Accuracies of contrasts between estimated breeding values of selection candidates from national cattle evaluations using pedigree or single-step genomic methodologies

D.P. Garrick\textsuperscript{1,2}, B.L. Golden\textsuperscript{1}, & D.J. Garrick\textsuperscript{1,2}

1. Theta Solutions, LLC, USA
2. A.L. Rae Centre for Genetics & Breeding, Massey University

Daniel@ThetaSolutionsLLC.com
EBV contrasts

- Evaluations produce individual EBVs for each trait (and typically a corresponding accuracy or reliability).

- Easy to find the “best” individual....
  - Rank ordering – e.g. sort by highest EBV (or index value).
    1. $EBV_{SireA}$ with accuracy/$R^2$.
    2. $EBV_{SireB}$ with accuracy/$R^2$.

- But how much better is it?
  - Contrasting EBVs of two (or more) animals.
    - $EBV_{SireA} - EBV_{SireB}$ with accuracy of?
Single Trait Birth Wt. Model

- 2,118,874 animals in the pedigree.
- 1,416,006 birthweight observations.
- 38,175 genotypes using the MSRP subset (Saatchi & Garrick 2014).

- MMEs solved (PCG) and sampled (MCMC) using BOLT software.

- 80,000 MCMC samples of plausible values of every effect stored.

- EBVs are posterior means and PEVs are posterior variances of the chain of samples.
Pedigree BLUP

\[ y = Xb + Zu + Z_m m + Z_p p + e \]

- Fixed effects (\(b\)), direct effects (\(u\)), maternal genetic (\(m\)), maternal permanent environment (\(p\)), and residual (\(e\)) effects
Super Hybrid Model

\[
\begin{bmatrix}
    y_n \\
    y_g
\end{bmatrix} =
\begin{bmatrix}
    X_n \\
    X_g
\end{bmatrix} b +
\begin{bmatrix}
    Z_n & 0 \\
    0 & Z_g
\end{bmatrix}
\begin{bmatrix}
    M_n \alpha + \varepsilon \\
    M_g \alpha
\end{bmatrix} + Zu + Zmm + Zp + e
\]

- Super Hybrid Model (Fernando et al. 2016) for genotyped (\(g\)) and non-genotyped (\(n\)) animals.

- Includes marker effects (\(\alpha\)).
- For non-genotyped animals, uses imputed markers (\(M_n\)) and fits an imputation error term (\(\varepsilon\)).
Accuracy of contrasts

- Calculated from diagonal and off-diagonal elements of the prediction error variance matrix.
  - PEV matrix is the inverse of the LHS of MME.
  - Inverse is computationally prohibitive, especially in single-step.

- Can approximate diagonal elements of inverse but PEV of contrasts rely on arbitrary off-diagonal elements

- Avoid approximation with MCMC (e.g. BOLT software).
  - GPU-accelerated single-site Gibbs sampler.
Accuracy of Contrasts

- One column per animal of interest that contains its chain of plausible EBVs.
- Make chain of contrast of samples, e.g. $EBV_{SireA} - EBV_{SireB}$

\[
\begin{bmatrix}
EBV_{1, \text{SireA}} & EBV_{1, \text{SireB}} & \cdots & EBV_{1, \text{SireZ}} \\
\vdots & \vdots & \ddots & \vdots \\
EBV_{N, \text{SireA}} & EBV_{N, \text{SireB}} & \cdots & EBV_{N, \text{SireZ}}
\end{bmatrix}
\begin{bmatrix}
1 \\
-1 \\
0 \\
\vdots \\
0
\end{bmatrix}
= \text{chain of } EBV_{SireA} - EBV_{SireB}
\]

- EBV of the contrast is the mean of the chain of contrasts.
- PEV of the contrast is the variance of the chain of contrasts.
Results – Individual EBVs

Results for “high” accuracy sires

<table>
<thead>
<tr>
<th>ID</th>
<th>EBV</th>
<th>R²</th>
<th>BIF</th>
<th>EBV</th>
<th>R²</th>
<th>BIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.71</td>
<td>0.97</td>
<td>0.84</td>
<td>0.84</td>
<td>0.97</td>
<td>0.84</td>
</tr>
<tr>
<td>B</td>
<td>3.04</td>
<td>0.96</td>
<td>0.81</td>
<td>3.15</td>
<td>0.97</td>
<td>0.81</td>
</tr>
</tbody>
</table>

High accuracy sires
Results – Individual EBVs

- Genomic information improves low accuracy.

<table>
<thead>
<tr>
<th>ID</th>
<th>PBLUP EBV</th>
<th>PBLUP R²</th>
<th>PBLUP BIF</th>
<th>SHM EBV</th>
<th>SHM R²</th>
<th>SHM BIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.71</td>
<td>0.97</td>
<td>0.84</td>
<td>0.84</td>
<td>0.97</td>
<td>0.84</td>
</tr>
<tr>
<td>B</td>
<td>3.04</td>
<td>0.96</td>
<td>0.81</td>
<td>3.15</td>
<td>0.97</td>
<td>0.81</td>
</tr>
<tr>
<td>C</td>
<td>0.10</td>
<td>0.60</td>
<td>0.37</td>
<td>-0.46</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>D</td>
<td>-0.03</td>
<td>0.59</td>
<td>0.36</td>
<td>0.36</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>E</td>
<td>1.17</td>
<td>0.60</td>
<td>0.37</td>
<td>1.47</td>
<td>0.73</td>
<td>0.48</td>
</tr>
<tr>
<td>F</td>
<td>-1.86</td>
<td>0.61</td>
<td>0.37</td>
<td>-1.81</td>
<td>0.73</td>
<td>0.48</td>
</tr>
<tr>
<td>G</td>
<td>3.48</td>
<td>0.62</td>
<td>0.38</td>
<td>3.15</td>
<td>0.72</td>
<td>0.49</td>
</tr>
<tr>
<td>H</td>
<td>-0.03</td>
<td>0.59</td>
<td>0.36</td>
<td>1.47</td>
<td>0.73</td>
<td>0.48</td>
</tr>
<tr>
<td>I</td>
<td>-2.91</td>
<td>0.59</td>
<td>0.36</td>
<td>-2.91</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>J</td>
<td>0.95</td>
<td>0.59</td>
<td>0.36</td>
<td>0.95</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>K</td>
<td>-1.73</td>
<td>0.59</td>
<td>0.36</td>
<td>-1.00</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>L</td>
<td>1.17</td>
<td>0.60</td>
<td>0.37</td>
<td>1.47</td>
<td>0.73</td>
<td>0.48</td>
</tr>
<tr>
<td>M</td>
<td>-0.72</td>
<td>0.62</td>
<td>0.38</td>
<td>-1.70</td>
<td>0.73</td>
<td>0.48</td>
</tr>
<tr>
<td>N</td>
<td>0.95</td>
<td>0.59</td>
<td>0.36</td>
<td>1.55</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>O</td>
<td>0.22</td>
<td>0.62</td>
<td>0.38</td>
<td>0.29</td>
<td>0.73</td>
<td>0.49</td>
</tr>
<tr>
<td>P</td>
<td>-1.21</td>
<td>0.58</td>
<td>0.35</td>
<td>-0.88</td>
<td>0.70</td>
<td>0.45</td>
</tr>
</tbody>
</table>

- **High accuracy sires**
- **2016 born males**

Benefit of genotyping young animals
Results – Contrasts

<table>
<thead>
<tr>
<th>Contrast</th>
<th>var(k’u)</th>
<th>PBLUP</th>
<th>SHM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PEV</td>
<td>R²</td>
<td>BIF</td>
</tr>
<tr>
<td>B-A</td>
<td>63.12</td>
<td>1.92</td>
<td>0.97</td>
</tr>
<tr>
<td>D-C</td>
<td>43.40</td>
<td>24.09</td>
<td></td>
</tr>
<tr>
<td>N-M</td>
<td>56.11</td>
<td>24.55</td>
<td>0.56</td>
</tr>
<tr>
<td>P-O</td>
<td>59.29</td>
<td>25.11</td>
<td>0.58</td>
</tr>
</tbody>
</table>

- High accuracy sires have high accuracy contrast (in this case)
### Results – Contrasts *same herd*

<table>
<thead>
<tr>
<th>Contrast</th>
<th>var($k'u$)</th>
<th>PBLUP</th>
<th>SHM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PEV</td>
<td>$R^2$</td>
<td>BIF</td>
</tr>
<tr>
<td>D-C</td>
<td>43.40</td>
<td>24.09</td>
<td>0.44</td>
</tr>
<tr>
<td>C-E</td>
<td>40.12</td>
<td>23.84</td>
<td>0.41</td>
</tr>
<tr>
<td>G-F</td>
<td>55.72</td>
<td>24.32</td>
<td>0.56</td>
</tr>
<tr>
<td>H-G</td>
<td>59.85</td>
<td>24.53</td>
<td>0.59</td>
</tr>
</tbody>
</table>

- Young selection candidates with same sire
- Young selection candidates with different sires
Results – Contrasts different herd

<table>
<thead>
<tr>
<th>Contrast</th>
<th>( \text{var}(k'u) )</th>
<th>PBLUP</th>
<th>( R^2 )</th>
<th>BIF</th>
<th>SHM</th>
<th>( R^2 )</th>
<th>BIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>J-I</td>
<td>44.28</td>
<td>24.78</td>
<td>0.44</td>
<td>0.25</td>
<td>16.98</td>
<td>0.62</td>
<td>0.38</td>
</tr>
<tr>
<td>L-K</td>
<td>44.51</td>
<td>24.26</td>
<td>0.46</td>
<td>0.26</td>
<td>16.68</td>
<td>0.63</td>
<td>0.39</td>
</tr>
<tr>
<td>N-M</td>
<td>56.11</td>
<td>24.55</td>
<td>0.56</td>
<td>0.34</td>
<td>16.87</td>
<td>0.70</td>
<td>0.45</td>
</tr>
<tr>
<td>O-P</td>
<td>59.29</td>
<td>25.11</td>
<td>0.58</td>
<td>0.35</td>
<td>17.77</td>
<td>0.70</td>
<td>0.45</td>
</tr>
</tbody>
</table>

- Young selection candidates with same sire
- Young selection candidates with different sires

- Genomic information improves accuracy of contrasts.
- Contrast between animals with the same sire have lower accuracy than those with different sires regardless of herd since PEV similar but \( \text{var}(k'u) \) lower.
- These herds are well-connected due to wide AI use.
Conclusions

- For comparing animals, it is the contrast (and the accuracy/reliability of contrasts) that matters.

- Accuracy of the individuals EBV’s are not an indication of the accuracy of the contrast.
  - Depends on prediction error co-variances which are influenced by “connectedness”.

- MCMC sampling of the MME using BOLT software is a computationally efficient method for national animal evaluations.
  - Same MCMC principles can be applied to selection indexes.
Special thanks to the American Hereford Association for allowing the use of their national single-step evaluation data.
EBV Accuracy

- Individual EBVs reported with a corresponding accuracy
  - Reliability $R^2$
  - Accuracy $r = \sqrt{R^2}$
  - BIF accuracy

- “Measure” of the amount of information that went into producing the EBV
  - Quantify the possible variation of the EBV

- Prediction error variance (PEV)
  - Elements of inverse of LHS of MME
  - Inverse is prohibitive to compute, especially in single-step
  - PEVs commonly approximated
Results - Contrasts

- For a contrast vector $k$, accuracy is computed as:
- $R^2 = 1 - \text{PEV}_k / (k'Gk)$ and $\text{BIF} = 1 - \sqrt{1 - R^2}$
- $\text{Var}(k'u) = k'Gk$, where $G = \text{var}(u) = \sigma_g^2 A$ for single trait
  where $A$ is the numerator relationship matrix

- For demonstration purposes $\text{var}(k'u)$ is taken to be the same in both the PBLUP and SHM (ie $k'Gk$)
Simplified hypothetical example

- Sire A and Sire B

- Both sires have many offspring.
- Both sires have high accuracy.

- But, offspring are in their own herds and geographically isolated from each other.

- A contrast of these sires might have a low or high accuracy.
  - Even in the absence of GxE
Overview

- Why EBV contrasts are important.

- Single trait birthweight model
  - Pedigree BLUP genetic evaluation.
  - Single-step Super Hybrid Model (SHM) genomic evaluation.

- Accuracy of contrasts.

- Results.
  - Contrast of high accuracy sires.
  - Contrasts of 2016 born males.

- Conclusions.