Interpretation of dominant and additive variances from genomic models

Z. G. Vitezica*, L. Varona†, A. Legarra††

* Université de Toulouse, INP, UMR 1289, 31326 Castanet-Tolosan, France
† Facultad de Veterinaria, Universidad de Zaragoza, 50013 Zaragoza, Spain
†† INRA, UR 631, SAGA, 31326 Castanet-Tolosan, France

zulma.vitezica@ensat.fr

Introduction

Genomic evaluation models typically fit only additive effects.

Dominance is of theoretical and practical interest:

- crosses,
- mating allocations.

Dominance effects have rarely been included in pedigree-based genetic evaluations. Genomic evaluations have renewed the interest in dominance (e.g., Toro and Varona, 2010; Wellmann and Bennewitz, 2012; Su et al., 2012).

Objectives

To show the equivalences between additive and dominant effects at the marker and the population levels.

To present how to compute from genotypes the covariances between individuals due to dominant deviations: **D**, Dominant genomic relationship matrix.

Theory

A model with additive and dominant **SNP effects**:

Additive effect

Dominant effect

$$y_i = \mu + \sum_{j=1}^{n} t_{ij} a_j + \sum_{j=1}^{n} x_{ij} d_j + e_i$$

is equal to {**1, 0, -1**} for {A_1A_1, A_1A_2, A_2A_2} genotypes

is equal to {**0, 1, 0**} for {A_1A_1, A_1A_2, A_2A_2} genotypes

In matrix form for a set of individuals,

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Ta} + \mathbf{Xd} + \mathbf{e}$$

Theory

Additive or **"breeding" values** (u) of individuals are generated by substitution effects (α) (Falconer, 1981)

The α involve both "biological" additive (a) and dominant (d) effects of the markers and the allelic frequency p

$$\alpha = a + d(q - p)$$

Dominance deviations (v) only include part of the biological dominant effects of the markers

Breeding values

The breeding value (u) for an individual is

$$u_{A_1A_1} = (2-2p)(a+d(q-p)) = (2-2p)\alpha$$

$$u_{A_1A_2} = (1-2p)(a+d(q-p)) = (1-2p)\alpha$$

$$u_{A_2A_2} = (-2p)(a+d(q-p)) = (-2p)\alpha$$

Substitution effect of the SNP

Additive effect of the SNP

Dominant effect of the SNP

$$\mathbf{u} = \mathbf{Z}\alpha \longrightarrow z_i = \begin{cases} (2-2p) \\ (1-2p) \\ -2p \end{cases} \text{ for genotypes } \begin{cases} A_1A_1 \\ A_1A_2 \\ A_2A_2 \end{cases}$$

VanRaden, 2008

Dominant deviations

Also, the dominant deviation (***v***) of an individual is

$$v_{A_1A_1} = -2q^2d$$

$$v_{A_1A_2} = 2pqd \quad \leftarrow \text{Dominant effect of the SNP}$$

$$v_{A_2A_2} = -2p^2d$$

So, the dominant deviations of a set of individuals are

$$\mathbf{v} = \mathbf{W}d \quad \rightarrow \quad w_i = \begin{cases} -2q^2 \\ 2pq \\ -2p^2 \end{cases} \text{ for genotypes } \begin{cases} A_1A_1 \\ A_1A_2 \\ A_2A_2 \end{cases}$$

Additive and dominance variances

The partition of the total variance $\sigma^2_G = \sigma^2_A + \sigma^2_D$

$\sigma^2_A = 2pq\alpha^2$ $\sigma^2_D = [2pqd]^2$

If *a* and *d* are considered random, the covariance of additive individual effects, ***u***, is

$$Cov(\mathbf{u}) = \frac{\mathbf{ZZ'}}{2\sum p_i q_i}\sigma^2_A = \mathbf{G}\sigma^2_A$$

G is the genomic additive relationship matrix

with $\sigma^2_A = 2\sum p_i q_i \sigma^2_a + 2\sum p_i q_i (q_i - p_i)^2 \sigma^2_d$

VanRaden, 2008

SNP variances for additive and dominant components

Genomic dominant relationship matrix

Also, the covariances across dominant deviations (***v***) are

$$Cov(\mathbf{v}) = \frac{\mathbf{WW}'}{\sum (2p_i q_i)^2}\sigma_D^2 = \mathbf{D}\sigma_D^2$$

D is the dominant genomic relationship matrix

As we have $\sigma_D^2 = \sum (2p_i q_i)^2 \sigma_d^2$ **SNP variance** for dominant component

Use in Mixed Model: DBLUP, GDBLUP

The models

Su et al. (2012) presented an alternative parameterization based on genotypic values of the individuals. We call this “genotypic” model.

« Breeding » model

« Genotypic » model

	A_1A_1	A_1A_2	A_2A_2
W	$-2q^2$	$2pq$	$-2p^2$
H	$-2pq$	$1-2pq$	$-2pq$

$$Cov(\mathbf{v}) = \frac{\mathbf{WW'}}{\sum(2p_iq_i)^2}\sigma_D^2 = \mathbf{D}\sigma_D^2$$

$$Cov(\mathbf{v}^*) = \frac{\mathbf{HH'}}{2\sum p_iq_i(1-2p_iq_i)}\sigma_{D^*}^2$$

These variances are different

The covariances are also different

It’s possible to go from genotypic to breeding model using the total genetic variance

The models

The **“breeding”** (or classical) and the **“genotypic”** models are equivalent models to explain the data (y) but their interpretation is different.

The **“breeding”** model is expressed in termed of breeding values and dominant deviations;
The “genotypic” model, in additive and dominant genotypic values.

σ_A^2 from the **“breeding”** model is the variance useful in selection and comparable with pedigree-based estimates.

Simulation data

- As in Toro and Varona, 2010 (Ne=100)
- 9,000 markers + 1,000 QTLs
- 2,100 individuals
- Var(A)=20, Var(D)=10, total phenotypic variance=100
- Results were the mean of 10 replicates

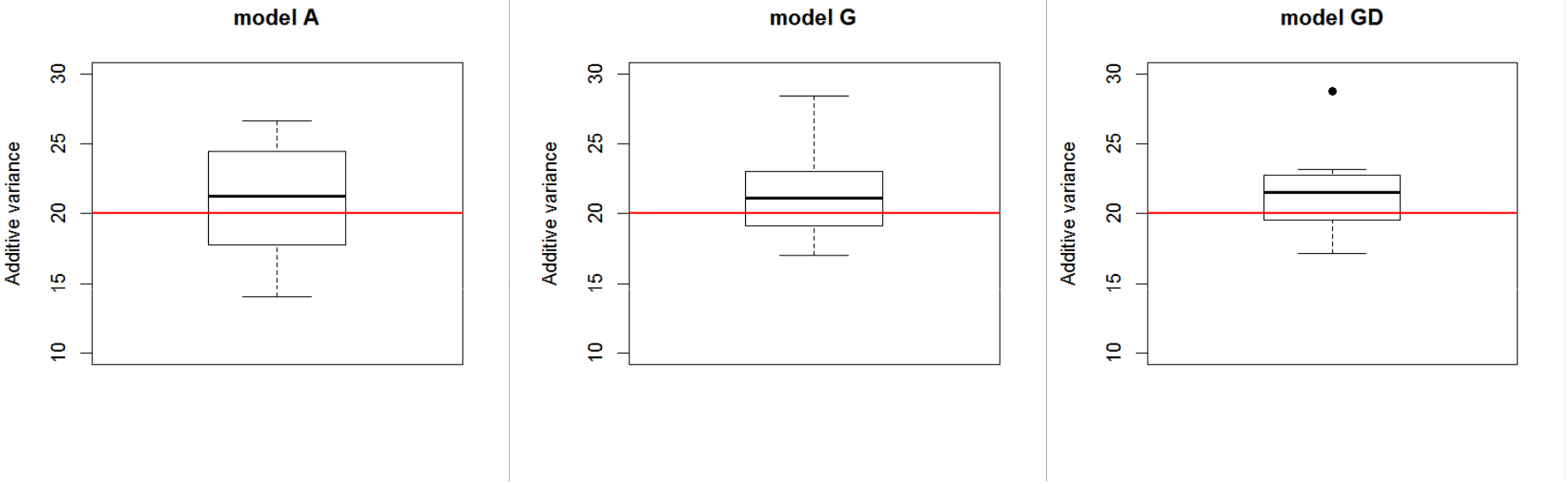
Estimation of variance components by REML (remlf90)

Model **A :** **pedigree relationships**

Model **G:** **Genomic additive relationships**

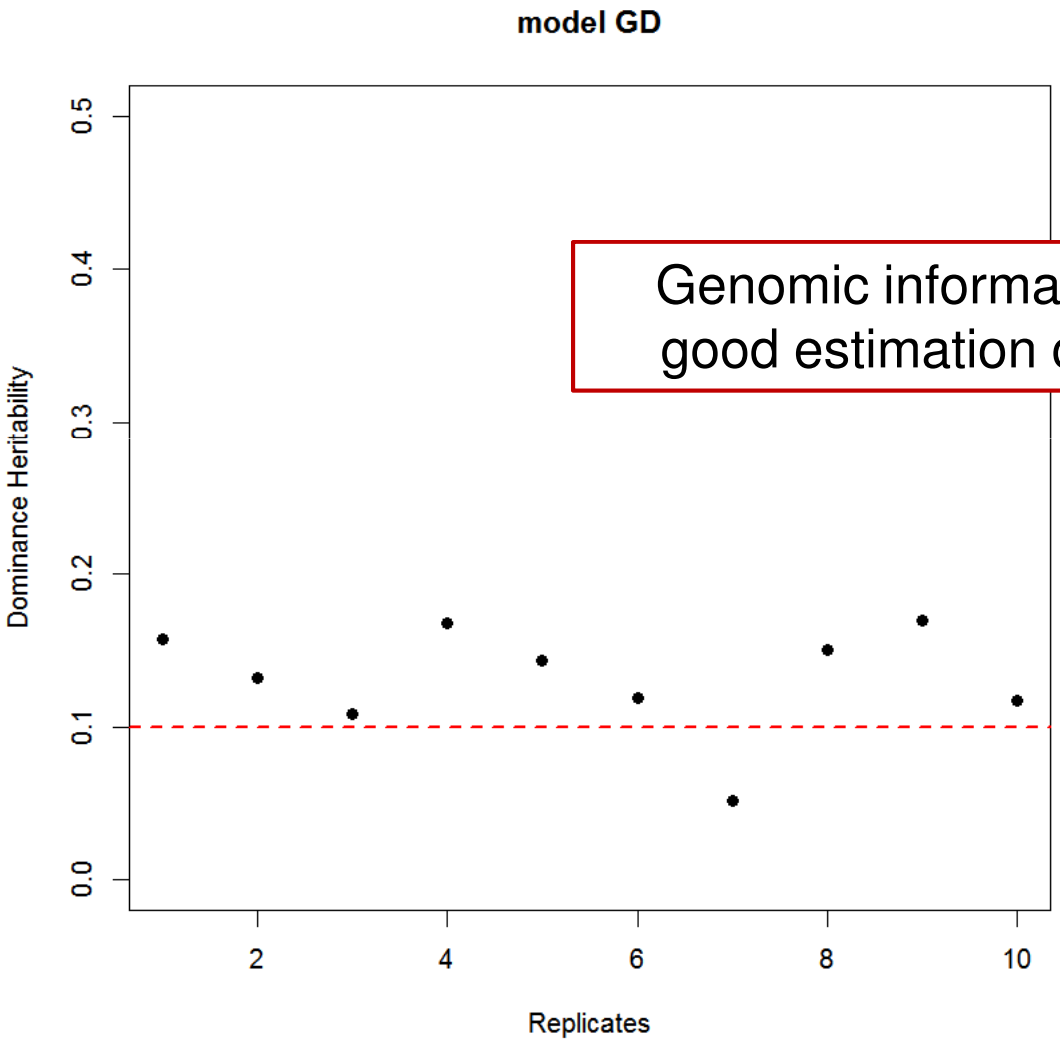
Model **GD:** **Genomic additive and Dominant relationships**

Results



The **additive genetic variance** was well estimated also when the model included dominant effects

Results



Genomic information allows obtaining a good estimation of **dominant variance**

Conclusions

The parameterization in terms of breeding values and substitution effects is more adequate than other parameterization for selection

Models using genomic additive and dominant relationships can estimate variance components correctly

Genomic models including dominance have the advantage that they provide a simple framework, compared with pedigree models

Acknowledgements

Financed by the INRA SELGEN Metaprogram, projet X-GEN and AGL2010-15903.

Project partially supported by the platform bioinformatics Toulouse Midi-Pyrenees.

Thank you for your attention ...

http://www.flickr.com/photos/16872979@N06

Theory

Remember (as in Falconer 1981)

Consider one locus. Following model (1) the genotypic value *G* of an individual is as follows:

$$G_{A_1A_1} = a \qquad G_{A_1A_2} = d \qquad G_{A_2A_2} = -a$$

where the values *a* and *d* are deviations from the midpoint of the two homozygous.

The genetic mean is therefore

$$E(G) = (p - q)a + 2pqd$$

where *p* is the frequency of A_1 and $q = 1\text{-}p$.

The substitution effect of the gene/marker is $\alpha = a + d(q - p)$

Results

Table 1. Accuracies (SDs) and inflations (SDs) computed from true and estimated breeding values for different effects and prediction models

Effect	Accuracy	Inflation
Additive		
Model A	0.58 (0.04)	0.98 (0.15)
Model G	0.68 (0.02)	0.96 (0.08)
Model GD	0.69 (0.02)	0.96 (0.08)
Model ADped	0.58 (0.03)	0.99 (0.15)
Dominance		
Model GD	0.44 (0.03)	0.91 (0.26)
Model ADped	0.32 (0.03)	1.23 (1.28)

Compared with A and ADped model, all genomic prediction methods (models G and GD) increased **accuracy** by about 18 % and 32% for additive and dominance effects, respectively.