Causes of inflation in genomic evaluations for young genotyped dairy bulls

Shogo Tsuruta[1], Daniela, A. L. Lourenco[1], Ignacy Misztal[1], and Tom J. Lawlor[2]

[1]University of Georgia, Athens, USA, [2]Holstein Association USA Inc., Brattleboro, VT

The University of Georgia.

Introduction

➢ With no daughter's record, why genomic (G) EBV for young genotyped bulls biased?

➢ Several possibilities: selection, missing or inaccurate pedigree, inaccurate genetic parameters, … , creating inconsistency between pedigree-base (**A**) and genomic (**G**) relationships?

Objectives

Investigate what causes biases (inflation and deflation) in GEBV for young genotyped animals in dairy cattle

Materials and Methods

Simulation

➢ **50 ♂ x 5000 ♀ for 10 generations**
➢ **29 autosomes**
➢ **60K SNP assuming 600 QTL**
➢ **10 replicates**
➢ **3 scenarios**

1)
◆ 20K genotyped animals in the last 4 generations
◆ 30K phenotypes in all 10 generations
◆ 40K animals in pedigree
◆ selecting top 10% and 50% in EBV from ♂ and ♀, respectively
◆ average inbreeding coefficient = 6.1%

2)
◆ the same population structure as in A) with random selection
◆ average inbreeding coefficient = 1.4%

3)
◆ 50K genotyped animals and 30K phenotypes in all 10 generations
◆ 55K animals in pedigree
◆ selecting top 10% and 50% in EBV from ♂ and ♀, respectively
◆ average inbreeding coefficient = 5.9%

➢ **True breeding values (TBV) = $b_0 + b_1 \times$ (G)EBV**
(G)EBV with no phenotypes and with genotypes in generations 9 and 10
$b_0 > 0.0$ => overestimation; $b_0 < 0.0$ => underestimation
$b_1 < 1.0$ => inflation; $b_1 < 1.0$ => deflation

Linear type traits in US Holsteins

➢ Data for the national genetic evaluation in August 2014 from *Holstein Association USA Inc.* (*Brattleboro, Vermont, USA*).
10M records for 10M animals
60K SNP markers for 570K genotyped animals

➢ **Daughter Yield Deviation (DYD) in 2014 = $b_0 + b_1 \times$ (G)EBV in 2010**
(G)EBV for young genotypes bulls with no phenotypes in 2010

Single-step genomic mixed model equation

➢ MME including unknown parent group (UPG) when parent(s) are missing

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z + H^{-1}\alpha} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

where

$$\mathbf{H^{-1}} = \begin{bmatrix} \mathbf{A^{11}} & \mathbf{A^{12}} \\ (\mathbf{A^{12}})' & \mathbf{A^{22}} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & (\mathbf{G^{-1} - \omega A_{22}^{-1}}) \end{bmatrix}$$

Results

Simulation

Sires: regression coefficient (b_1) on TBV[1] = $b_0 + b_1 \times$ (G)EBV[2]

Sire's generation[3]	Inbreeding[4] in **A**[5]		No inbreeding in **A**		Missing dams[6]	
	EBV	GEBV	EBV	GEBV	EBV	GEBV
BLUP selection with genotypes in the last 4 generations						
All	1.00	0.96	1.00	0.90	1.00	0.94
10	0.94	0.83	0.93	0.70	0.98	0.82
9	0.97	0.83	0.97	0.71	0.99	0.82
8	0.99	0.85	0.98	0.73	1.21	0.80
7	1.00	0.87	1.00	0.77	1.03	0.91
No selection with genotypes in the last 4 generations						
All	0.99	0.88	0.99	0.87	0.95	0.86
10	0.95	0.85	0.95	0.83	0.94	0.85
9	0.99	0.87	0.99	0.86	0.99	0.87
8	1.03	0.89	1.02	0.87	0.98	0.88
7	0.98	0.90	0.98	0.88	0.98	0.89
BLUP selection with genotypes in all generations						
All	1.00	0.96	1.00	0.98	1.00	1.00
10	0.94	0.83	0.93	0.71	0.98	0.84
9	0.97	0.83	0.97	0.71	0.99	0.83
8	0.99	0.85	0.98	0.71	1.21	0.83
7	1.00	0.87	1.00	0.76	1.03	0.86

Cows: regression coefficient (b_1) on TBV[1] = $b_0 + b_1 \times$ (G)EBV[2]

Cow's generation[3]	Inbreeding[4] in **A**[5]		No inbreeding in **A**		Missing dams[6]	
	EBV	GEBV	EBV	GEBV	EBV	GEBV
BLUP selection using genotypes in the last 4 generations						
All	1.00	1.00	1.00	0.99	0.99	1.01
10	0.96	0.98	0.94	0.87	0.90	0.99
9	0.98	0.99	0.96	0.89	0.91	0.99
8	0.99	0.89	0.99	0.77	1.27	0.85
7	0.99	1.01	0.98	0.94	1.00	1.05
No selection with genotypes in the last 4 generations						
All	1.00	1.00	1.00	0.99	0.99	0.99
10	0.99	0.98	0.98	0.97	0.98	0.98
9	1.00	0.99	0.99	0.98	0.99	0.99
8	1.00	0.91	1.00	0.90	0.96	0.91
7	1.00	1.00	1.00	0.99	1.00	1.00
BLUP selection with genotypes in all generations						
All	1.00	1.00	1.00	0.98	0.99	0.99
10	0.96	0.98	0.94	0.86	0.90	0.97
9	0.98	0.99	0.96	0.88	0.91	0.99
8	0.99	0.89	0.99	0.75	1.27	0.87
7	0.99	1.01	0.98	0.91	1.00	1.00

[1] True breeding values
[2] (Genomic) estimated breeding values
[3] Sires in generations 8, 9 and 10 with no daughters; cows in generations 9 and 10 with no phenotypes
[4] Inbreeding with no UPG
[5] Pedigree-based relationship matrix for all animals
[6] Missing dams in pedigree for generations 8, 9, and 10 with UPG

Linear type traits in US Holsteins

Bulls: regression coefficient (b_1) on DYD = $b_0 + b_1 \times$ (G)EBV sorted by genetic progress (GP)

Trait	Genetic progress	PA	Inbreeding In A	Inbreeding UPG	Inbreeding UPG bull	No inbreeding w = 0.95
Rear udder height	2.39	0.71	0.76	0.96	0.86	0.86
Rear udder width	2.25	0.71	0.77	0.96	0.88	0.83
Fore udder attachment	2.08	0.69	0.78	1.03	0.90	0.95
Udder depth	1.61	0.73	0.84	1.08	0.94	1.01
Foot angle	1.43	0.73	0.75	1.02	0.86	0.98
Rear legs rear view	1.40	0.60	0.69	0.98	0.81	0.90
Feet & legs score	1.26	0.54	0.61	0.82	0.71	0.76
Udder cleft	1.24	0.88	0.83	1.13	0.96	1.06
Teat placement	1.09	0.82	0.85	1.14	0.94	1.08
Stature	1.06	0.80	0.84	1.03	0.93	0.97
R teat placement	1.05	0.92	0.89	1.22	0.99	1.15
Dairy form	0.96	0.79	0.86	1.09	0.98	1.03
Rump width	0.82	0.82	0.84	1.10	0.96	1.04
Strength	0.45	0.81	0.86	1.16	0.95	1.05
Body depth	0.43	0.79	0.84	1.05	0.92	1.00
Rump angle	0.05	0.81	0.85	1.13	0.95	1.10
Rear legs side view	-0.14	0.85	0.84	1.21	0.95	1.15
Teat length	-0.34	0.75	0.90	1.15	0.98	1.11
Mean	1.01	0.76	0.81	1.07	0.92	1.00
Correlation with GP	GP	-0.42	-0.51	-0.60	-0.45	-0.69

Conclusions

➢ Selection => more genetic progress => more inflation in genetic (PA) and genomic (GEBV) predictions
➢ Inconsistency between **A** and $\mathbf{A_{22}}$ => inflation in genomic predictions
➢ Inconsistency between **G** and $\mathbf{A_{22}}$ => inflation in genomic predictions
➢ Inflation in GEBV for sires with no daughters and cows with no phenotypes
➢ No bias in EBV
➢ Missing pedigree (dams) => not a source of inflation
➢ Options to reduce the inflation in GEBV:
1) include inbreeding in **A** considering UPG
2) use smaller additive genetic variances or current (co)variance component estimates
3) adjust the weight ($\omega < 1.0$) on $\mathbf{A_{22}^{-1}}$
4) Use all available pedigree information
5) Define UPG that can be estimated with no bias

References

Aguilar, I., & I. Misztal. 2008. J. Dairy Sci. 91:1669-1672.

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta & T.J. Lawlor. 2010. J. Dairy Sci. 93:743-752.

Legarra, A., I. Aguilar, & I. Misztal. 2009. J. Dairy Sci. 92:4656-4663.

Mrode, R.A., & G.J.T. Swanson. 2004. Livest. Prod. Sci. 86(1-3):253-260.

Sargolzaei M. & F.S. Schenkel. 2009.. Bioinformatics. 25: 680–681.

Tsuruta, S., I. Misztal., I. Aguilar & T.J. Lawlor. 2011. J. Dairy Sci. 94:4198-4204.

Vitezica, Z.G., I. Aguilar, I. Misztal & A. Legarra. 2011. Genet. Res. 93:357-366.