

CORRECTION

Open Access



# Correction to: Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method

Andres Legarra<sup>1\*</sup> and Antonio Reverter<sup>2</sup>

## Correction to: *Genet Sel Evol* (2018) 50:53

<https://doi.org/10.1186/s12711-018-0426-6>

After publication of original article [1], the authors noticed that there was an error.

In the section “Quadratic forms of estimated breeding values” we wrote, based on Taylor series approximations, that:

“In the remainder of this paper, we assume that the expectation of a ratio of quadratic forms is equal to the ratio of the expectations. The “Appendix” shows that this holds when the number of individuals included in the statistics is large (several hundred or more) or when they are not structured into very large sibships. Otherwise, as shown in the “Appendix”, both the true regression coefficient  $b = cov(\hat{\mathbf{u}}_p, \mathbf{u})/var(\hat{\mathbf{u}}_p)$  and its estimator  $\hat{b} = cov(\hat{\mathbf{u}}_p, \hat{\mathbf{u}}_w)/var(\hat{\mathbf{u}}_p)$  have an expectation less than 1, even when the model is perfect and the EBV have the right dispersion.”

This statement is wrong, because a further expansion of the Taylor series (until its 3rd term) shows that the expectation of both  $b$  and  $\hat{b}$  is 1. The correct paragraph should be:

“In the remainder of this paper, we assume that the expectation of a ratio of quadratic forms is equal to the ratio of the expectations. The “Correction” below

shows that this holds (up to the second order of the Taylor series expansion) when the BLUP properties hold, as assumed throughout the paper.”

Moreover, the Abstract reads “Contrary to common belief, the regression of true on estimated breeding values is (on expectation) lower than 1 for small or related validation sets, due to family structures.” whereas it should read “Up to the second term of a Taylor series expansion, the regression of true on estimated breeding values is (on expectation) equal to 1, even for small or related validation sets.”

Further proof (equivalent to the Appendix) is below. According to [2] the second-order Taylor series expansion of a ratio is  $E\left(\frac{X}{Y}\right) \approx \frac{E(X)}{E(Y)} - \frac{Cov(X,Y)}{E(Y)^2} + \frac{Var(Y)E(X)}{E(Y)^3}$ . Consider the bias in the estimation of the ratio  $b = \frac{cov(\hat{\mathbf{u}}_p, \mathbf{u})}{var(\hat{\mathbf{u}}_p)}$  using the estimator  $\hat{b} = \frac{cov(\hat{\mathbf{u}}_p, \hat{\mathbf{u}}_w)}{var(\hat{\mathbf{u}}_p)}$ . In our case:

$$\hat{b}_{w,p} = \frac{cov(\hat{\mathbf{u}}_p, \hat{\mathbf{u}}_w)}{var(\hat{\mathbf{u}}_p)} = \frac{\frac{1}{n}(\hat{\mathbf{u}}_p - \bar{\hat{\mathbf{u}}}_p)'(\hat{\mathbf{u}}_w - \bar{\hat{\mathbf{u}}}_w)}{\frac{1}{n}(\hat{\mathbf{u}}_p - \bar{\hat{\mathbf{u}}}_p)'(\hat{\mathbf{u}}_p - \bar{\hat{\mathbf{u}}}_p)}$$

Therefore, we expand  $E\left(\frac{X}{Y}\right)$  for  $= \frac{1}{n}(\hat{\mathbf{u}}_p - \bar{\hat{\mathbf{u}}}_p)'(\hat{\mathbf{u}}_w - \bar{\hat{\mathbf{u}}}_w) = \frac{1}{n}(\hat{\mathbf{u}}_p' \mathbf{S}' \hat{\mathbf{S}} \hat{\mathbf{u}}_w) = \frac{1}{n}(\hat{\mathbf{u}}_p \mathbf{S} \hat{\mathbf{u}}_w)$ , where  $\mathbf{S} = \mathbf{I} - \frac{1}{n} \mathbf{J}$  and  $Y = \frac{1}{n}(\hat{\mathbf{u}}_p - \bar{\hat{\mathbf{u}}}_p)'(\hat{\mathbf{u}}_p - \hat{\mathbf{u}}_p) = \frac{1}{n}(\hat{\mathbf{u}}_p \mathbf{S} \hat{\mathbf{u}}_p)$ . The expression for the covariance of bilinear forms (under normality) is:

The original article can be found online at <https://doi.org/10.1186/s12711-018-0426-6>.

\*Correspondence: andres.legarra@inra.fr

<sup>1</sup> INRA, UMR1388 GenPhySE, 31326 Castanet-Tolosan, France

Full list of author information is available at the end of the article



$$\begin{aligned} &Cov(\mathbf{x}'_1 \mathbf{A}_{12} \mathbf{x}_2, \mathbf{x}_3 \mathbf{A}_{34} \mathbf{x}_4) \\ &= tr(\mathbf{A}_{12} \mathbf{C}_{23} \mathbf{A}_{34} \mathbf{C}_{41} + \mathbf{A}_{12} \mathbf{C}_{24} \mathbf{A}_{43} \mathbf{C}_{31}) \\ &+ \mu'_1 \mathbf{A}_{12} \mathbf{C}_{23} \mathbf{A}_{34} \mu_4 + \mu'_1 \mathbf{A}_{12} \mathbf{C}_{24} \mathbf{A}_{43} \mu_3 \\ &+ \mu'_2 \mathbf{A}_{21} \mathbf{C}_{13} \mathbf{A}_{34} \mu_4 + \mu'_2 \mathbf{A}_{21} \mathbf{C}_{14} \mathbf{A}_{43} \mu_3 \end{aligned}$$

(**C** is the covariance matrix across  $\mathbf{x}_i$ ; chapter 2 Equation 58 in [3]). Applied to our case, this yields:

$$Cov(X, Y) = \frac{2}{n^2} tr(\mathbf{S}(\mathbf{G} - \mathbf{C}_p^{uu}) \mathbf{S}(\mathbf{G} - \mathbf{C}_p^{uu})).$$

The terms linked to the means disappear, as before, because they have the form  $\mu \mathbf{1}' \mathbf{S}'(\mathbf{G} - \mathbf{C}_p^{uu}) \mathbf{S} \mathbf{1} \mu$ , which has a value of 0. Then, we have:

$$\begin{aligned} E(Y) &= E\left(\frac{1}{n}(\hat{\mathbf{u}}_p - \bar{\mathbf{u}}_p)'(\hat{\mathbf{u}}_p - \bar{\mathbf{u}}_p)\right) \\ &= \frac{1}{n} tr(\mathbf{S}(\mathbf{G} - \mathbf{C}_p^{uu})), \\ Var(Y) &= Cov(Y, Y) \\ &= \frac{2}{n^2} tr(\mathbf{S}(\mathbf{G} - \mathbf{C}_p^{uu}) \mathbf{S}(\mathbf{G} - \mathbf{C}_p^{uu})), \\ E(X) &= E\left(\frac{1}{n}(\hat{\mathbf{u}}_p - \bar{\mathbf{u}}_p)'(\hat{\mathbf{u}}_w - \bar{\mathbf{u}}_w)\right) \\ &= \frac{1}{n} tr(\mathbf{S}(\mathbf{G} - \mathbf{C}_p^{uu})), \end{aligned}$$

which are all the elements needed. To simplify notation, consider  $\mathbf{K} = \mathbf{G} - \mathbf{C}_p^{uu}$ . Putting all together gives:

$$\begin{aligned} &-\frac{Cov(X, Y)}{E(Y)^2} + \frac{Var(Y)E(X)}{E(Y)^3} \\ &= -\frac{\frac{2}{n^2} tr(\mathbf{SKSK})}{\left(\frac{1}{n} tr(\mathbf{SK})\right)^2} + \frac{\frac{2}{n^2} tr(\mathbf{SKSK}) \frac{1}{n} tr(\mathbf{SK})}{\left(\frac{1}{n} tr(\mathbf{SK})\right)^3} \\ &= -\frac{2tr(\mathbf{SKSK})}{(tr(\mathbf{SK}))^2} + \frac{2tr(\mathbf{SKSK})tr(\mathbf{SK})}{(tr(\mathbf{SK}))^3} \end{aligned}$$

which cancels out. Therefore, up to the 3rd order approximation, the bias in the estimation of  $b = \frac{cov(\hat{\mathbf{u}}_p, \mathbf{u})}{var(\hat{\mathbf{u}}_p)}$  using the estimator  $\hat{b} = \frac{cov(\hat{\mathbf{u}}_p, \hat{\mathbf{u}}_w)}{var(\hat{\mathbf{u}}_p)}$  is 0. In addition, this shows that the expected value of  $b = \frac{cov(\hat{\mathbf{u}}_p, \mathbf{u})}{var(\hat{\mathbf{u}}_p)}$  is also 1 since the same derivation as above holds with  $\mathbf{u}$  in the place of  $\mathbf{u}_w$ .

**Acknowledgements**

The authors thank Peter Sullivan (Lactanet, Guelph, ON, Canada) for pointing out that a second order expansion should give a ratio of 1.

**Author details**

<sup>1</sup> INRA, UMR1388 GenPhySE, 31326 Castanet-Tolosan, France. <sup>2</sup> CSIRO Agriculture and Food, 306 Carmody Rd., St. Lucia, QLD 4067, Australia.

Published online: 22 November 2019

**References**

- Legarra A, Reverter A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet Sel Evol*. 2018;50:53. <https://doi.org/10.1186/s12711-018-0426-6>.
- Stuart A, Ord K. Kendall's advanced theory of statistics. London: Arnold Hodder Headline Group; 1998.
- Searle S. Linear models. New York: John Wiley; 1971.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

