Choice of parameters for single-step genomic evaluation for type

Ignacy Misztal, Ignacio Aguilar, Shogo Tsuruta University of Georgia Andres Legarra, INRA Toulouse Tom Lawlor, Holstein Association, USA

Genomic evaluations in dairy

- Multi-step (e.g., VanRaden et al., 2009)
 - Works
 - Several approximations (deregression, uncorrelated residuals, weights for index with PA)
 - Biases
 - Hard to apply for complicated models
- Single-step (Aguilar et al., 2010)
 - Much simpler no approximations
 - Works with any model
 - Biases present but reduced with modifications

Goal: Investigate several modifications in single-step evaluation with respect to bias and accuracy

Matrix that combines genomic and pedigree relationships

$$\mathbf{H} = \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} \mathbf{G} - \mathbf{A}_{22} \quad \mathbf{I} \quad \mathbf{I} \begin{bmatrix} \mathbf{A}_{22}^{-1} \mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

- A pedigree-based relationship matrix
- **G** genomic relationship matrix
- 1- ungenotyped animals
- 2- genotyped animals

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

Aguilar et al. (2010)

Legarra et al. (2009)

Analyzes of final scores (Aguilar et al., 2010)

- 10.5 million records US Holsteins
- Genotypes for 6,508 bulls
- Repeatability model
- G with 0.5 gene frequencies
- Modified BLUP90IOD (Tsuruta et al.)

Quality of several predictions (Aguilar et al., 2010)

Prediction in	DD2009			
2004	R ² δ (Regr)			
Parent Avg	24	0.76		
Multistep	40	0.86		
Single-step				
$G^{-1}-A_{22}^{-1}$	41	0.76		
$0.7(G^{-1}-A_{22}^{-1})$	40	0.88		
$0.5(G^{-1}-A_{22}^{-1})$	39	0.92		

Why adjustments needed?

- G on wrong scale?
- Adjustment for parental index too small or too large (as response to preferential treatment)?
- Lack of polygenic component?

Unmodified H⁻¹

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}.$$

$$Var(\mathbf{u}_2) = \mathbf{G}$$

Polygenic and genomic effects

$var(\mathbf{u}_2) = \alpha \mathbf{G} + \beta \mathbf{A}_{22}$



DYD R^2 as function of α and β

	- 0	0.1 -	0.2 -	0.3 -	0.4 -	0.5 -	0.6 -
1.2 -	0.404	0.404	0.403	0.401	0.397	0.394	0.39
1.1 -	0.407	0.407	0.406	0.402	0.399	0.395	0.391
1 -	0.409	0.409	0.407	0.404	0.402	0.397	0.393
ರ 0.9 -	0.411	0.412	0.411	0.406	0.402	0.398	0.394
0.8 -	0.412	0.413	0.41	0.407	0.403	0.399	0.394
0.7 -	0.413	0.415	0.412	0.408	0.404	0.399	0.395
0.6	0.413	0.415	0.412	0.408	0.404	0.399	0.394

DYD Regression coefficient as function of α and β

0.6 -	0.82	0.828	0.833	0.827	0.825	0.808	0.806
0.7 -	0.8	0.809	0.808	0.802	0.794	0.786	0.785
0.8 -	0.781	0.788	0.784	0.781	0.774	0.767	0.759
ರ <i>0.9</i> –	0.764	0.769	0.778	0.761	0.755	0.748	0.742
1 -	0.747	0.751	0.748	0.744	0.754	0.732	0.725
1.1 -	0.731	0.734	0.738	0.727	0.721	0.716	0.724
1.2 -	0.716	0.718	0.722	0.728	0.708	0.702	0.696
	- 0	0.1 -	0.2 -	ъ 0.3 –	0.4 -	0.5 -	0.6 -

 $var(\mathbf{u}_2) = \alpha \mathbf{G} + \beta \mathbf{A}_{22}$

α	β	R ²	Regression
0.95	0.05	0.41	0.75
0.8	0.2	0.41	0.78
0.6	0.1	0.42	0.83
1.2	0.5	0.39	0.70

Optimal G if multiplied by 0.6

Single weight

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) \end{bmatrix}$$

$$\operatorname{Var}\left(\mathbf{u}_{2}\right) = \left[\lambda \mathbf{G}^{-1} + 1 - \lambda \quad \mathbf{A}_{22}^{-1} \right]^{-1}.$$

$$\mathbf{u}_2 \mid \mathbf{A}_{22}, \mathbf{G} \sim \mathrm{N}(0, \frac{\mathbf{G}}{\lambda}) = \mathrm{N}(0, \frac{\mathbf{A}_{22}}{1-\lambda})$$

 λ – fraction of information from genomics

Weights on G⁻¹ and A₂₂⁻¹

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau \mathbf{G}^{-1} - \boldsymbol{\omega} \mathbf{A}_{22}^{-1} \end{bmatrix}.$$

$$\operatorname{Var}\left(\mathbf{u}_{2}\right) = \left[\tau \mathbf{G}^{-1} + 1 - \omega \quad \mathbf{A}_{22}^{-1} \right]^{-1}.$$

$$\mathbf{u}_2 | \mathbf{A}_{22}, \mathbf{G} \sim \mathrm{N}(0, \frac{\mathbf{G}}{\tau}) = \mathrm{N}(0, \frac{\mathbf{A}_{22}}{1-\omega})$$

au – scaling factor ω – fraction of information from genomics

DYD R^2 as function of τ and ω

			-	9	9	9	9	9	9	9	9	9	I
			0	-	4	ά	4	ģ	ģ	~	0	6	-
1	.5	_	0.38	0.384	0.389	0.394	0.398	0.403	0.407	0.411	0.413	0.416	0.415
1	.4		0.379	0.384	0.389	0.393	0.398	0.402	0.407	0.411	0.414	0.416	0.415
1	.3		0.379	0.383	0.388	0.392	0.397	0.402	0.407	0.411	0.414	0.416	0.415
1	.2	-	0.377	0.382	0.386	0.391	0.397	0.401	0.406	0.41	0.414	0.416	0.414
1	.1		0.376	0.381	0.385	0.39	0.395	0.4	0.406	0.41	0.414	0.415	0.413
	1	-	0.373	0.378	0.383	0.389	0.394	0.399	0.405	0.41	0.413	0.414	0.411
0	.9	-	0.371	0.376	0.382	0.387	0.393	0.398	0.404	0.409	0.412	0.413	0.41
÷0	.8	-	0.369	0.373	0.379	0.385	0.39	0.397	0.402	0.407	0.41	0.412	0.405
0	.7	-	0.365	0.371	0.376	0.382	0.388	0.394	0.4	0.404	0.407	0.407	0.401
0	.6		0.361	0.367	0.372	0.378	0.385	0.391	0.398	0.401	0.404	0.403	0.394
0	.5	_	0.355	0.36	0.367	0.373	0.379	0.385	0.391	0.396	0.399	0.397	0.386
0	.4	-	0.348	0.353	0.36	0.366	0.372	0.378	0.384	0.389	0.392	0.39	0.374
0	.3	-	0.338	0.343	0.349	0.356	0.364	0.368	0.374	0.381	0.383	0.38	0.353
0	.2	-	0.322	0.327	0.333	0.338	0.346	0.352	0.359	0.366	0.369	0.361	0.306
0	.1	-	0.298	0.301	0.305	0.31	0.316	0.322	0.329	0.335	0.332	0.299	0.176

DYD Regression coefficient as function of τ and ω

		-	9	9	9	9	0	0	9	Ŷ.	9	Ĩ
		0	-		ŝ	4	6	6	~	00	6	-
1.5	5 -	1.04	1.027	1.021	1.007	0.995	0.978	0.957	0.932	0.898	0.865	0.819
1.4	4 -	1.033	1.023	1.019	1.006	0.988	0.973	0.95	0.927	0.892	0.858	0.81
1.3	} -	1.034	1.022	1.014	0.998	0.986	0.968	0.947	0.921	0.887	0.851	0.799
1.2	? -	1.027	1.018	1.003	0.99	0.982	0.96	0.943	0.914	0.879	0.842	0.788
1.1	- 1	1.022	1.015	1.004	0.985	0.973	0.953	0.937	0.91	0.875	0.827	0.774
1	- 1	1.012	1.002	0.993	0.979	0.965	0.948	0.926	0.902	0.868	0.82	0.759
0.9) –	1.01	0.996	0.992	0.974	0.964	0.946	0.925	0.894	0.858	0.808	0.753
₩0.8	3 -	1.006	0.99	0.982	0.974	0.953	0.937	0.916	0.886	0.846	0.804	0.72
0.7	7 -	0.996	0.989	0.973	0.96	0.945	0.932	0.906	0.873	0.832	0.775	0.703
0.6	5 -	0.985	0.98	0.964	0.958	0.942	0.922	0.904	0.862	0.816	0.755	0.667
0.5	5 -	0.976	0.963	0.958	0.948	0.932	0.91	0.882	0.847	0.797	0.731	0.645
0.4	4 -	0.965	0.949	0.947	0.933	0.917	0.893	0.864	0.827	0.776	0.703	0.592
0.3	} -	0.945	0.937	0.926	0.923	0.914	0.873	0.843	0.813	0.751	0.671	0.536
0.2	2 -	0.915	0.908	0.897	0.883	0.876	0.845	0.817	0.787	0.725	0.631	0.443
0.1	1 -	0.895	0.859	0.847	0.835	0.829	0.803	0.781	0.745	0.682	0.545	0.235

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau \mathbf{G}^{-1} - \boldsymbol{\omega} \mathbf{A}_{22}^{-1} \end{bmatrix}.$$

т	ω	R ²	Regression
1.0	1.0	0.41	0.75
1.5	0.9	0.42	0.87
1.5	0.6	0.41	0.96
1.5	0.4	0.40	1.00
1.0	0.4	0.39	0.97
0.6	0.4	0.39	0.94

Optimal G if multiplied by 0.66



$w_1PA + w_2GEBV - w_3GPI$

PA = Parent AverageGEBV = Genomic EBVGPI = Parental Index for genotyped animals

Smaller ω reduces u_i

Large genomic information

For sons of popular bulls: $g^{ii} >> 2$, $g^{ii} >> a_{22}^{ii}$

$$u_{i} \approx \frac{\sum_{j,j\neq i} (-\tau g^{ij})u_{j}}{\tau g^{ii}} = \frac{\sum_{j,j\neq i} (-g^{ij})u_{j}}{g^{ii}} = GEBV$$

If gⁱⁱ large, genomic predictions almost independent of many parameters

Conclusions

- R² affected little by wide variations of parameters (dairy only)
- Biases (inflation) strongly affected by weight on the parental index of genotyped bulls
- R² slightly higher if G scaled correctly
- Effect of polygenic effect small for SNP50k

Acknowledgements

Paul VanRaden Rohan Fernando **Bill Muir** George Wiggans Curt VanTassel **Bruce** Tier Guilherme Rosa

NIFA grants Holstein Association Beef Cattle Consortium **Cobb-Vantress Smithfield Premium** Genetics PIC

💫 🞜 🏛 🤮 😂 🍳 🔧 ?



-1074.54, -8.32335