

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 (deregession, 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

Legarra et al. (2009)

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

Aguilar et al. (2010)

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 2004	DD2009	
	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

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

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 $\mathbf{H}^{-1}$

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

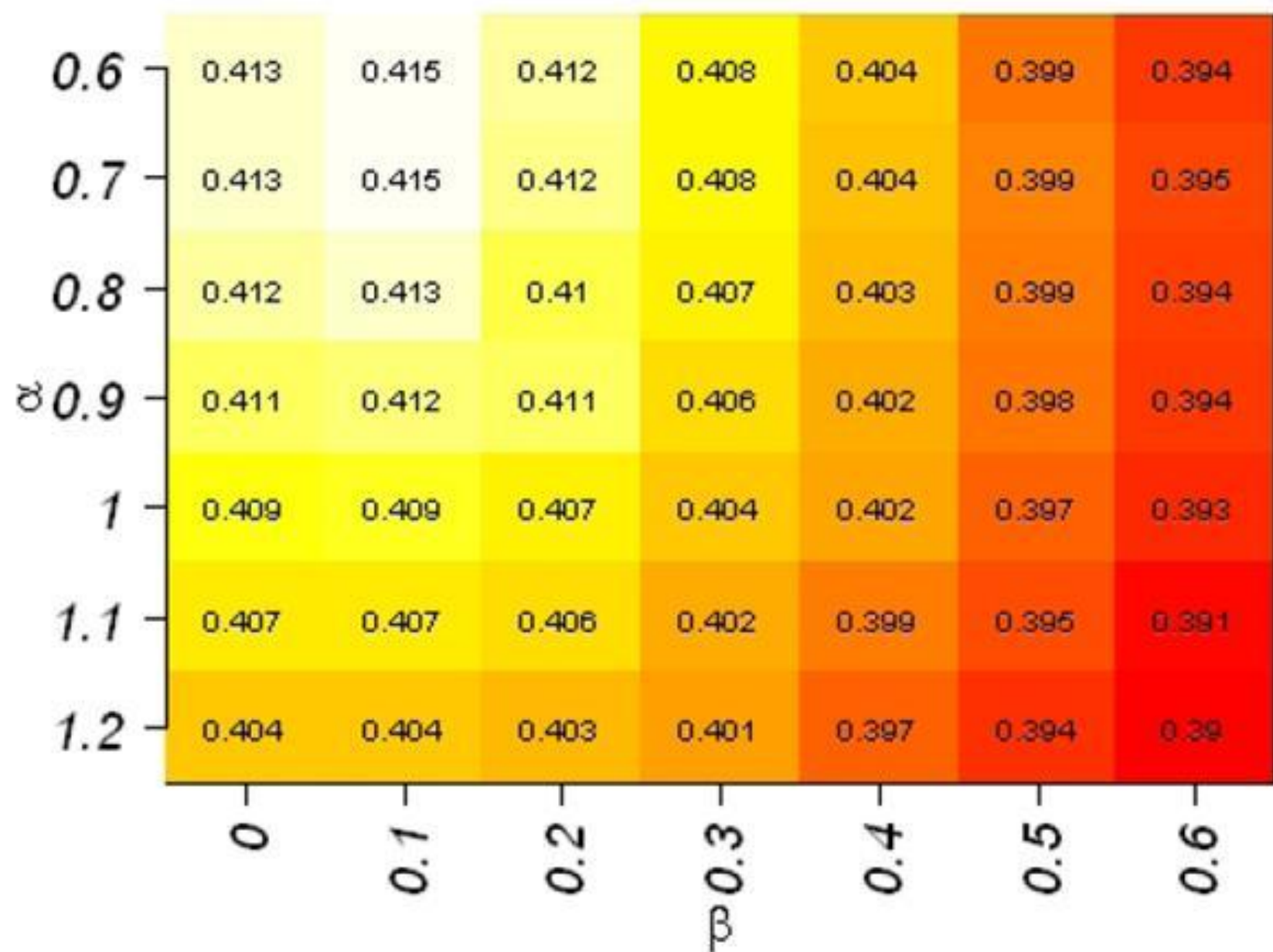
$$\text{Var}(\mathbf{u}_2) = \mathbf{G}$$

Polygenic and genomic effects

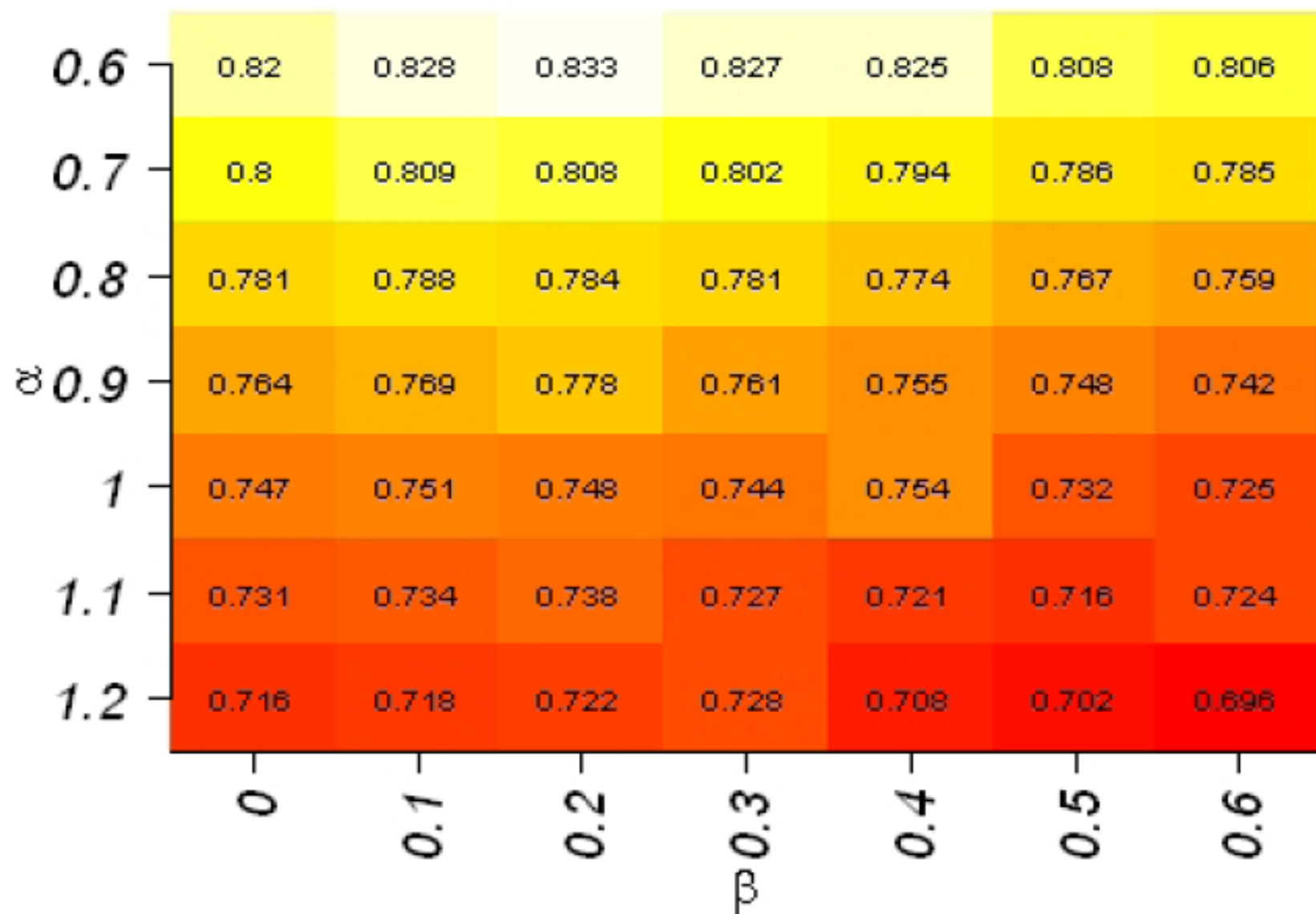
$$\text{var}(\mathbf{u}_2) = \alpha \mathbf{G} + \beta \mathbf{A}_{22}$$

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

DYD R^2 as function of α and β



DYD Regression coefficient as function of α and β



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

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

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

$$\text{Var}(\mathbf{u}_2) = \begin{bmatrix} \lambda \mathbf{G}^{-1} + 1 - \lambda & \mathbf{A}_{22}^{-1} \end{bmatrix}^{-1}.$$

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

λ – fraction of information from genomics

Weights on $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$

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

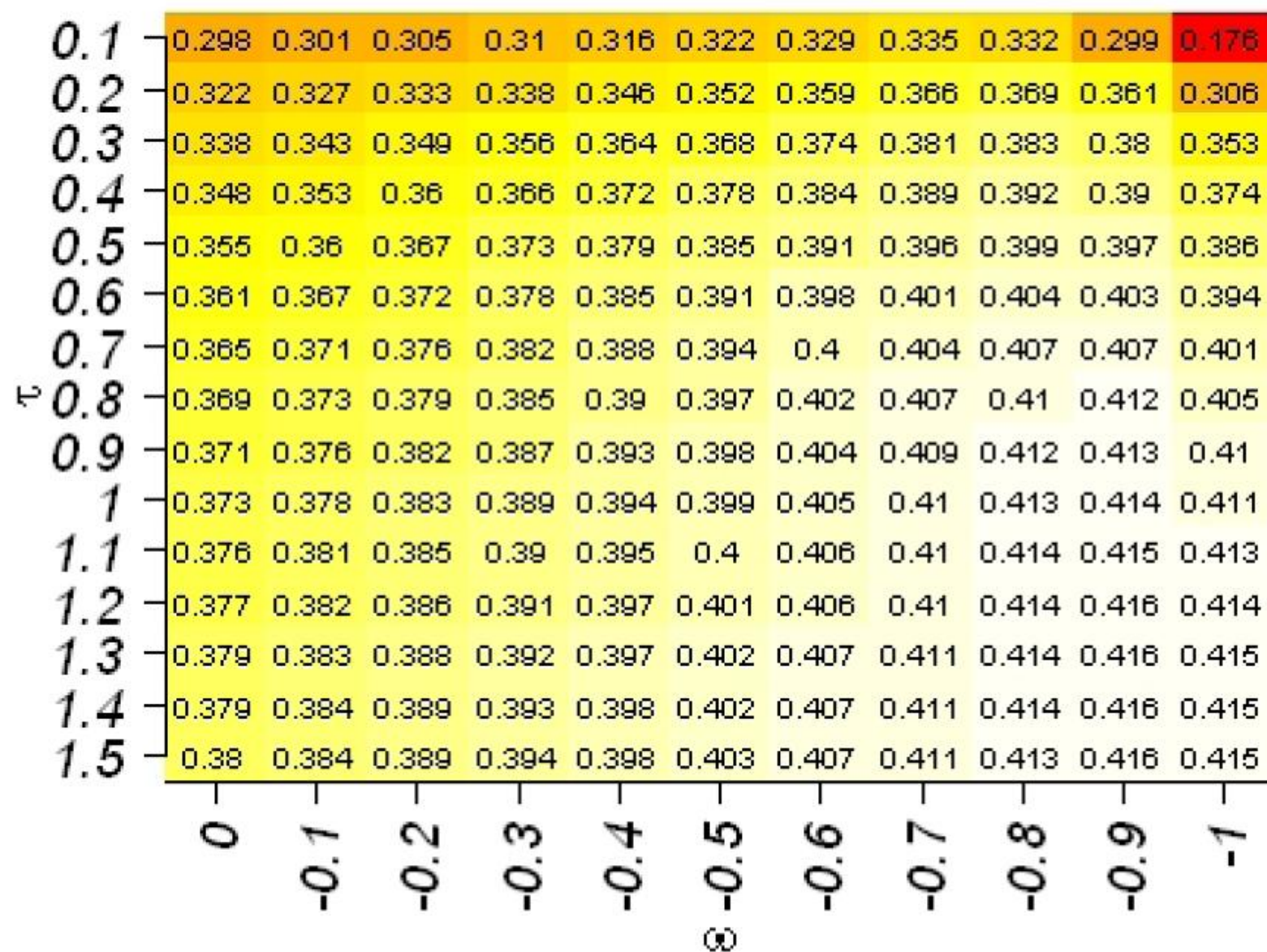
$$\text{Var}(\mathbf{u}_2) = \begin{bmatrix} \tau \mathbf{G}^{-1} + 1 - \omega & \mathbf{A}_{22}^{-1} \end{bmatrix}^{-1}.$$

$$\mathbf{u}_2 \mid \mathbf{A}_{22}, \mathbf{G} \sim \text{N}\left(0, \frac{\mathbf{G}}{\tau}\right) \quad \text{N}\left(0, \frac{\mathbf{A}_{22}}{1-\omega}\right)$$

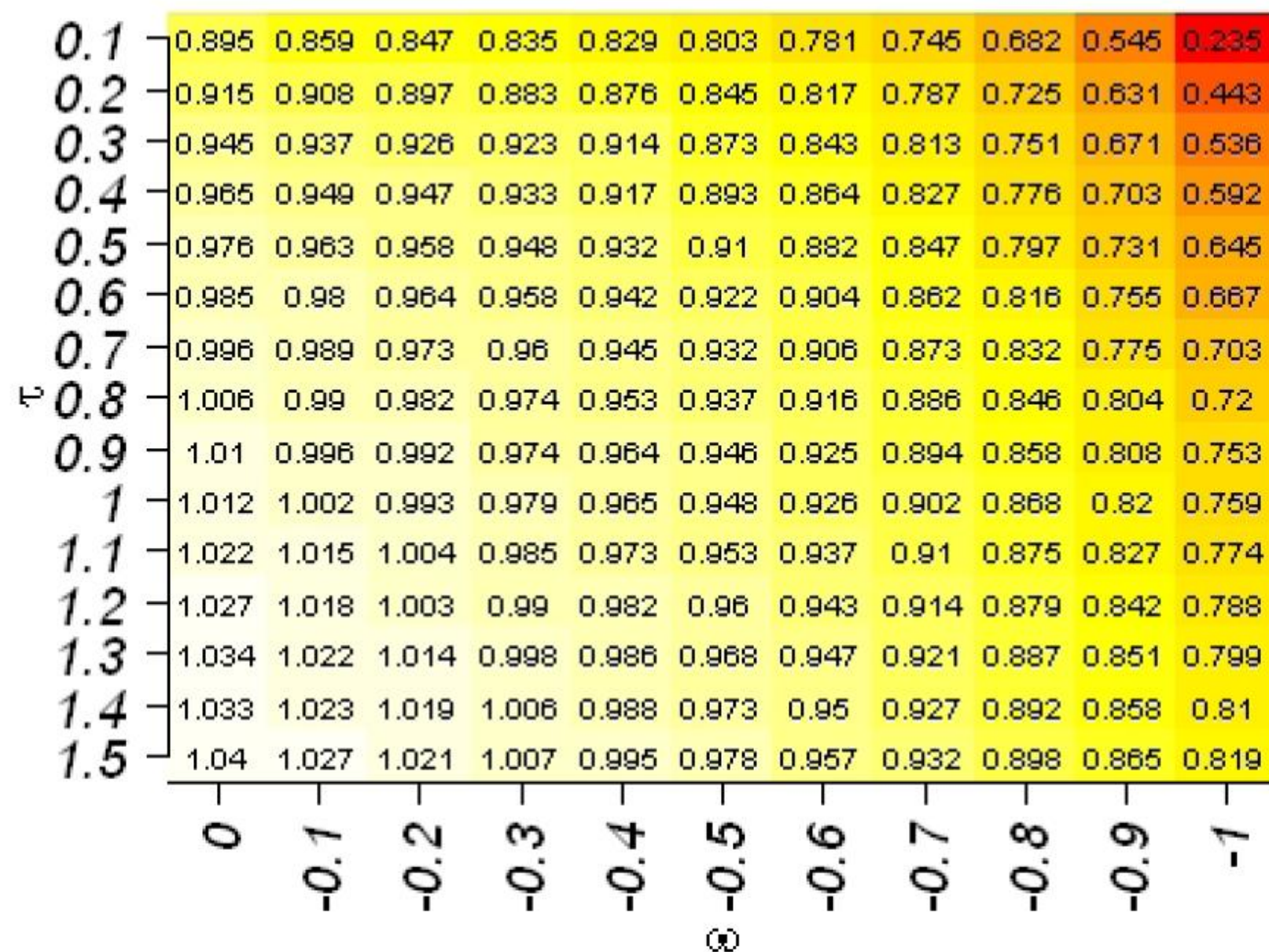
τ – scaling factor

ω – fraction of information from genomics

DYD R^2 as function of τ and ω



DYD Regression coefficient as function of τ and ω



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

T	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

EBV for young animals

$$u_i = \frac{u_s + u_d + \sum_{j, j \neq i} (-\tau g^{ij} + \omega a_{22}^{ij}) u_j}{2 + \tau g^{ii} - \omega a_{22}^{ii}} =$$

$$w_1 PA + w_2 GEBV - w_3 GPI$$

PA = Parent Average

GEBV = Genomic EBV

GPI = Parental Index for genotyped animals

Smaller ω reduces u_i

Large genomic information

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

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

If g^{ii} large, genomic predictions almost independent of many parameters

Conclusions

- R^2 affected little by wide variations of parameters (dairy only)
- Biases (inflation) strongly affected by weight on the parental index of genotyped bulls
- R^2 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

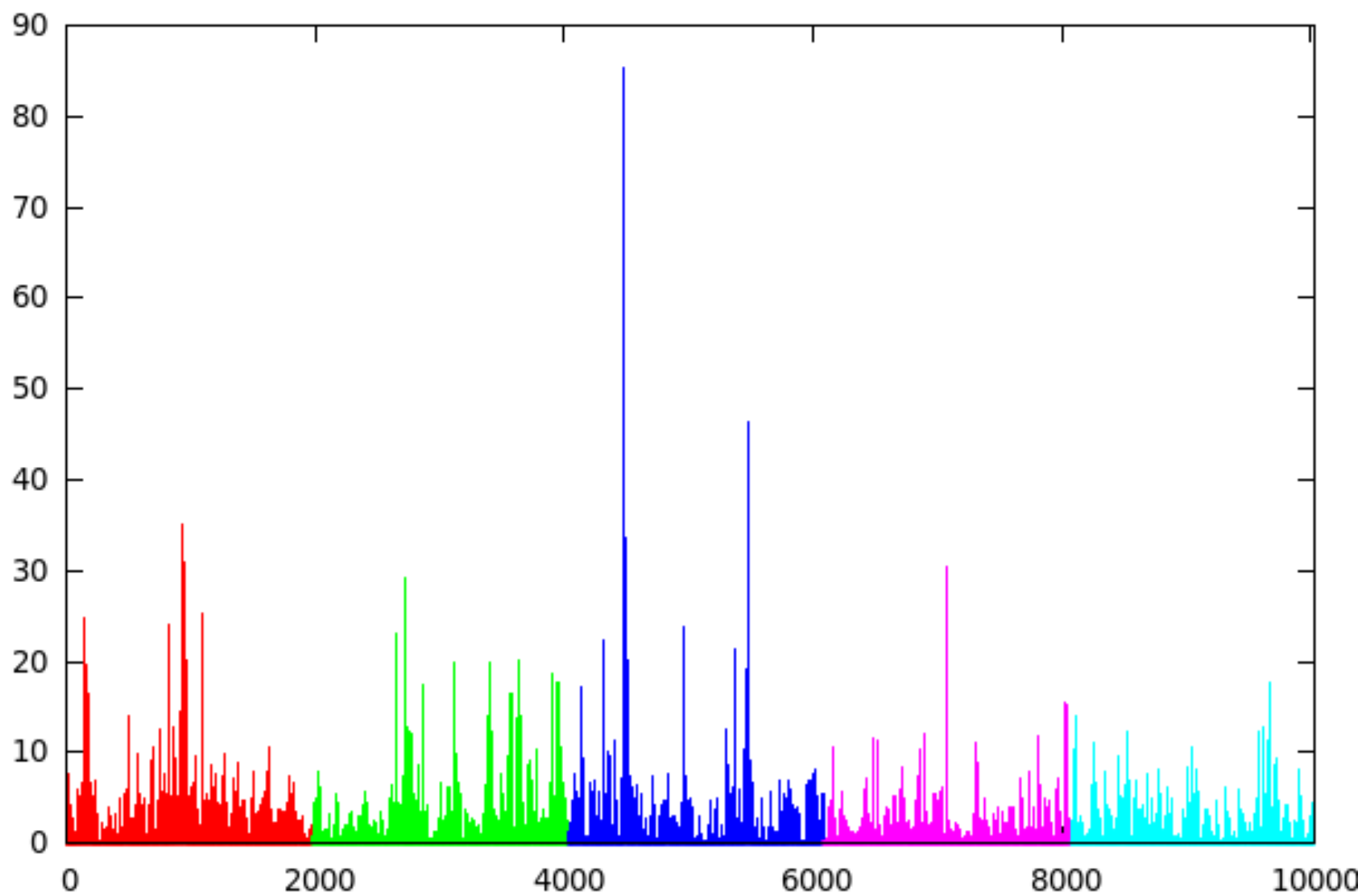
Holstein Association

Beef Cattle Consortium

Cobb-Vantress

Smithfield Premium
Genetics

PIC



-1074.54, -8.32335