Prediction of indirect genetic effects with genomic information

Authors: **Bjarke Grove Poulsen**, O. F. Christensen, H. M. Nielsen, T. Ostersen, B. Ask

Session: Various topics in pig production

Conference: 70th EAAP, Ghent, Belgium

Contact: bgp@seges.dk

AARHUS UNIVERSITY

Miljø- og Fødevareministeriet

gudp

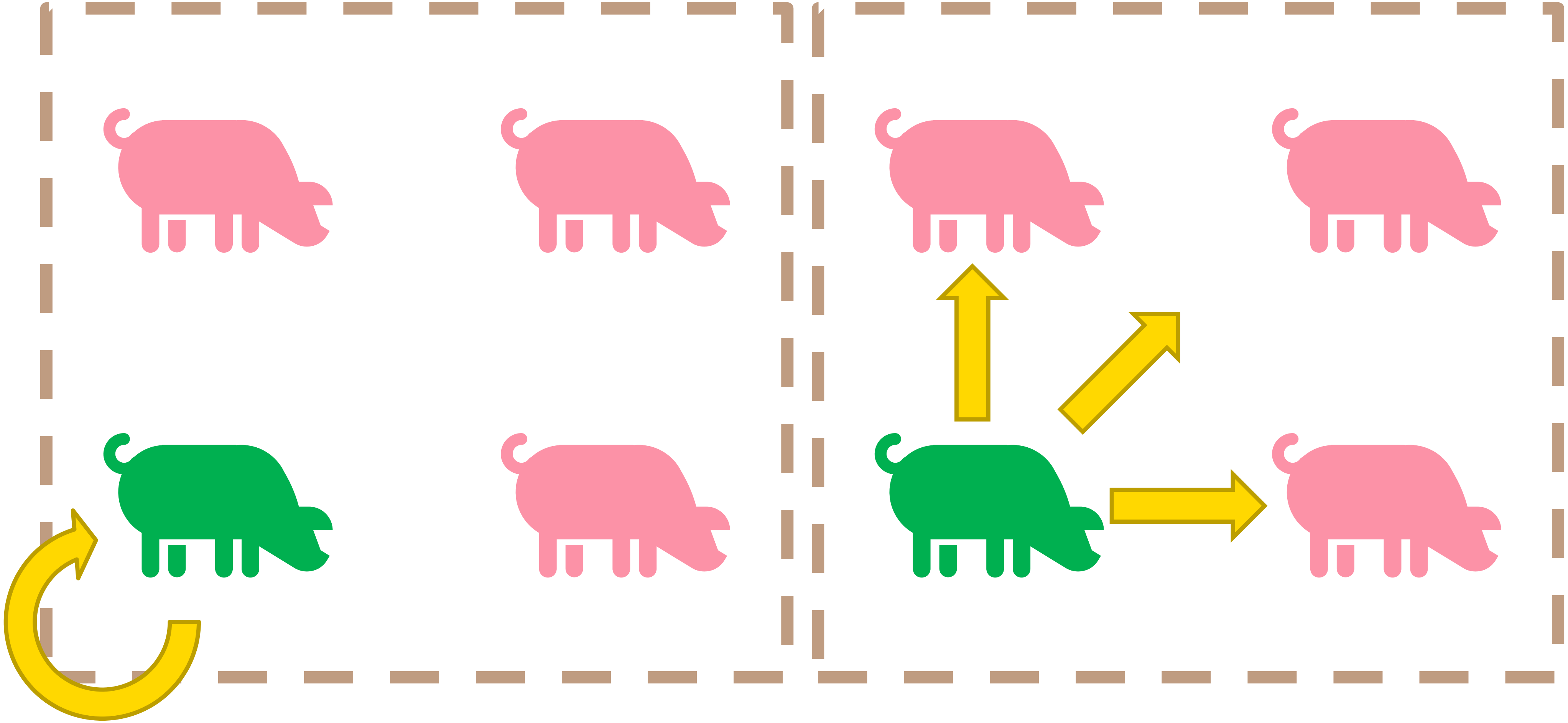
SEGES

Danish Pig Research Centre

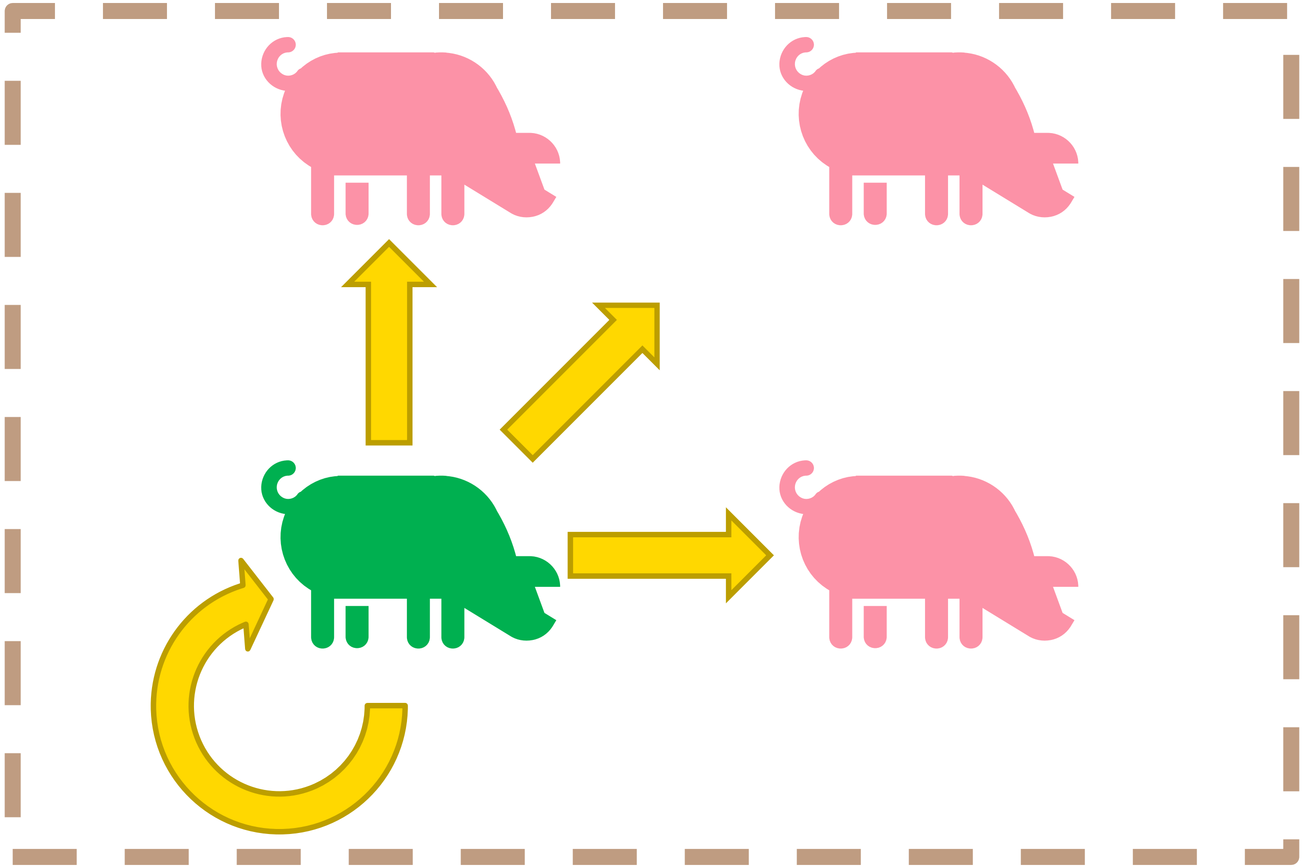
Spoiler alert

The results are unexpected!

Direct genetic effects and indirect genetic effects



Total genetic effects = Direct + (n-1) x Indirect



How total genetic effects are modelled

- The animal model with both direct- and indirect genetic effects:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}^{\mathrm{D}}\mathbf{a}^{\mathrm{D}} + \mathbf{Z}^{\mathrm{I}}\mathbf{a}^{\mathrm{I}} + \mathbf{R}^{\mathrm{I}}\mathbf{e}^{\mathrm{I}} + \mathbf{R}^{\mathrm{D}}\mathbf{e}^{\mathrm{D}}$$

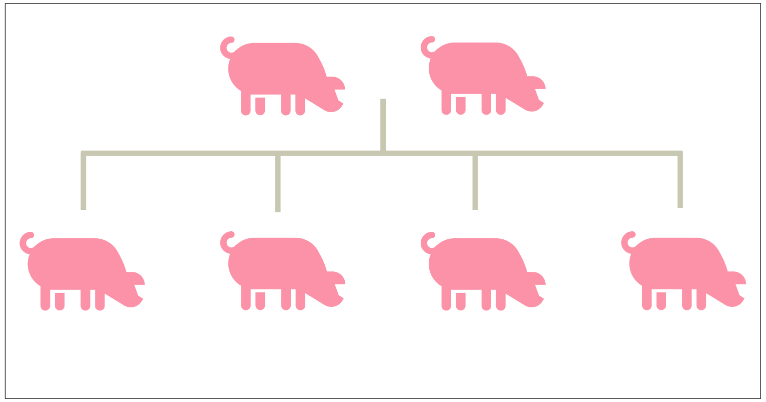
$$\begin{bmatrix} a^{\mathrm{D}} \\ a^{\mathrm{S}} \end{bmatrix} \sim N\left(\begin{bmatrix} a^{\mathrm{D}} \\ a^{\mathrm{S}} \end{bmatrix}, \left[\begin{bmatrix} \sigma_{\mathrm{d}}^2 & \sigma_{\mathrm{di}} \\ \sigma_{\mathrm{di}} & \sigma_{\mathrm{i}}^2 \end{bmatrix} \otimes \mathbf{C} \right] \right).$$

$\mathbf{a}^{\mathrm{D}}$	Direct genetic effects
$\mathbf{a}^{\mathrm{S}}$	Indirect genetic effects
$\mathbf{C}$	Relationship matrix

The Relationship Matrix

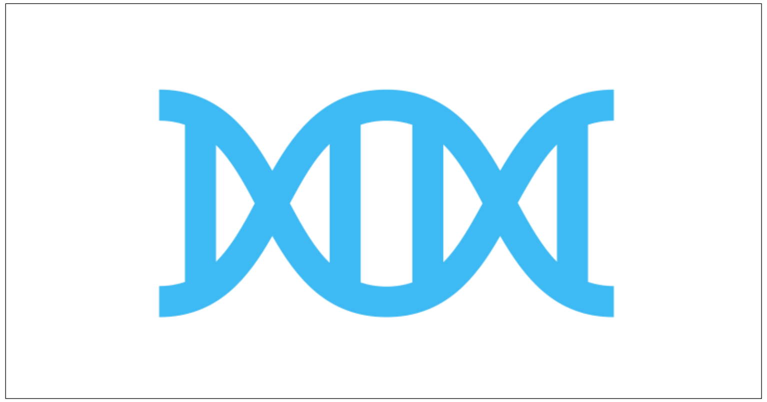
Previous studies on indirect genetic effects

- Numerator relationship matrix
- Pedigree-based
- Promising results!



The alternative

- Genomic relationship matrix
- SNP-based



Classic genomic animal models

- Don’t use genomic information for parameter estimation
- Use genomic information for prediction of genetic effects

SEGES Danish Pig Research Centre
AARHUS UNIVERSITY

Research Question

We know that genomic information improves prediction with the classical animal model…

Is the same true for indirect genetic models?

The Data

Variable	Description
Phenotype	Average daily gains between 30kg and 94kg
Group sizes	10 for boars and 9 for gilts
First/Last birth months of pigs	July 2015 – October 2018

# Number of	Amount
Pigs	11,420 pigs
Pigs with phenotype	11,255 pigs
Pigs with genotype information	10,998 pigs
Single Nuclear Polymorphisms (SNP)	34,123 SNP
Groups	1,179 groups
Herds	2 herds

SEGES Danish Pig Research Centre

AARHUS UNIVERSITY

The procedure

Two types of statistical models

- A classical animal model
- An indirect animal model

Variance components were estimated with pedigree information

Two types of relationship matrices

- A pedigree-based
- A combined pedigree and genomic (single-step)

The Single-Step Relationship Matrix

The genomic relationship matrix was defined as:

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{2\sum p_j\left(1 - p_j\right)}$$

The genomic relationship matrix was redefined:

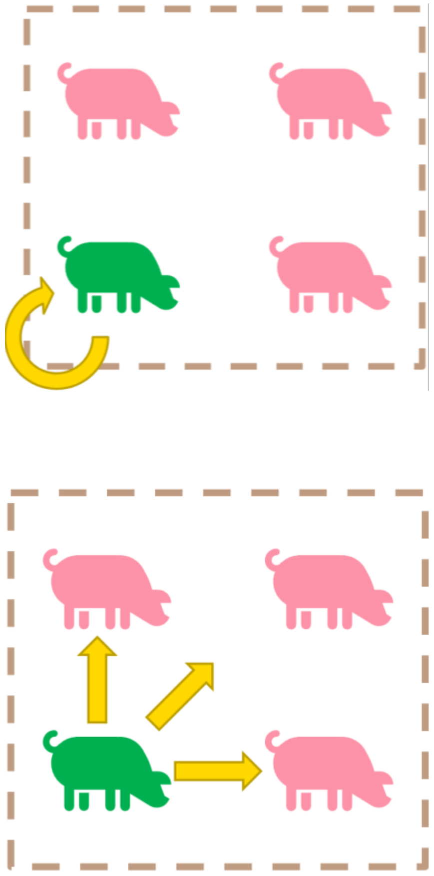
$$\mathbf{G}_{\mathrm{corrected}} = \mathbf{G}(\mathrm{w}) = \mathrm{w}\mathbf{G} + (1 - \mathrm{w})\mathbf{A}, \qquad w \in \{0, 0.05, 0.10 \ldots 0.99\}$$

The single-step relationship matrix was defined as:

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

The Indirect Statistical Model

	Fixed Effect	Random Effect
Direct	**Direct Fixed Effects** • Birth year • Birth month • Average pen mates pr. day • Start weight corrected for age	**Direct Random Effects** • Genetic • Litter • Herd-year-month • Residual
Indirect	**Indirect Fixed Effects** • Start weight as deviation from group mean corrected for date	**Indirect Random Effects** • Genetic • Litter • Residual



The Indirect Statistical Model

A relationship matrix

The assumed distribution for random effects:

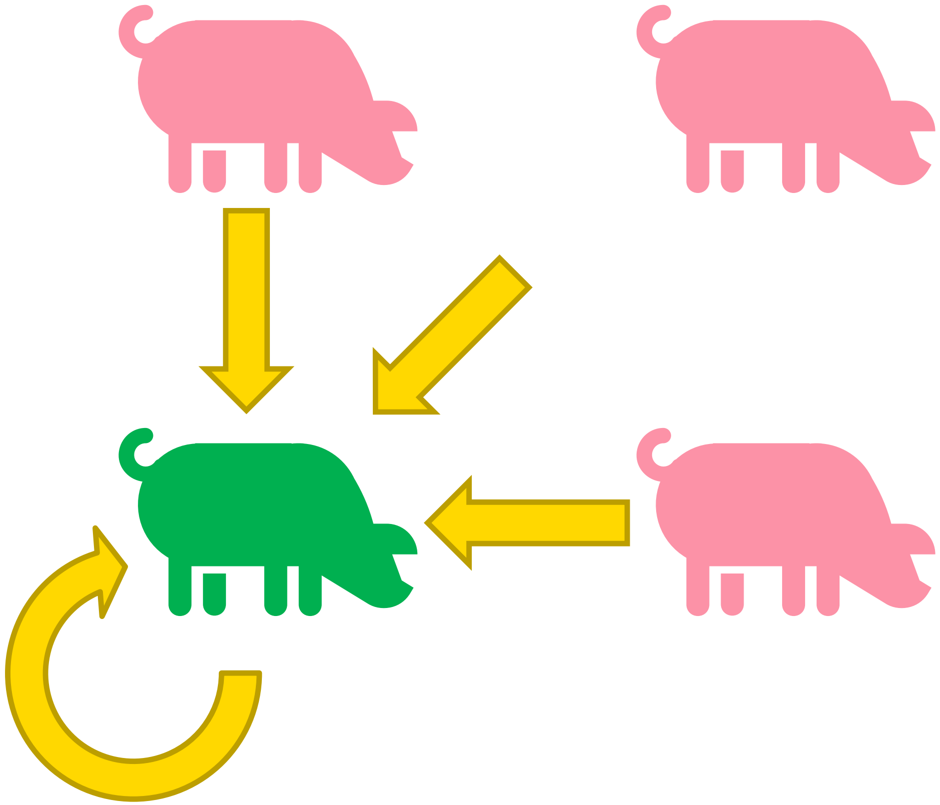
$$\begin{bmatrix} \boldsymbol{\mathit{Direct\ genetic}} \\ \boldsymbol{\mathit{Indirect\ genetic}} \\ \boldsymbol{\mathit{Herd - year - month}} \\ \boldsymbol{\mathit{Direct\ litter}} \\ \boldsymbol{\mathit{Indirect\ Litter}} \\ \boldsymbol{\mathit{Indirect\ animal}} \\ \boldsymbol{\mathit{Direct\ animal}} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \begin{bmatrix} \sigma_{\mathrm{a}}^2 & \sigma_{\mathrm{as}} \\ \sigma_{\mathrm{as}} & \sigma_{\mathrm{s}}^2 \end{bmatrix} \otimes \mathbf{C} & & & & & \\ \mathbf{0} & \sigma_{\mathrm{u}}^2 \otimes \mathbf{I}_{\mathrm{u}} & & & & \\ \mathbf{0} & \mathbf{0} & \sigma_{\mathrm{l}}^2 \otimes \mathbf{I}_{\mathrm{l}} & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{\mathrm{ls}}^2 \otimes \mathbf{I}_{\mathrm{ls}} & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{\mathrm{es}}^2 \otimes \mathbf{I}_{\mathrm{es}} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{\mathrm{e}}^2 \otimes \mathbf{I}_{\mathrm{e}} \end{bmatrix} \right)$$

The Training Data and Validation Data

- Training data: All groups with pigs born prior to January 1st, 2018.
- Validation data: All other groups of pigs

Training Data (83.8 %) **Validation Data (16.2 %)**

Prediction



Predictive performance of genetic effects

$$\mathbf{r}_i = correlation(\mathbf{y} - \mathbf{Xb}, \mathbf{g})$$

Two types of expected levels

1. Direct Genetic Level: $\mathbf{g} = \boldsymbol{a}_{Direct}$
2. Total Genetic Level: $\mathbf{g} = \boldsymbol{a}_{direct} + \sum_{1}^{n-1} \boldsymbol{a}_{indirect}$

eDGE = $\mathrm{r}_{Direct,ClassicModel}$

DGE = $\mathrm{r}_{Direct,IndirectModel}$

TGE = $\mathrm{r}_{Total,IndirectModel}$

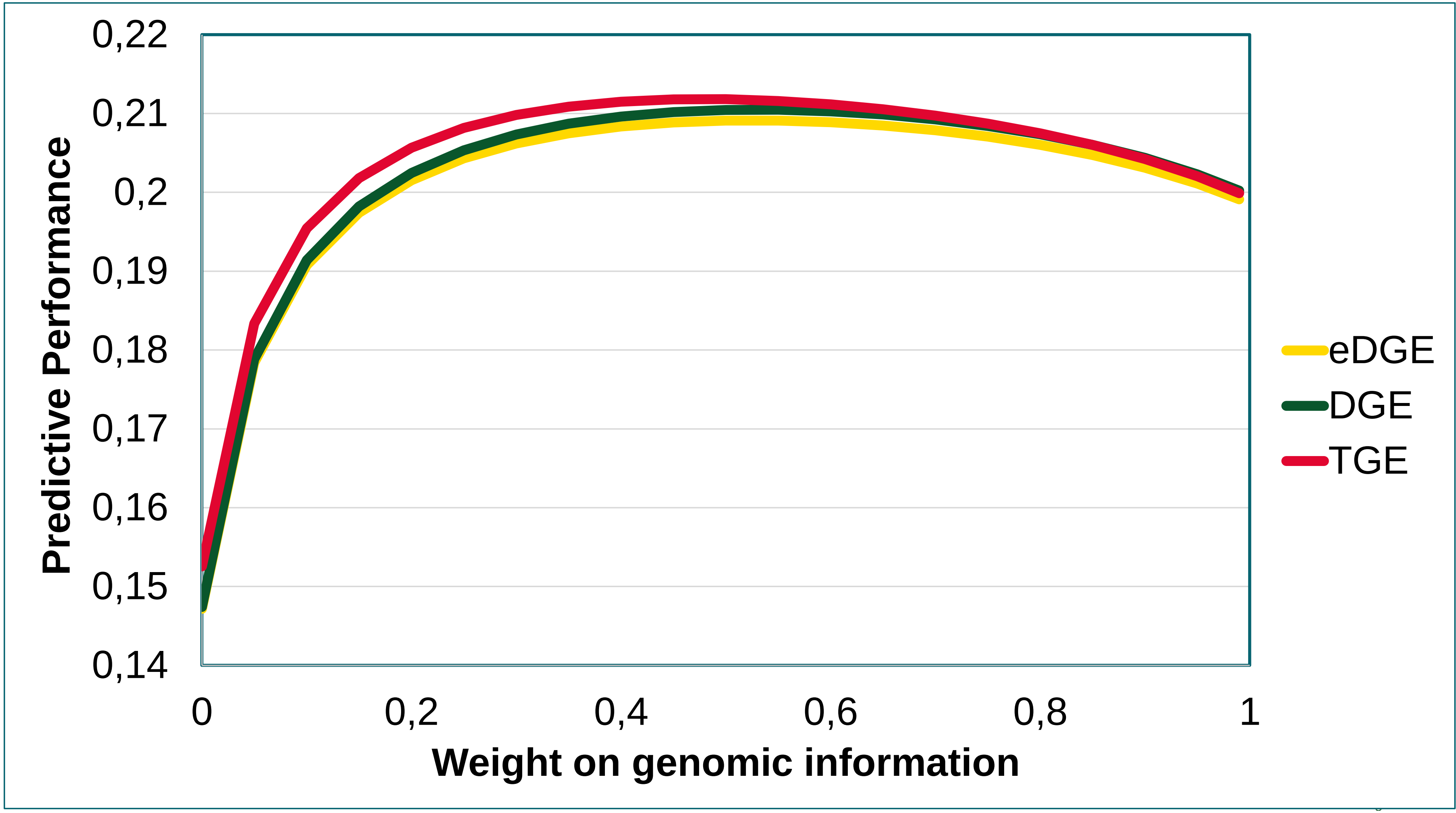
SEGES Danish Pig Research Centre

AARHUS UNIVERSITY

The indirect model had the best fit to the data

	CLASSIC	INDIRECT
$P(\Delta LogL)$		0.025

Prediction Results

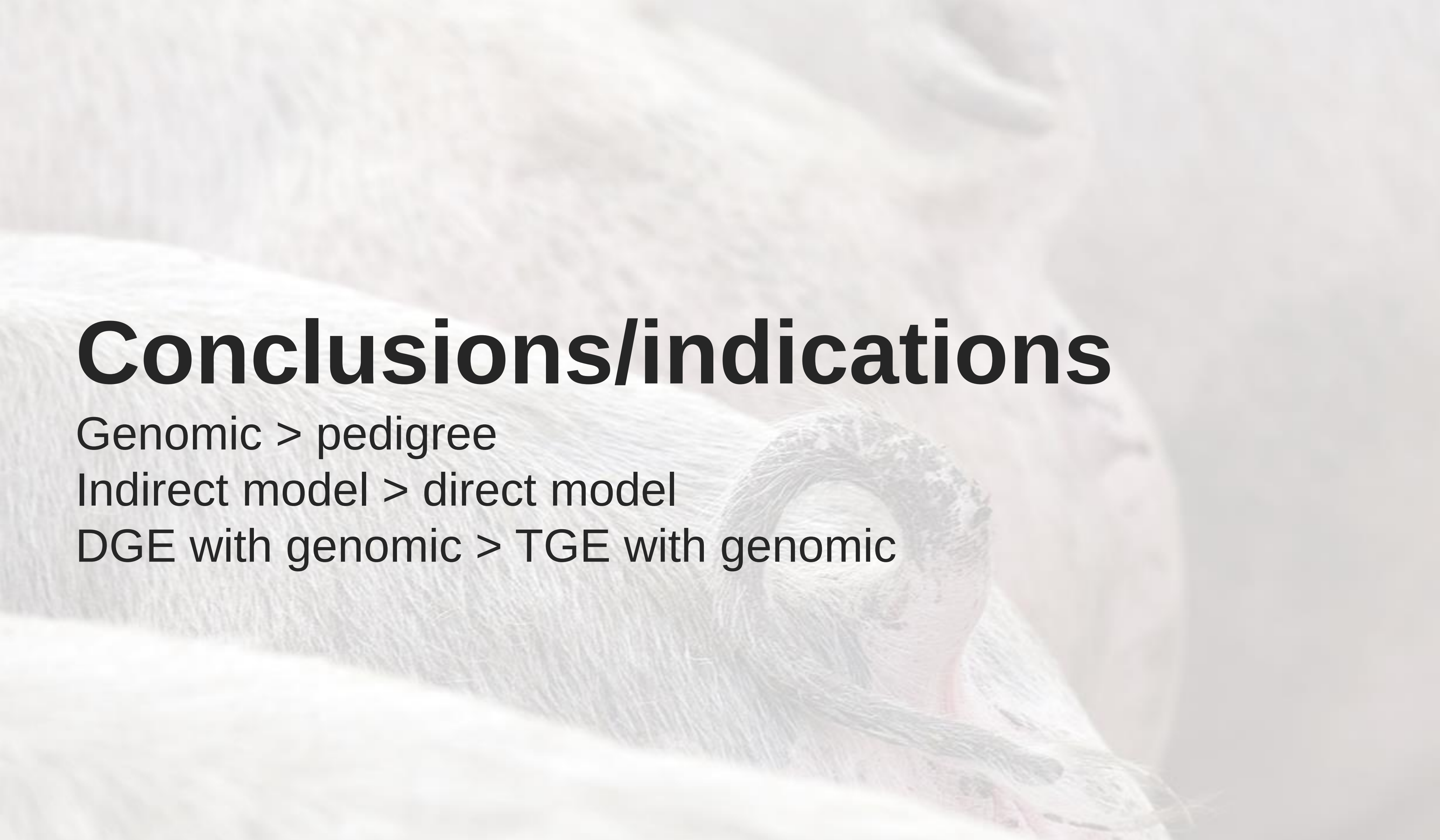


We don’t know why we see this pattern

Genomic prediction should be better than classic prediction for both types of genetic effects!

What makes genomic prediction favor direct genetic effects

- Pedigree information better explains indirect genetic effects?
- There are essential confounding effects that are unaccounted for?
    - Complicated interaction patterns
- Too little data?
    - Only two herd
    - 1K groups



Conclusions/indications

Genomic > pedigree

Indirect model > direct model

DGE with genomic > TGE with genomic

AARHUS UNIVERSITY

Miljø- og Fødevareministeriet

gudp

SEGES

Danish Pig Research Centre

The indirect model had the best fit to the data

	CLASSIC	INDIRECT
σ_a^2	3339(±450)	3349(±451)
σ_s^2		34.4(±16.3)
ρ_{as}		0.06(±0.18)
σ_u^2	951(±319)	908(±306)
σ_l^2	636(±131)	633(±131)
σ_e^2	10971(±289)	10989(±295)
σ_{ls}^2	37.7(±11.1)	19.1(±12.2)
σ_{es}^2	112.9(±18.1)	97.4(±21.8)
σ_p^2	17188(±420)	17597(±495)
σ_{tbv}^2		6201(±1694)
h^2	0.19(±0.02)	0.19(±0.02)
T^2		0.35(±0.09)
r	16.6 %	16.6 %
$-2LogL$	119074.46	119067.12
$P(\Delta LogL)$		0.025