We Need to Talk about Interactions

F. Rosales

The following is trivial and perhaps not all that rigorous. Regardless, I hope the argument is clear.

1 OLS

Consider the usual model

$$\boldsymbol{Y} = \alpha_0\mathbf{1} + \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \alpha_3\boldsymbol{x}_3, \quad \boldsymbol{x}_3 = \boldsymbol{x}_1 \circ \boldsymbol{x}_2 \tag{1}$$

and its centred version

$$\begin{aligned} \boldsymbol{Y} &= \beta_0\mathbf{1} + \beta_1\tilde{\boldsymbol{x}}_1 + \beta_2\tilde{\boldsymbol{x}}_2 + \beta_3\tilde{\boldsymbol{x}}_3 \\ &= \beta_0\mathbf{1} + \beta_1(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}_1) + \beta_2(\boldsymbol{x}_2 - \bar{\boldsymbol{x}}_2) + \beta_3\{(\boldsymbol{x}_1 \circ \boldsymbol{x}_2) - \overline{(\boldsymbol{x}_1 \circ \boldsymbol{x}_2)}\} \\ &= \underbrace{\{\beta_0 - \beta_1\bar{\boldsymbol{x}}_1 - \beta_2\bar{\boldsymbol{x}}_2 - \beta_3\overline{(\boldsymbol{x}_1 \circ \boldsymbol{x}_2)}\}}_{\text{const.}}\mathbf{1} + \beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 + \beta_3\boldsymbol{x}_3, \end{aligned} \tag{2}$$

which has the same estimator as (1), except for the estimator for the intercept. Now consider the alternative model

$$\begin{aligned} \boldsymbol{Y} &= \gamma_0\mathbf{1} + \gamma_1\tilde{\boldsymbol{x}}_1 + \gamma_2\tilde{\boldsymbol{x}}_2 + \gamma_3\tilde{\tilde{\boldsymbol{x}}}_3 \\ &= \gamma_0\mathbf{1} + \gamma_1(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}_1) + \gamma_2(\boldsymbol{x}_2 - \bar{\boldsymbol{x}}_2) + \gamma_3\{(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}_1) \circ (\boldsymbol{x}_2 - \bar{\boldsymbol{x}}_2)\} \\ &= \underbrace{(\gamma_0 - \gamma_1\bar{\boldsymbol{x}}_1 - \gamma_2\bar{\boldsymbol{x}}_2 + \gamma_3\bar{\boldsymbol{x}}_1\bar{\boldsymbol{x}}_2)}_{\text{const.}}\mathbf{1} + \underbrace{(\gamma_1 - \bar{\boldsymbol{x}}_2)}_{\text{const.}}\boldsymbol{x}_1 + \underbrace{(\gamma_2 - \bar{\boldsymbol{x}}_1)}_{\text{const.}}\boldsymbol{x}_2 + \gamma_3\boldsymbol{x}_1 \circ \boldsymbol{x}_2, \end{aligned} \tag{3}$$

which, has the same estimator for the interaction term only, and the justification of example 1 in the paper is trivial. Moreover, since models (1), (2) and (3) are the same model, just with a different parametrization, the resulting $\hat{\boldsymbol{Y}}$ should be invariant.

2 RRBLUP

Consider the problem

$$\hat{\boldsymbol{\gamma}} = \text{argmin}_{\gamma}\{\{\boldsymbol{Y} - \gamma_0\mathbf{1} - \gamma_1\boldsymbol{x}_1 - \gamma_2\boldsymbol{x}_2 - \gamma_3(\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda(\gamma_1^2 + \gamma_2^2 + \gamma_3^2)\} \tag{4}$$

$$\hat{\boldsymbol{\gamma}} = \text{argmin}_{\gamma}\{\{\boldsymbol{Y} - (\gamma_0 - \gamma_1\bar{\boldsymbol{x}}_1 - \gamma_2\bar{\boldsymbol{x}}_2 + \gamma_3\bar{\boldsymbol{x}}_1\bar{\boldsymbol{x}}_2)\mathbf{1} - (\gamma_1 - \bar{\boldsymbol{x}}_2)\boldsymbol{x}_1 - (\gamma_2 - \bar{\boldsymbol{x}}_1)\boldsymbol{x}_2 - \gamma_3(\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda(\gamma_1^2 + \gamma_2^2 + \gamma_3^2)\} \tag{5}$$

This RRBLUP problem is basically the OLS problem with a penalty on all terms except for the intercept. If one assumes that the penalty is fixed to $\lambda = 1$, as it is in example 1 of the paper, the behaviour of the estimated coefficients can be anticipated. Namely, if one extends

1. (1) to (4). The expected result is all shrinked estimators except for the intercept. In the paper the values 1.81 and 1.83 are reported.
2. (3) to (5). The expected result is, once again, to obtain all shrinked estimators except for the intercept. In this case the value reported is 0.334 in both models.
3. (4) to (5). The models are completely different.

Note however that under this reasoning $\hat{\boldsymbol{Y}}$ must be the same in both (4) and (5), which is not the case in the paper. Check empirically once again.

If instead the alternative model is considered

$$\hat{\boldsymbol{\gamma}} = \text{argmin}_{\gamma}\{\{\boldsymbol{Y} - \gamma_0\mathbf{1} - \gamma_1\boldsymbol{x}_1 - \gamma_2\boldsymbol{x}_2 - \gamma_3(\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda\gamma_3^2\} \tag{6}$$

$$\hat{\boldsymbol{\gamma}} = \text{argmin}_{\gamma}\{\{\boldsymbol{Y} - (\gamma_0 - \gamma_1\bar{\boldsymbol{x}}_1 - \gamma_2\bar{\boldsymbol{x}}_2 + \gamma_3\bar{\boldsymbol{x}}_1\bar{\boldsymbol{x}}_2)\mathbf{1} - (\gamma_1 - \bar{\boldsymbol{x}}_2)\boldsymbol{x}_1 - (\gamma_2 - \bar{\boldsymbol{x}}_1)\boldsymbol{x}_2 - \gamma_3(\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda\gamma_3^2\}, \tag{7}$$

as done in the last part of example 1. Now we can make the following comparisons

1. (4) to (6). The expect result is that the estimators $\hat{\gamma}_0$ and $\hat{\gamma}_3$ should remain the same, while all the others must change. This does not hold empirically. Check once again.
2. (5) to (7). The expect result is that the estimators $\hat{\gamma}_0$ and $\hat{\gamma}_3$ should remain the same, while all the others must change. In the paper we have 0.334 for the intercept in both models and -0.575 and -0.570 for models (5) and (7) respectively.
3. (6) to (7). The expect result is that both models produce the same estimator for $\hat{\gamma}_3$, which is verified.