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

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

Hence, Model (1), or any of its parametrizations, are solved via

$$\hat{\boldsymbol{\delta}} = \text{argmin}_{\boldsymbol{\delta}} \mathcal{V}(\boldsymbol{\delta}), \quad \mathcal{V}(\boldsymbol{\delta}) := \{\boldsymbol{Y} - \delta_0 \mathbf{1} - \delta_1 \boldsymbol{x}_1 - \delta_2 \boldsymbol{x}_2 - \delta_3 (\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2. \tag{4}$$

Since (4) has the same form for models (1)-(3), the solver leads to the same solution, so $\hat{\boldsymbol{Y}}$ must be the same in all cases. Regarding the parameters we note that

1. (1) and (2) must have exactly the same estimators for additive and interaction effects, but the intercept should be different. This is observed in the paper.
2. (1) and (3) must have different estimators for the additive effects and the intercept, but the coefficient related to interactions should be the same. This is also observed in the paper.

2 RRBLUP

We can regularize the estimated coefficients via penalizatio by

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \mathrm{argmin}_{\boldsymbol{\alpha}} \mathcal{V}(\boldsymbol{\alpha}), \\ \mathcal{V}(\boldsymbol{\alpha}) &:= \{\boldsymbol{Y} - \alpha_0 \mathbf{1} - \alpha_1 \boldsymbol{x}_1 - \alpha_2 \boldsymbol{x}_2 - \alpha_3 (\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda(\alpha_1^2 + \alpha_2^2 + \alpha_3^2) \end{aligned} \tag{5}$$

or

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= \mathrm{argmin}_{\boldsymbol{\gamma}} \mathcal{V}(\boldsymbol{\gamma}), \\ \mathcal{V}(\boldsymbol{\gamma}) &:= \{\boldsymbol{Y} - (\gamma_0 - \gamma_1 \bar{\boldsymbol{x}}_1 - \gamma_2 \bar{\boldsymbol{x}}_2 + \gamma_3 \bar{\boldsymbol{x}}_1 \bar{\boldsymbol{x}}_2)\mathbf{1} - (\gamma_1 - \bar{\boldsymbol{x}}_2)\boldsymbol{x}_1 - \\ &\quad (\gamma_2 - \bar{\boldsymbol{x}}_1)\boldsymbol{x}_2 - \gamma_3 (\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda(\gamma_1^2 + \gamma_2^2 + \gamma_3^2) \end{aligned} \tag{6}$$

These RRBLUP problems are basically the OLS problem with a penalty on all terms except for the intercept for the extensions of solvers for model (1) and (3) respectively.

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 (5). 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 (6). 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. (5) to (6). The models are completely different regarding parameter estimation.

Note however that under this reasoning $\hat{\boldsymbol{Y}}$ must be the same in both (5) and (6), which is not the case in the paper. Check empirically once again.

If instead the alternative models are considered

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \mathrm{argmin}_{\boldsymbol{\alpha}} \mathcal{V}(\boldsymbol{\alpha}), \\ \mathcal{V}(\boldsymbol{\alpha}) &:= \{\boldsymbol{Y} - \alpha_0 \mathbf{1} - \alpha_1 \boldsymbol{x}_1 - \alpha_2 \boldsymbol{x}_2 - \alpha_3 (\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda \alpha_3^2 \end{aligned} \tag{7}$$

or

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= \operatorname{argmin}_{\boldsymbol{\gamma}} \mathcal{V}(\boldsymbol{\gamma}), \\
\mathcal{V}(\boldsymbol{\gamma}) &:= \{\boldsymbol{Y} - (\gamma_0 - \gamma_1 \bar{\boldsymbol{x}}_1 - \gamma_2 \bar{\boldsymbol{x}}_2 + \gamma_3 \bar{\boldsymbol{x}}_1 \bar{\boldsymbol{x}}_2)\mathbf{1} - (\gamma_1 - \bar{\boldsymbol{x}}_2)\boldsymbol{x}_1 - \\
&\quad (\gamma_2 - \bar{\boldsymbol{x}}_1)\boldsymbol{x}_2 - \gamma_3(\boldsymbol{x}_1 \circ \boldsymbol{x}_2)\}^2 + \lambda \gamma_3^2,
\end{aligned} \tag{8}$$

as done in the last part of example 1, we observe some other things:

1. (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. This does not hold empirically. Check once again.

2. (6) to (8). 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 (6) and (8) respectively.

3. (7) to (8). The expect result is that both models produce the same estimator for $\hat{\gamma}_3$ but all the others should be different, which is verified.