

# Estimation of heritabilities and genetic correlations in very large datasets using predictivities within and across traits

Ignacy Misztal, Shogo Tsuruta, Zuleica Trujano, Mary Kate Hollifield, Gopal Gowane, Daniela Lourenco



UNIVERSITY OF  
**GEORGIA**

# Challenge of parameter estimation in genomic era

## Potential negative effects of genomic selection

Ignacy Misztal<sup>1</sup>  and Daniela Lourenco 

Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA

<sup>1</sup>Corresponding author: [ignacy@uga.edu](mailto:ignacy@uga.edu)

- Possibly rapidly changing parameters
- Need estimates using complete data including genomic
- Computing issues with REML and Bayesian methods
- Can we estimate parameters without size restriction, generation by generation?

# Realized and theoretical accuracies

Realized accuracy  $acc = corr(y - Xb, \hat{u})/h$

Legarra et al. (2008)

$y - Xb$  - adjusted phenotype

$\hat{u}$  - breeding value obtained without  
that phenotype

$h^2$  - heritability

Theoretical accuracy  $acc = \sqrt{\frac{Nh^2}{Nh^2 + M_e}}$

Daetwyler et al. (2008)

$N$  – number of genotyped animals with phenotypes

$M_e$  – number of independent chromosome segments

$M_e \approx 5k$  (chickens, pigs),  $10k$  (beef),  $15k$  (Holsteins)

Pocrnic et al. (2017)

# Formula for estimating heritability

$$\widehat{h^2} = \frac{c^2 + \sqrt{c^4 + 4c^2 M_e / N_{ref}}}{2} \mp \frac{3c}{\sqrt{N_{val}}}$$

$$c = \text{corr}(y - Xb, \hat{u})$$

$N_{ref}$  – animals in reference population

$M_e$  – effective chromosome segments, ~5k in pigs and chicken, ~15k in cattle

$N_{val}$  – number of animals in validation population

# Heritability for milk in Holsteins



**J. Dairy Sci. 104:5843–5853**

**<https://doi.org/10.3168/jds.2020-19789>**

© 2021, The Authors. Published by Elsevier Inc. and Fass Inc. on behalf of the American Dairy Science Association®.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## **Genomic predictions for yield traits in US Holsteins with unknown parent groups**

**A. Cesarani,<sup>1\*</sup>  Y. Masuda,<sup>1</sup>  S. Tsuruta,<sup>1</sup>  E. L. Nicolazzi,<sup>2</sup> P. M. VanRaden,<sup>3</sup>  D. Lourenco,<sup>1</sup>   
and I. Misztal<sup>1</sup> **

<sup>1</sup>Department of Animal and Dairy Science, University of Georgia, Athens 30602

<sup>2</sup>Council on Dairy Cattle Breeding, Bowie, MD 20716

<sup>3</sup>Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

Reference: 580k Validation 381k

Starting  $h^2 = 0.35$  Me=15k chromosome segments

Predictivity = 0.55

Calculated  $h^2 = 0.33$

# Formula for genetic correlations

$corr(y_i - Xb_i, \hat{u}_j)$  Predictivity of trait i by trait j

$$corr_{ij} = \frac{corr(y_i - Xb_i, \hat{u}_j)}{h_i acc_j} \mp \frac{1}{h_i acc_j \sqrt{N_{val}}}$$

Under correct model:  $corr_{ij} = corr_{ji}$

# Procedure

- Select reference population, number of genotyped  $N > 10,000$
- Select validation population, number of genotyped  $N_{\text{val}} > 5000$
- Estimate GEBV with phenotypes of reference population and genotypes of both populations; treat traits uncorrelated
- Calculate predictivities within and across traits

- Recalculate heritabilities 
$$\widehat{h^2} = \frac{c^2 + \sqrt{c^4 + 4c^2 M_e / N}}{2}$$

- Recalculate accuracies 
$$acc = \text{corr}(y - Xb, \hat{u}) / h$$

- Calculate genetic correlations 
$$\text{corr}_{ij} = \frac{\text{corr}(y_i - Xb_i, \hat{u}_j)}{h_i \text{ acc}_j} \pm \frac{1}{h_i \text{ acc}_j \sqrt{N}}$$

# Data simulation

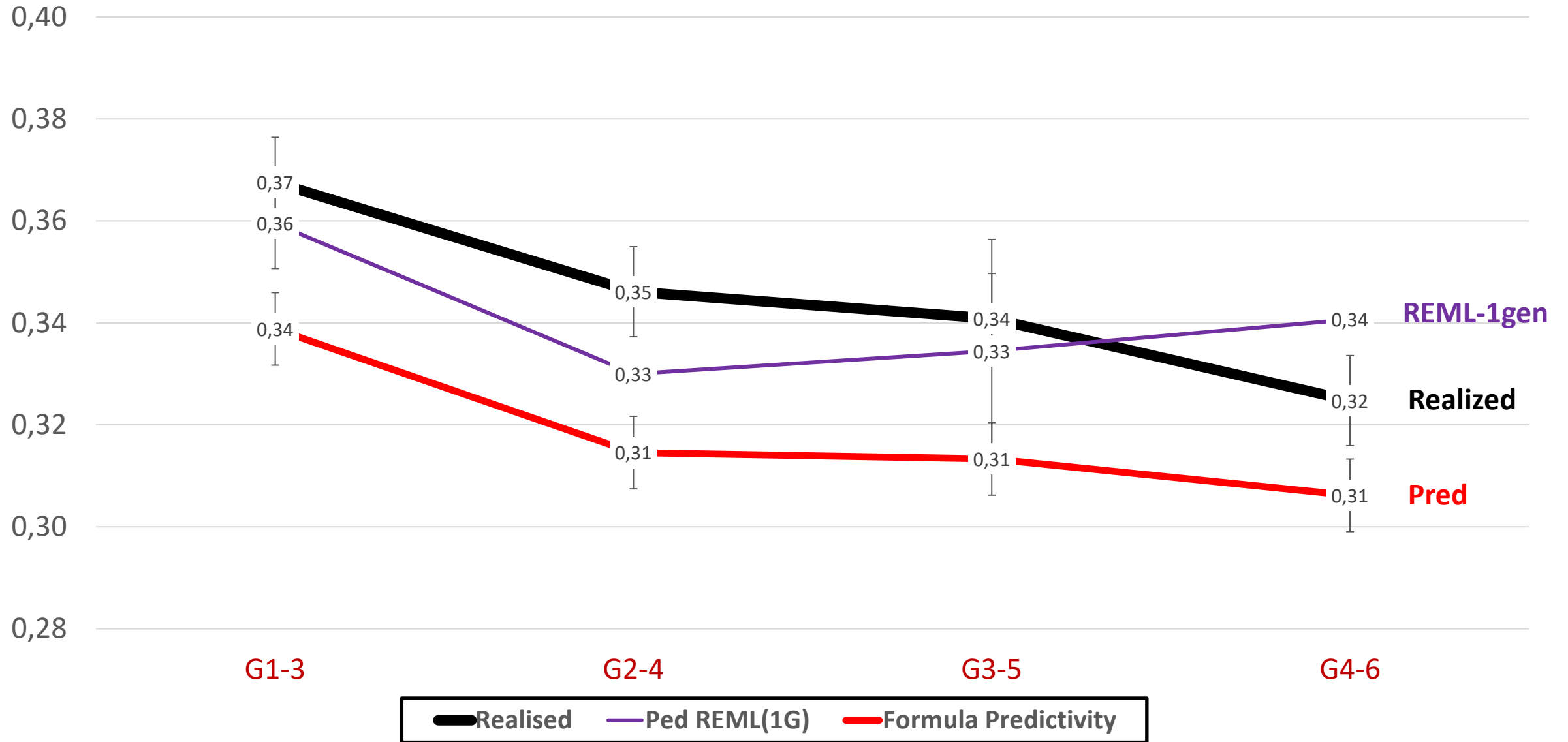
- $y_i^j, u_i^j$  - phenotype and breeding values of trait  $i$  in  $j$ -th generation
- Two uncorrelated traits
  - “production” with  $h^2 = 0.4$
  - “base fitness” with  $h^2 = 0.1$
- Evolving fitness trait 
$$u_3^j = \alpha_j \left( u_2^j - \beta u_1^j (\overline{u_1^j} - u_1^0) \right)$$
  - $\alpha_j$  - scaling factor so that  $var(u_3^j) = const$
  - $\beta$  - chosen to change genetic correlation between traits 1 and 3 by about -0.1 each generation.



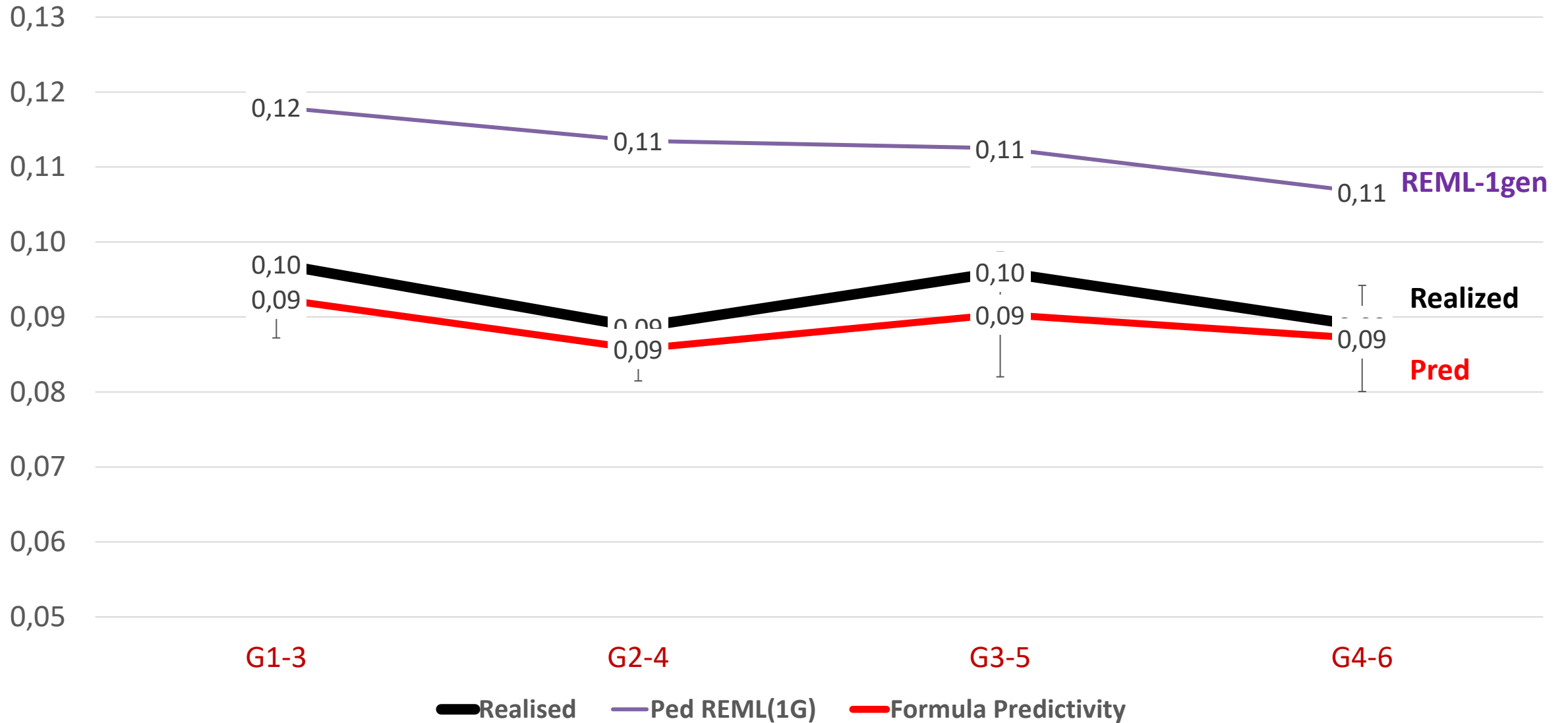
# Data simulation (2)

- Simulation by AlphaSim
  - 40k genotyped and phenotyped per generation
  - 6 generations
  - GBLUP selection in each generation
  - Effective population size 50
- Analyzes
  - Realized parameters for each generation
  - Pedigree REML
  - Parameters by predictivity, use 2 reference generations

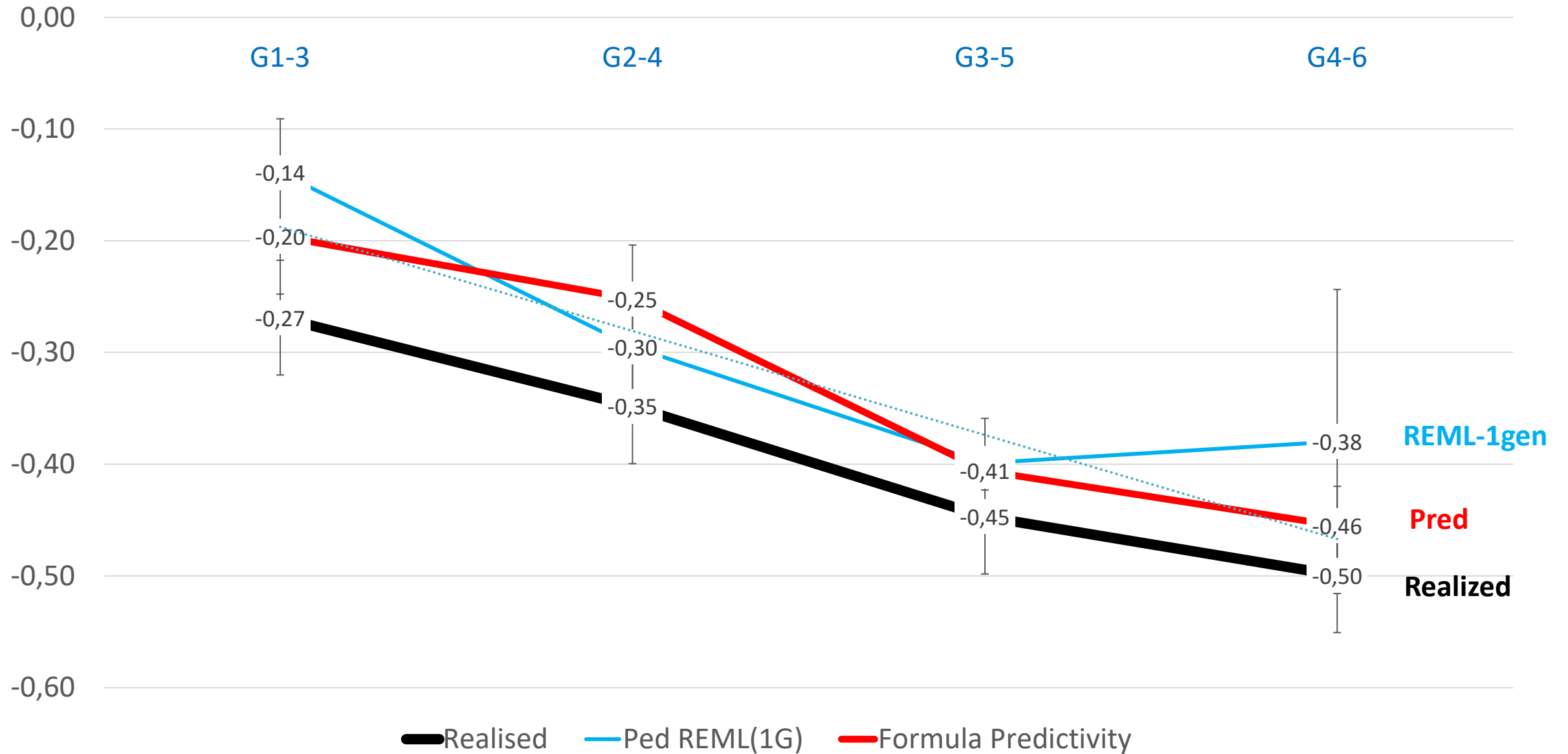
# Heritability for “Production” Trait



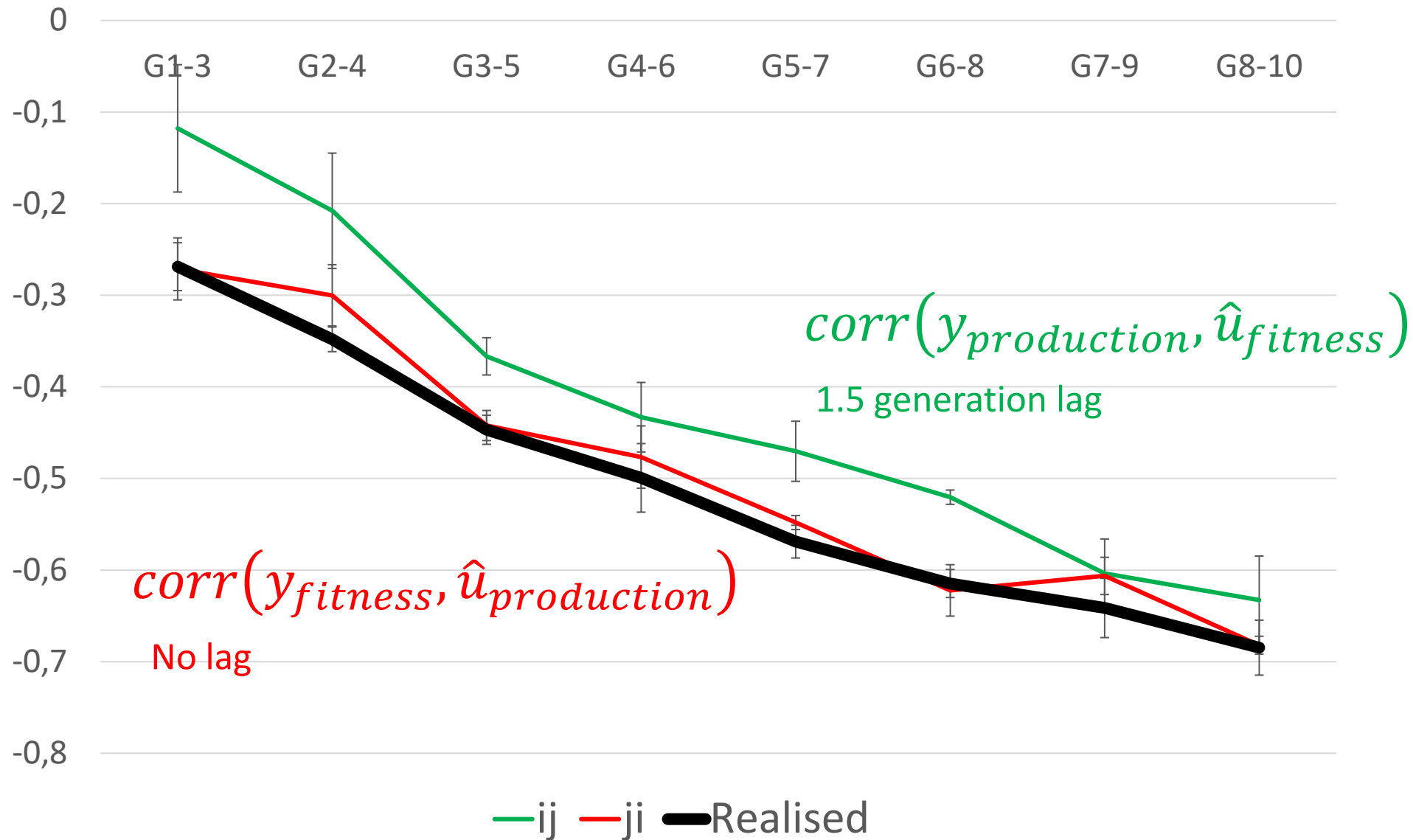
# Heritability for “Fitness” Trait



# Genetic correlations



# Two predictivities: IJ and JI



# Comments

- Results with pedigree REML OK because of simple selection
- Order of genetic correlation important
- Predictivity sensitive to selected genotyping – no problem with current commercial data
- ssGBLUP: use number of both genotyped and phenotyped

# Conclusions

- Can estimate heritability from theoretical and realized accuracies
- Can estimate genetic correlations by predictivity across traits
- Any data size and data slice
- Applicable to models where predictivity is applicable

# UGA AB&G team

