+$$
$$24\,S_{22}\,S_{20}^2\;+\;3\,S_{30}^2\,S_{02}\;-\;6\,S_{30}\,S_{21}\,S_{11}\;+\;24\,S_{30}\,S_{12}\,S_{20}\;-\;21\,S_{21}^2\,S_{20}\}\;+$$
$$4\,\gamma_1^3\,\gamma_2\{12\,S_{40}\,S_{11}\,S_{02}\;-\;4\,S_{31}\,(2\,S_{20}\,S_{02}\;+\;S_{11}^2)\;-\;12\,S_{22}\,S_{20}\,S_{11}\;+\;12\,S_{13}\,S_{20}^2\;+$$
$$3\,S_{30}\,S_{21}\,S_{02}\;+\;9\,S_{30}\,S_{12}\,S_{11}\;+\;6\,S_{30}\,S_{03}\,S_{20}\;-\;15\,S_{21}^2\,S_{11}\;-\;3\,S_{21}\,S_{12}\,S_{20}\}\;+$$
$$6\gamma_1^2\gamma_2^2\{4\,S_{40}\,S_{02}^2\;+\;8\,S_{31}\,S_{11}\,S_{02}\;-\;12\,S_{22}\,(S_{20}\,S_{02}\;+\;S_{11}^2)\;+\;8\,S_{13}\,S_{20}\,S_{11}\;+$$
$$4\,S_{04}\,S_{20}^2\;+\;5\,S_{30}\,S_{12}\,S_{02}\;+\;7\,S_{30}\,S_{03}\,S_{11}\;-\;2\,S_{21}^2\,S_{02}\;-\;13\,S_{21}\,S_{12}\,S_{11}\;+$$
$$5\,S_{21}\,S_{03}\,S_{20}\;-\;2\,S_{12}^2\,S_{20}\}\;+\;4\,\gamma_1\,\gamma_2^3\{12\,S_{31}\,S_{02}^2\;-\;12\,S_{22}\,S_{11}\,S_{02}\;-\;4\,S_{13}\,(2\,S_{20}\,S_{02}\;+$$
$$S_{11}^2)\;+\;12\,S_{04}\,S_{20}\,S_{11}\;+\;6\,S_{30}\,S_{03}\,S_{02}\;-\;3\,S_{21}\,S_{12}\,S_{02}\;+\;9\,S_{21}\,S_{03}\,S_{11}\;-$$
$$15\,S_{12}^2\,S_{11}\;+\;3\,S_{12}\,S_{03}\,S_{20}\}\;+\;\gamma_2^4\{24\,S_{22}\,S_{02}^2\;-\;48\,S_{13}\,S_{11}\,S_{02}\;+\;4\,S_{04}\,(S_{20}\,S_{02}\;+$$
$$5\,S_{11}^2)\;+\;24\,S_{21}\,S_{03}\,S_{02}\;-\;21\,S_{12}^2\,S_{02}\;-\;6\,S_{12}\,S_{03}\,S_{11}\;+\;3\,S_{03}^2\,S_{20}\}\Big]$$

The estimated variance of the ratio of the least squares estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ for the multiple regression of $y_i = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \varepsilon_i$ where $\varepsilon \sim N(0, \sigma^2)$ is approximately equal to

$$\frac{\hat{\sigma}^2}{n\,\hat{\beta}_2^4} \left\{ \hat{\beta}_1^2 \frac{S_{20}}{S_{20}\,S_{02} - S_{11}^2} + 2\,\hat{\beta}_1\,\hat{\beta}_2 \frac{S_{11}}{S_{20}\,S_{02} - S_{11}^2} + \hat{\beta}_2^2 \frac{S_{02}}{S_{20}\,S_{02} - S_{11}^2} \right\},$$

where $\hat{\sigma}^2$ is the estimated error variance of the fitted model. Substitute expectations of $\hat{\sigma}^2$, $\hat{\beta}_1^2$, $\hat{\beta}_2^2$ and $\hat{\beta}_2^4$ under the "true" model into the above expression, the expectation under the "true" model of the estimated variance calculated from the fitted model is

$$\frac{e^{\tau^2} - 1}{n\,\gamma_2^4\,(S_{20}\,S_{02} - S_{11}^2)} (S_{20}\,\gamma_1^2 + 2\,S_{11}\,\gamma_1\,\gamma_2 + S_{02}\,\gamma_2^2) + \frac{e^{\tau^2} - 1}{n\,\gamma_2^4\,(S_{20}\,S_{02} - S_{11}^2)^2}$$
$$(S_{30}\,\gamma_1^3 + 3\,S_{21}\,\gamma_1^2\,\gamma_2 + 3\,S_{12}\,\gamma_1\,\gamma_2^2 + S_{03}\,\gamma_2^3) + \frac{e^{\tau^2} - 1}{12\,n\,\gamma_2^4\,(S_{20}\,S_{02} - S_{11}^2)^2}$$
$$\Big[\gamma_1^4\,\{4\,(S_{40} + 6\,S_{20}^2)\,(S_{20}\,S_{02} - S_{11}^2) + 3\,S_{30}^2\,S_{02} - 6\,S_{30}\,S_{21}\,S_{11} + 3\,S_{21}^2\,S_{20}\} +$$
$$4\,\gamma_1^3\,\gamma_2\,\{4\,(S_{31} + 6\,S_{20}\,S_{11})\,(S_{20}\,S_{02} - S_{11}^2) + 3\,S_{30}\,S_{21}\,S_{02} - 3\,S_{30}\,S_{12}\,S_{11} -$$
$$3\,S_{21}^2\,S_{11} + 3\,S_{21}\,S_{12}\,S_{20}\} + 6\,\gamma_1^2\,\gamma_2^2\,\{4\,(S_{22} + 2\,(2\,S_{11}^2 + S_{20}\,S_{02}))$$
$$(S_{20}\,S_{02} - S_{11}^2) + S_{30}\,S_{12}\,S_{02} - S_{30}\,S_{03}\,S_{11} + 2\,S_{21}^2\,S_{02} - 5\,S_{21}\,S_{12}\,S_{11} +$$
$$S_{21}\,S_{03}\,S_{20} + 2\,S_{12}^2\,S_{20}\} + 4\,\gamma_1\,\gamma_2^3\,\{4\,(S_{13} + 6\,S_{11}\,S_{02})\,(S_{20}\,S_{02} - S_{11}^2) +$$
$$3\,S_{21}\,S_{12}\,S_{02} - 3\,S_{21}\,S_{03}\,S_{11} - 3\,S_{12}^2\,S_{11} + 3\,S_{12}\,S_{03}\,S_{20}\} + \gamma_4^2\{4\,(S_{04} + 6\,S_{02}^2)$$
$$(S_{20}\,S_{02} - S_{11}^2) + 3\,S_{12}^2\,S_{02} - 6\,S_{12}\,S_{03}\,S_{11} + 3\,S_{03}^2\,S_{20}\}\Big].$$