Behavior of method LR to benchmark genetic evaluations

Fernando Macedo & Andrés Legarra
GenPhySE, INRA.
fernando.macedo@inra.fr
EAAP, August 2018, Dubrovnik, Croatia.
Summary

• Why another method to benchmark genetic evaluations?

• Method LR

• Simulations

• First results

• Conclusions
Why another method to benchmark genetic evaluations?

• In genomic evaluations cross validation is the most used tool for benchmarking

• All golden standard have problems:
  • Pre-corrected phenotypes may be not well corrected
  • Daughter Yield Deviations are not always available or might be inaccurate
  • Some traits (like maternal effects) don’t have direct observation related to animals

• Need simple general tools for varied situations in animal breeding systems

• Legarra & Reverter (2017) proposed a new method based on comparisons of EBV from partial (old) data vs whole (old+new) data.
  • Does not require “true” breeding values
  • Does not require pre-corrected phenotypes
Method LR

EBVs ($\hat{u}_p$) Young males without daughters

EBVs ($\hat{u}_w$) Same males with daughters

Statistics

BIAS $\mu_{wp}$
SLOPE $b_{wp}$
Correlation $\rho_{wp}$
...
Method LR: Estimators.

Bias
\[ \mu_{wp} = \bar{u}_p - \bar{u}_w . \] Expected value of 0 in absence of bias.

Slope of the regression EBVw on EBVp
\[ b_{w,p} = \frac{cov(\bar{u}_p, \bar{u}_w)}{var(u_p)} . \] With a value of 1 in unbiased procedure.

Correlation between EBVp and EBVw.
Direct estimator of relative increase of accuracy from partial to whole.
\[ \rho_{p,w} = \frac{cov(\bar{u}_w, \bar{u}_p)}{\sqrt{var(u_w)var(u_p)}} . \] The expected value is \( E(\rho_{p,w}) \approx \frac{acc_p}{acc_w} \).
Objective

Testing the estimators of bias, slope and accuracy using simulated selection schemes in several scenarios:

1. The genetic evaluation model is the correct one.
2. The genetic evaluation model is wrong.
Simulation details

Dairy sheep like scheme.
Simulation was performed with QMSim software (version 1.10) (Sargolzaei & Schenkel, 2009).

Parameters used:
• $h^2$ simulated: 0.05, 0.10, 0.25, 0.50
• 20 replicates for each $h^2$
• Records only in females
• 10 generations
• Total animals in each replicate around 500,000
• Selection by higher EBV's
The genetic evaluation model is wrong

2 Strategies:

1. Contemporary groups with phenotypic trend
   • Around 90 CG/Generation with about 500 animals each.
   • Simulated: True effect of CG as random with time trend
   • Estimated: in BLUP as fixed effect

2. Using different $h^2$ in BLUP evaluations to those used for simulation (results not shown)
   • e.g. simulated $h^2=0.10$ and evaluation $h^2=0.05$
Within each replicate:

e.g.

At the end of generation 5: estimate EBV of young males (without progeny) $\hat{u}_p$

At the end of generation 6: estimate EBV of the same males (with progeny) $\hat{u}_w$

Compute statistics:

Bias

$$\mu_{wp} = \bar{u}_p - \bar{u}_w$$

Slope

$$b_{w,p} = \frac{\text{cov} (\bar{u}_p, \bar{u}_w)}{\text{var} (\bar{u}_p)}$$

Correlation

$$\rho_{p,w} = \frac{\text{cov} (\bar{u}_w, \bar{u}_p)}{\sqrt{\text{var} (\bar{u}_w) \text{var} (\bar{u}_p)}}$$

In this work we estimate the statistics for generations 5 to 9:

<table>
<thead>
<tr>
<th>5 vs 6</th>
<th>6 vs 7</th>
<th>7 vs 8</th>
<th>8 vs 9</th>
<th>9 vs 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>p 0</td>
<td>w 0</td>
<td>p 0</td>
<td>w 0</td>
<td></td>
</tr>
<tr>
<td>5 vs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>vs</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p 0</td>
<td>w 0</td>
<td>p 0</td>
<td>w 0</td>
<td></td>
</tr>
<tr>
<td>0 vs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>vs</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
First Results

Using the correct evaluation model
Estimated $\Rightarrow \mu_{w,p} = \overline{u}_p - \overline{u}_w$

True $\Rightarrow \mu_{u,p} = \overline{u}_p - \overline{u}$

True Bias = 0.19
Estimated Bias = 0.18
replicate=14
G9 (partial) vs G10 (whole)

$h^2=0.5$
BIAS

Estimated Bias ⇒ $\mu_{w,p} = \bar{u}_p - \bar{u}_w$

True Bias ⇒ $\mu_{u,p} = \bar{u}_p - \bar{u}$

$h^2 = 0.05$
SLOPE $b_{w,p}$

Estimated $\Rightarrow b_{w,p} = \frac{\text{cov}(\hat{u}_p, \hat{u}_w)}{\text{var}(u_p)}$

True $\Rightarrow b_{u,p} = \frac{\text{cov}(\hat{u}_p, u)}{\text{var}(u_p)}$

$h^2=0.50$

$h^2=0.05$
Relative accuracy gain

Estimated $\Rightarrow \rho_{p,w} = \frac{\text{cov}(\widehat{u}_w, \widehat{u}_p)}{\sqrt{\text{var}(\widehat{u}_w)\text{var}(\widehat{u}_p)}}$

True $\Rightarrow \frac{\text{acc}_p}{\text{acc}_w}$

$h^2 = 0.50$

Estimated

$h^2 = 0.05$

True
Using the wrong evaluation model
Wrong evaluation fitting CG as fixed when they have a time trend

Estimated bias: $\mu_{w,p} = \bar{u}_p - \bar{u}_w$

True bias: $\mu_{u,p} = \bar{u}_p - \bar{u}$

$h^2 = 0.10$
Wrong evaluation fitting CG as fixed when they have a time trend

SLOPE $b_{w,p}$

Estimated $\Rightarrow b_{w,p} = \frac{\text{cov}(\hat{u}_p, \hat{u}_w)}{\text{var}(\hat{u}_p)}$

True $\Rightarrow b_{u,p} = \frac{\text{cov}(\hat{u}_p, u)}{\text{var}(\hat{u}_p)}$

$h^2 = 0.10$
Wrong evaluation fitting CG as fixed when they have a time trend

**Relative accuracy gain**

Estimated $\Rightarrow \rho_{p,w} = \frac{\text{cov}(\hat{u}_w, \hat{u}_p)}{\sqrt{\text{var}(\hat{u}_w)\text{var}(\hat{u}_p)}}$

True $\Rightarrow \frac{\text{acc}_p}{\text{acc}_w}$

$h^2 = 0.10$
Conclusions

• The proposed method LR estimates well bias, slope and accuracy when the model is in concordance with the reality.

But when the model has differences with the reality:

• With wrong model for contemporary groups
  • It is not possible to estimate bias or slope.
  • Accuracies can be estimated but not well

• With wrong heritabilities:
  • The bias could be under or over estimated
  • The slope is uninformative about the reality.
  • Accuracies can be estimated
Acknowledgements

Authors thank the following institutions that finance this research:

• Poctefa Project ARDI
• La region Occitanie
• INRA - Metaprogram SelGen

As well as to **Computing platform Bioinfo-Genotoul** for providing bioinformatics support.
Thank you for your attention!

Questions or comments?
Wrong $h^2$ in genetic evaluations simulated with $h^2 0.10$ and evaluated with $h^2 0.05$

**BIAS**

Estimated $\Rightarrow \mu_{w,p} = \bar{u}_p - \bar{u}_w$

True $\Rightarrow \mu_{u,p} = \bar{u}_p - \bar{u}$
Wrong $h^2$ in genetic evaluations simulated with $h^2 0.10$ and evaluated with $h^2 0.05$

**SLOPE $b_{w,p}$**

Estimated $\Rightarrow b_{w,p} = \frac{cov(\hat{u}_p, \hat{u}_w)}{var(\hat{u}_p)}$

True $\Rightarrow b_{u,p} = \frac{cov(\hat{u}_p, u)}{var(\hat{u}_p)}$
Wrong $h^2$ in genetic evaluations simulated with $h^2 0.10$ and evaluated with $h^2 0.05$

Relative accuracy gain

Estimated $\Rightarrow \rho_{p,w} = \frac{\text{cov}(\hat{u}_w, \hat{u}_p)}{\sqrt{\text{var}(\hat{u}_w)\text{var}(\hat{u}_p)}}$

True $\Rightarrow \frac{\text{acc}_p}{\text{acc}_w}$