

Individual Reliabilities of Molecular Breeding Values

Stephen D. Kachman



BIF Genetic Prediction Workshop

December 13, 2013

Introduction

- Reliability and MBV
- Global versus Individual
- Model assumptions
- Parametric form
- Implications

Reliability and MBV

- Reliability is the squared genetic correlation

$$R_{MBV}^2 = \frac{\text{cov}(MBV_i, u_i)^2}{\text{var}(MBV_i) \text{var}(u_i)}$$

- Covariance between the MBV and the true breeding value $\text{cov}(MBV_i, u_i)$
- Variance of the MBV $\text{var}(MBV_i)$
- Genetic variance $\text{var}(u_i)$
- Proportion of the genetic variance accounted for by the MBV

- Scaling the MBV

$$\sqrt{R_{MBV}^2 \frac{\text{var}(u_i)}{\text{var}(MBV_i)}}$$

- Blending (Ignoring scaling and base adjustments)

$$EBV + \frac{1 - R_{EBV}^2}{1 - R_{MBV}^2 R_{EBV}^2} [MBV - R_{MBV}^2 EBV]$$

- GE-EPD Reliability

$$R_{EBV}^2 + R_{MBV}^2 (1 - R_{EBV}^2) \left[\frac{1 - R_{EBV}^2}{1 - R_{EBV}^2 R_{MBV}^2} \right]$$

Correlated Traits

- Genetic Covariance Matrix (Ignoring scaling)

$$\sigma_g^2 \begin{pmatrix} 1 & R_{MBV}^2 \\ R_{MBV}^2 & R_{MBV}^2 \end{pmatrix}$$

Correlated Traits

- Genetic Covariance Matrix (Ignoring scaling)

$$\sigma_g^2 \begin{pmatrix} 1 & R_{MBV}^2 \\ R_{MBV}^2 & R_{MBV}^2 \end{pmatrix}$$

- Standard assumptions
 - ▶ MBV reliability is a constant
 - ▶ MBV is uncorrelated with the prediction error $u_i - \hat{u}_i$

Animals in training

- Information coming from the genotypes of animals used in training

Animals in training

- Information coming from the genotypes of animals used in training
- Information is also coming from the phenotypes of animals that went into training
 - ▶ Phenotypes of the genotyped animals
 - ▶ EPDs of the genotyped animals
 - ★ Phenotypes of all animals used in the NCE

- Estimated SNP effects are based

$$EBV = TBV + (EBV - TBV)$$

- A MBV also predicts the prediction error

- Estimated SNP effects are based

$$EBV = TBV + (EBV - TBV)$$

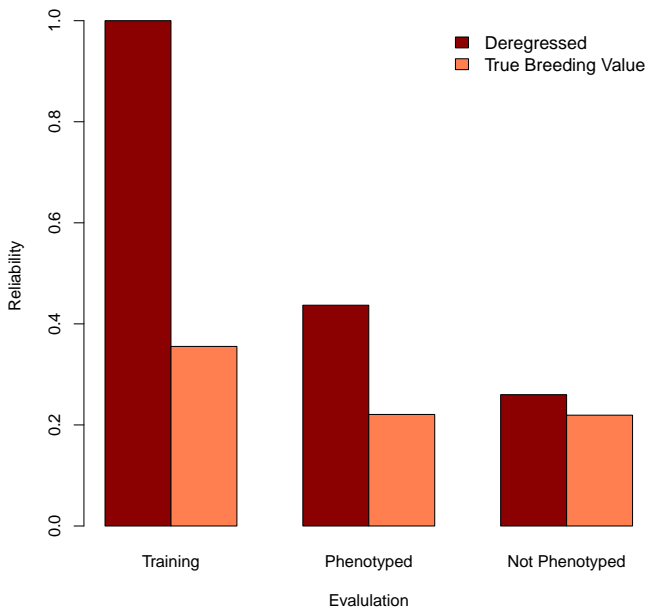
- A MBV also predicts the prediction error

Minimize prediction error feedback

- High accuracy animals in training
- Separation between training and evaluation

Impact of training on both estimated and true reliability

- 50K SNP panel
- 1,300 genotyped animals in training
 - ▶ 300: Foundation animals
 - ▶ 1,000: Current sires and dams
- Phenotypes
 - ▶ Historic
 - ▶ Half the progeny
- Bayes B using deregressed EBV
- Evaluated using complete phenotypic information



Global versus Individual

- Global reliability assumes that all animals being evaluated have the same reliability

$$EBV + \frac{1 - R_{EBV}^2}{1 - R_{MBV}^2 R_{EBV}^2} [MBV - R_{MBV}^2 EBV]$$

- High accuracy animals
MBV will have a minimal impact
- Low accuracy animals
Where we want to get it right
- Know reliability decreases the further you are away from training
How to model that change?

Approach

- Construct a simplified model
 - ▶ Capture the main features
 - ▶ Keep the math manageable
 - ▶ Keep it as general as possible
- Predict SNP effects
 - ▶ Based on EBV information from genotyped animals in training
 - ▶ Accounting for the covariances between the EBV
- Incorporate approximations to end up with a manageable result without losing the main features

Model assumptions

- SNP covariate matrix \mathbf{M}
- SNP effect vector \mathbf{b}

$$\text{Vec}(\mathbf{M}) \sim (\mathbf{0}, \mathbf{L} \otimes \mathbf{A})$$

$$\mathbf{b} \sim (\mathbf{0}, \mathbf{I}\sigma_b^2)$$

- Genomic relationship matrix

$$\mathbf{G} = \mathbf{M}\mathbf{M}' / \text{tr}(\mathbf{L})$$

$$E(\mathbf{G}) = \mathbf{A}$$

- Breeding values

$$\mathbf{u} = \mathbf{M}\mathbf{b}$$

$$\mathbf{u} \sim (\mathbf{0}, \mathbf{A}\sigma_u^2)$$

$$\sigma_u^2 = \text{tr}(\mathbf{L})\sigma_b^2$$

Data for training

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathbf{u} \sim (0, \mathbf{A}\sigma_u^2)$$

$$\mathbf{e} \sim (0, \mathbf{R})$$

$$\begin{pmatrix} \mathbf{u} \\ \hat{\mathbf{u}} \end{pmatrix} \sim \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A}\sigma_u^2 & \text{var}(\hat{\mathbf{u}}) \\ \text{var}(\hat{\mathbf{u}}) & \text{var}(\hat{\mathbf{u}}) \end{pmatrix} \right)$$

- Predict SNP effects using $\hat{\mathbf{u}}_1$ and \mathbf{M}_1

Parametric form

- Approximate Covariance between the MBV and the BV

$$\frac{\text{tr}(\mathbf{L})^2 + \text{tr}(\mathbf{L}^2)}{\text{tr}(\mathbf{L})^2} \mathbf{a}'_{1i} \mathbf{A}_{11}^{-1} \text{var}(\hat{\mathbf{u}}_1) \mathbf{A}_{11}^{-1} \mathbf{a}_{1i} + \frac{\text{tr}(\mathbf{L}^2) \text{tr}[\mathbf{A}_{11}^{-1} \text{var}(\hat{\mathbf{u}}_1)]}{\text{tr}(\mathbf{L})^2} a_{ii}$$

- Approximate Variance of the MBV

$$\frac{\text{tr}(\mathbf{L})^2 + \text{tr}(\mathbf{L}^2)}{\text{tr}(\mathbf{L})^2} \mathbf{a}'_{1i} \mathbf{A}_{11}^{-1} \text{cov}(\hat{\mathbf{u}}_1, \hat{u}_i) + \frac{\text{tr}(\mathbf{L}^2) \text{tr}[\mathbf{A}_{11}^{-1} \text{var}(\hat{\mathbf{u}}_1)]}{\text{tr}(\mathbf{L})^2} a_{ii}$$

Two unknown parameters

- Approximate Covariance between the MBV and the BV

$$\left[R_S^2 \frac{\mathbf{a}'_{1i} \mathbf{A}_{11}^{-1} \text{var}(\hat{\mathbf{u}}_1) \mathbf{A}_{11}^{-1} \mathbf{a}_{1i}}{a_{ii} \sigma_u^2} + R_B^2 \right] a_{ii} \sigma_u^2$$

- Approximate Variance of the MBV

$$\left[R_S^2 \frac{\mathbf{a}'_{1i} \mathbf{A}_{11}^{-1} \text{cov}(\hat{\mathbf{u}}_1, \hat{u}_i)}{a_{ii} \sigma_u^2} + R_B^2 \right] a_{ii} \sigma_u^2$$

- Both terms are very similar as expected

$$\mathbf{a}'_{1i} \mathbf{A}_{11}^{-1} \text{var}(\hat{\mathbf{u}}_1) \mathbf{A}_{11}^{-1} \mathbf{a}_{1i} = \text{var}(BLUP(u_i | \hat{\mathbf{u}}_1))$$

$$\mathbf{a}'_{1i} \mathbf{A}_{11}^{-1} \text{cov}(\hat{\mathbf{u}}_1, \hat{u}_i) = \text{cov}(BLUP(u_i | \hat{\mathbf{u}}_1), \hat{u}_i)$$

- Similarity to training

$$\frac{\text{var}(BLUP(u_i | \hat{\mathbf{u}}_1))}{a_{ii} \sigma_u^2}$$

Reliability of the BLUP of the EBV based on EBV of the genotyped animals in training

- Approximate MBV Reliability will decrease from a maximum of $R_S^2 + R_B^2$ for very high accuracy animals
- To a minimum of R_B^2 for animals who are unrelated to training

- Look at R_B^2

$$\text{tr} [\mathbf{A}_{11}^{-1} \text{var}(\hat{\mathbf{u}}_1)]$$

- ▶ Reliability of animals in training
- ▶ Diversity in the training population
- Look at multiplier for R_S^2
 - ▶ Want animals in training close to the animals being targeted for evaluation

Summary

- These are approximate results
- Do provide a way to account for distance from training when using MBV in genetic evaluations
- Even genotyped animals in training aren't expected to have the same reliability
- Provides a means to identify who we want to genotype based on who we want to evaluate
- Also need the $\text{cov}(MBV_i, \hat{u}_i)$ which still needs work to put it into a usable form