

Assortative Mating

Teresa de Candia

University of Colorado at Boulder

Assortative Mating (AM)

- Occurs when individuals choose mates that resemble them
- Height: spousal phenotypic correlation ~ 0.2 .
- If a trait is heritable, then breeding values will be correlated across spouses. Assuming spousal $r \sim 0.2$ and $h^2 \sim .7$, breeding values $r \sim .14$
- For polygenic traits, individual with an increasing allele at locus x will be more likely to mate with an individual with an increasing allele at that or any other causal locus
- This **leads to gene effects being correlated within offspring**
- This, in turn, leads to increases of actual genetic (and total phenotypic) variance in the population

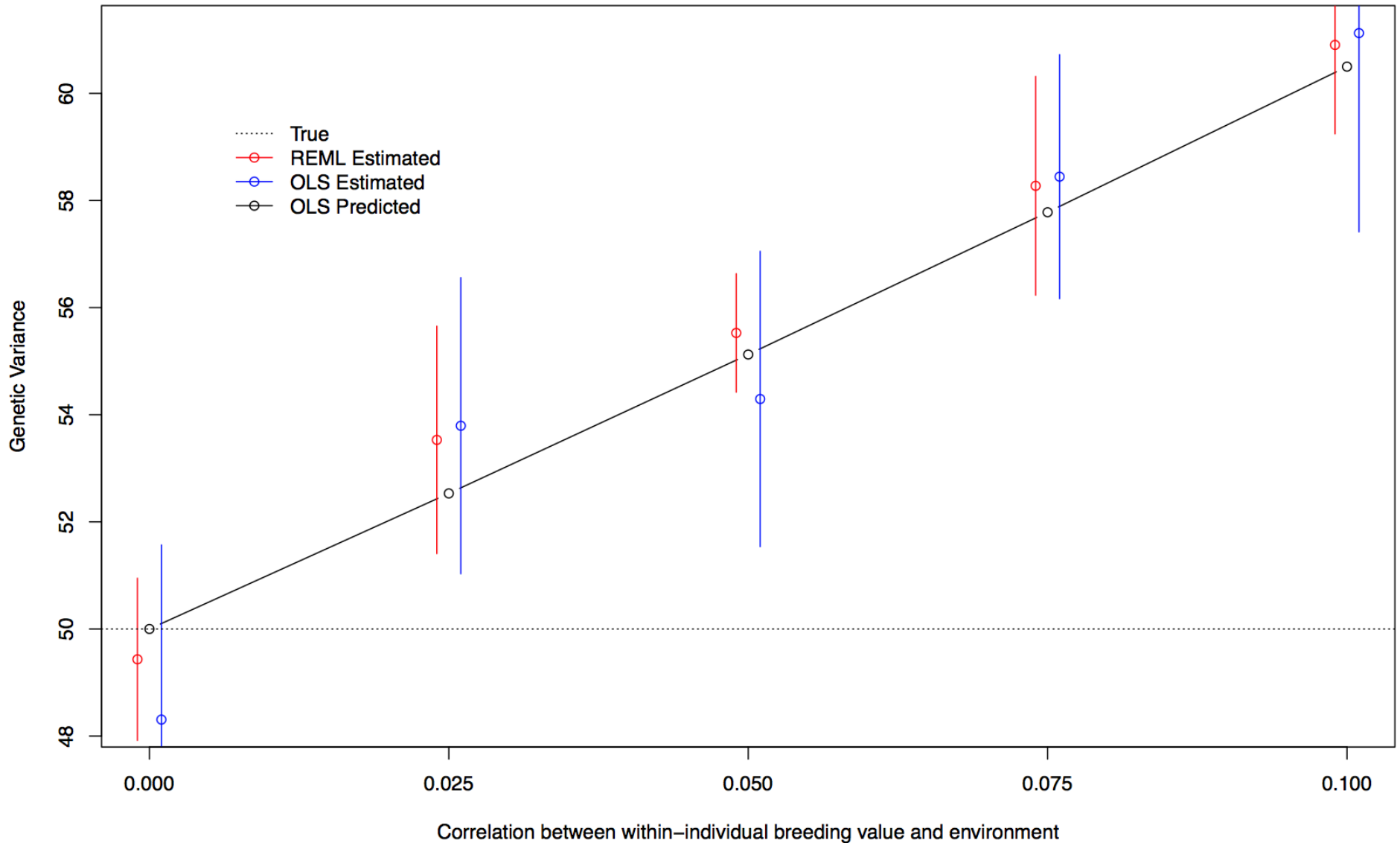
Broken Assumptions

- As long as gene effects are correlated with one another or with environmental effects within individuals, assumptions of GCTA, HE regression, and other such models do not hold

Broken Assumptions

- As long as gene effects are correlated with one another or with environmental effects within individuals, assumptions of GCTA, HE regression, and other such models do not hold
- **rGE within individuals:** For a special case of assortative mating in which there is *vertical cultural transmission*, GRM is confounded with unmodeled environmental similarity. This confound results in overestimates of genetic variance. This is similar to confounds occurring when environmental effects correlate with genes in samples with stratification or cryptic relatedness.

Using Simulated Genes and Phenotypes: Overestimates due to rGE (not specific to AM)



Broken Assumptions

- As long as gene effects are correlated with one another or with environmental effects within individuals, assumptions of GCTA, HE regression (HE-r), and other such models do not hold
- **rGE within individuals: rGE within individuals:** For a special case of assortative mating in which there is *vertical cultural transmission*, GRM is confounded with unmodeled environmental similarity. This confound results in overestimates of genetic variance. This is similar to confounds occurring when environmental effects correlate with genes in samples with stratification or cryptic relatedness.
- **rGG within individuals:** Assortative mating for heritable traits induces correlations among genetic effects. Parametrization of GRM does not properly account for these. This results in variance across pi-hats that is too low compared with variance across phenotypic cross-products

Broken Assumptions



- As long as gene effects are correlated with one another or with environmental effects within individuals, assumptions of GCTA, HE regression (HE-r), and other such models do not hold
- **rGE within individuals: rGE within individuals:** For a special case of assortative mating in which there is *vertical cultural transmission*, GRM is confounded with unmodeled environmental similarity. This confound results in overestimates of genetic variance. This is similar to confounds occurring when environmental effects correlate with genes in samples with stratification or cryptic relatedness.
- **rGG within individuals:** Assortative mating for heritable traits induces correlations among genetic effects. Parametrization of GRM does not properly account for these. This results in variance across pi-hats that is too low compared with variance across phenotypic cross-products

Pi-hats compared with cross-products of breeding values

- Let's assume a polygenic trait under AM with a heritability of 1 and equal amounts of variance explained by each CV. Individual i will have phenotype $CV_{1i} + CV_{2i} + CV_{3i} + \dots$ and

$$P_{ij} = (CV_{1i} + CV_{2i} + CV_{3i} + \dots) * (CV_{1j} + CV_{2j} + CV_{3j} + \dots)$$

$$= \sum CV_i \sum CV_j$$

$$\Pi_{ij} = CV_{1i}CV_{1j} + CV_{2i}CV_{2j} + CV_{3i}CV_{3j} + \dots = \sum CV_i CV_j$$

- For each pair of alleles m and n ,

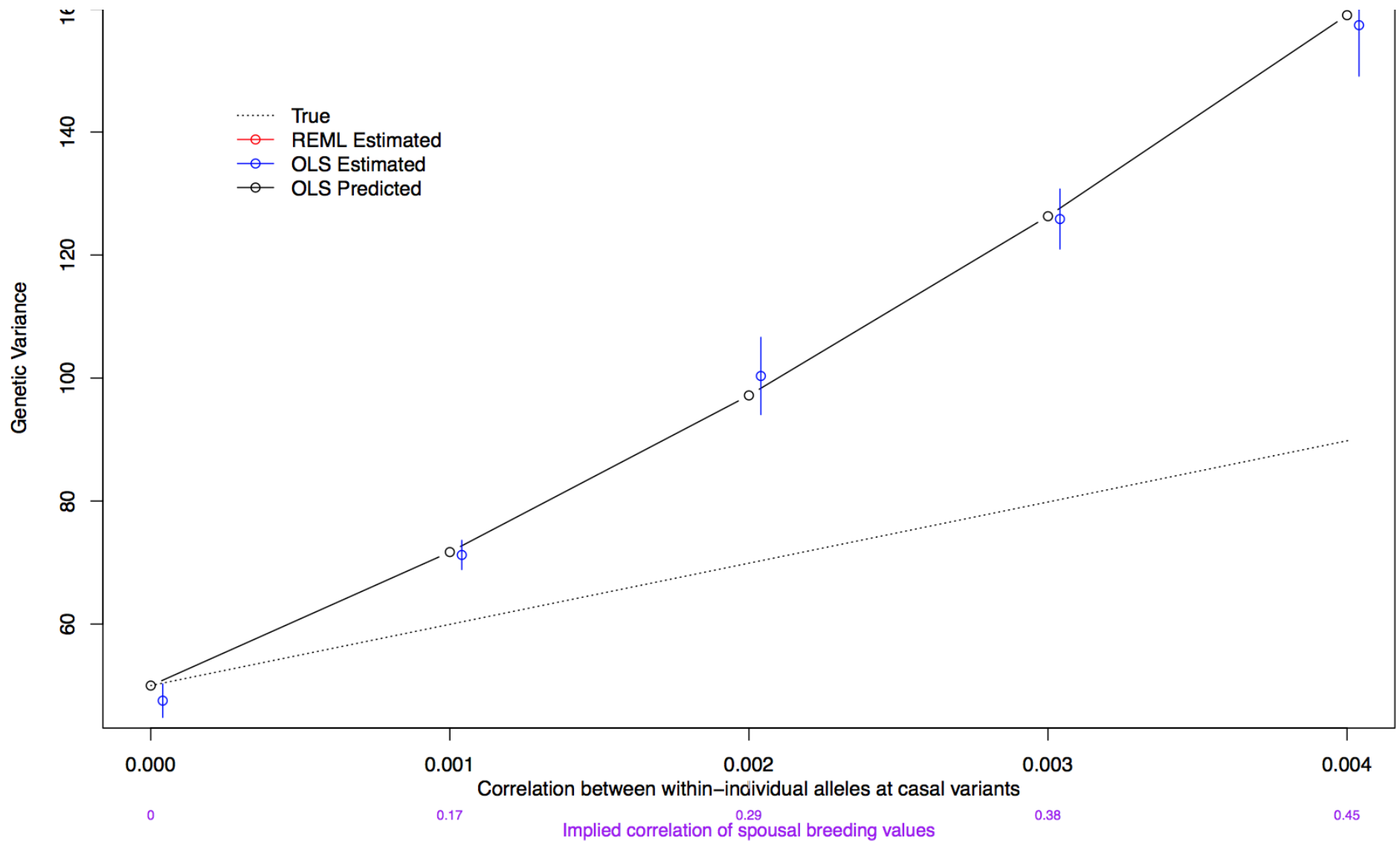
$$\text{If } r(CV_{mi}, CV_{ni}) = 0, \text{ then } V(\sum CV_i \sum CV_j) = V(\sum CV_i CV_j)$$

$$\text{Else, } V(\sum CV_i \sum CV_j) > V(\sum CV_i CV_j)$$

- By quantifying this relationship (at equilibrium), we can predict overestimates expected from fitting OLS

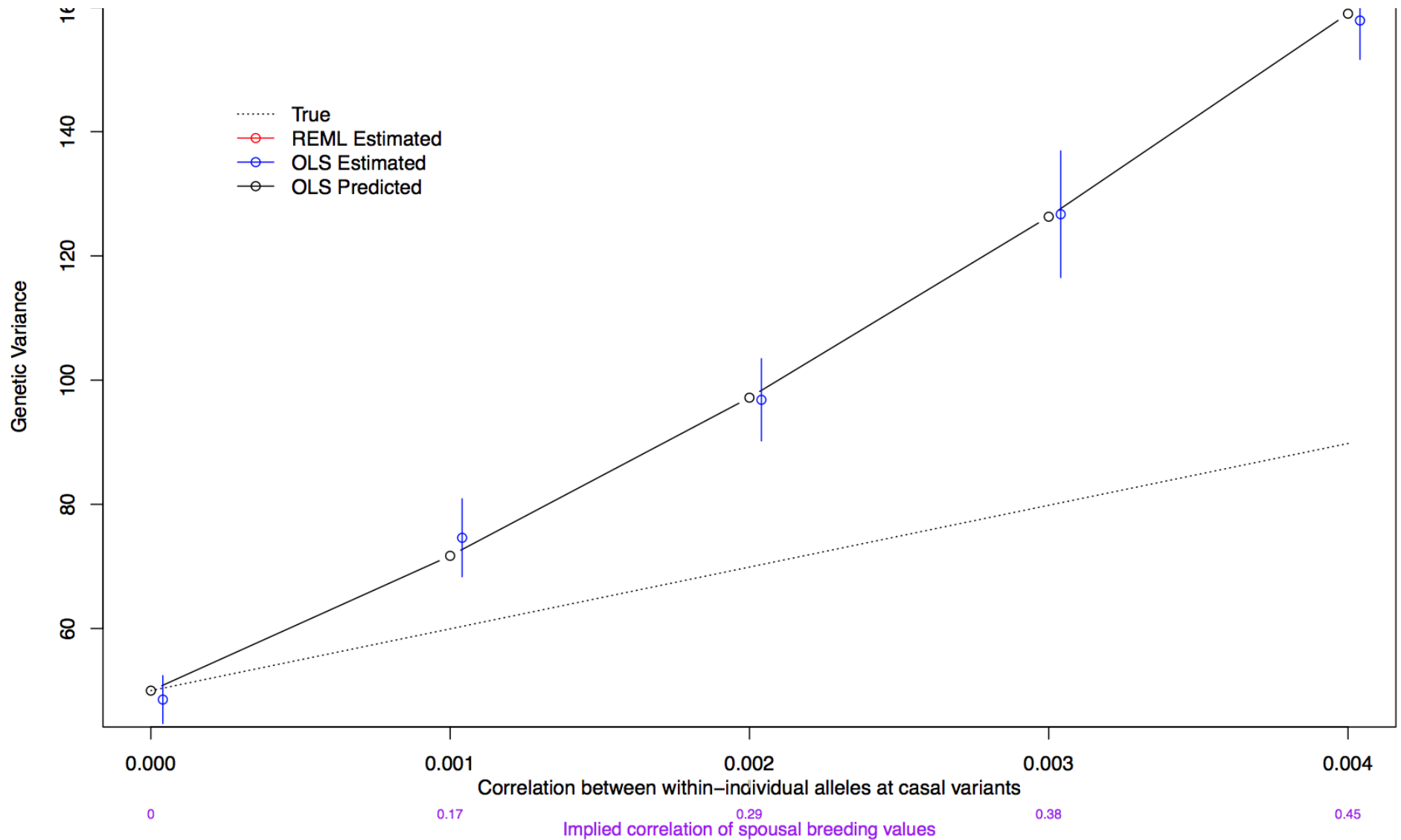
Using Simulated Genes and Phenotypes: OLS (HE-r) Overestimates Predictable

CVs=100, other markers=0



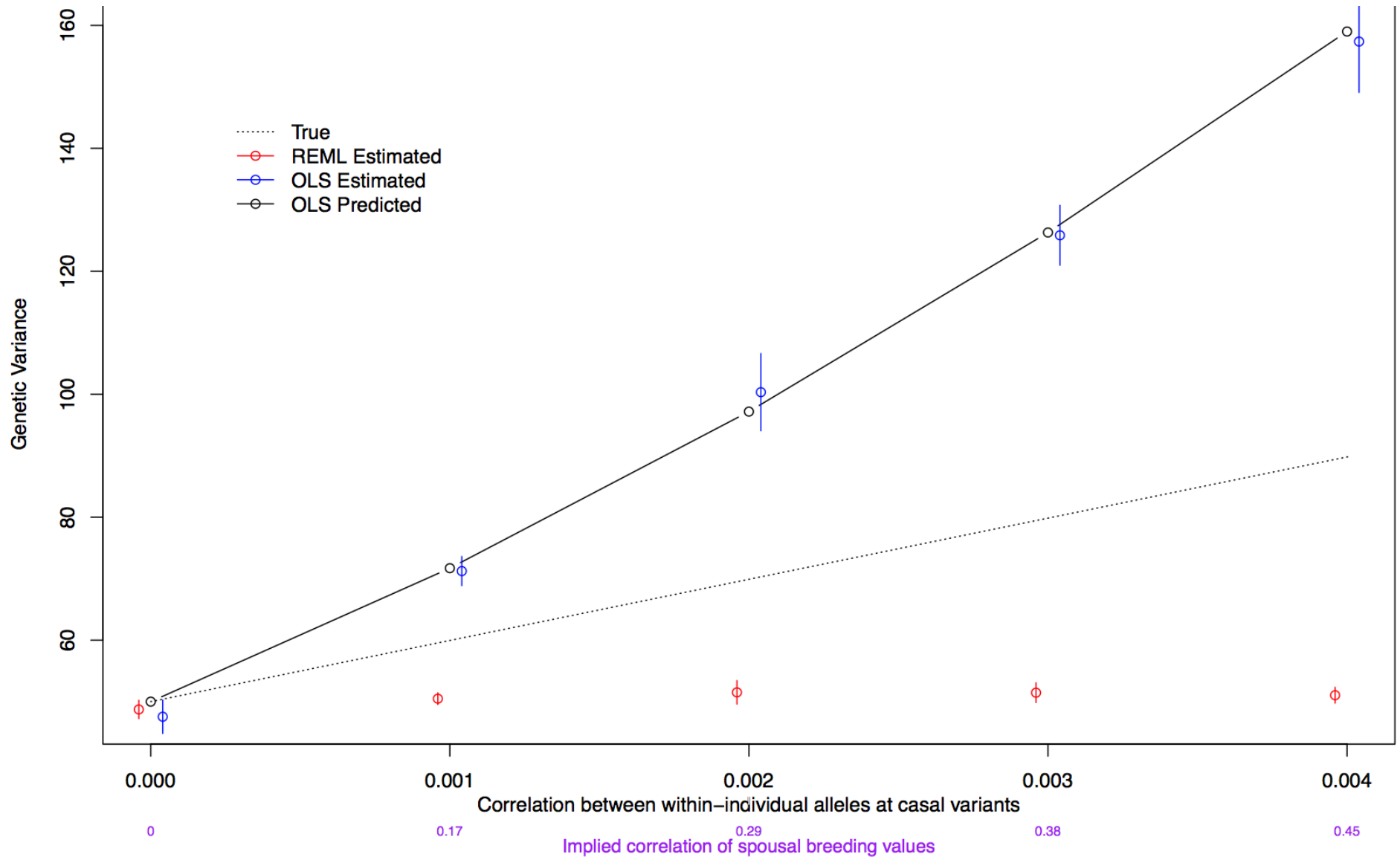
Using Simulated Genes and Phenotypes: OLS (HE-r) Overestimates Predictable

CVs=100, other markers=10k

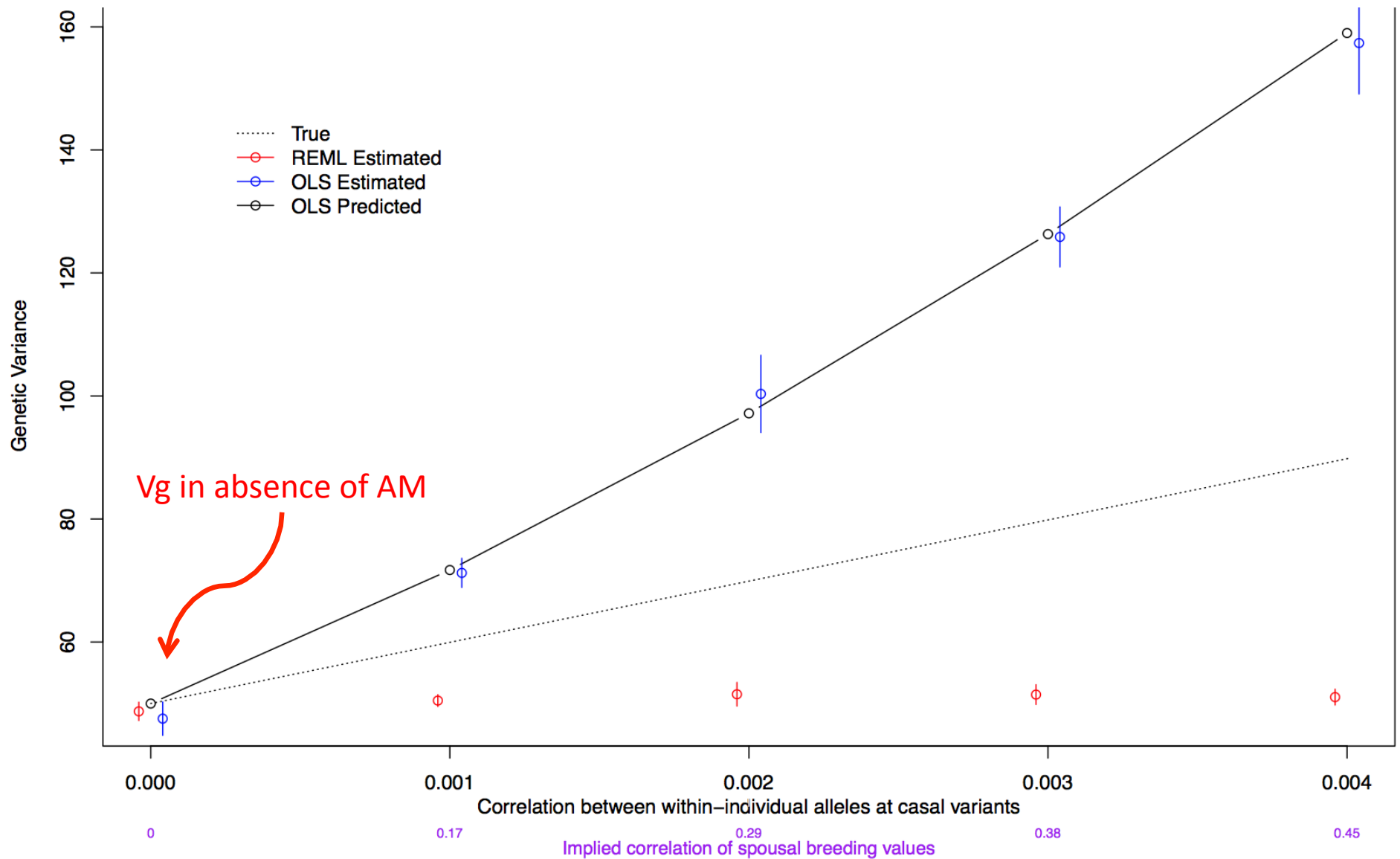


Using Simulated Genes and Phenotypes: REML/FIML (e.g., GCTA) Estimates

CVs=100, other markers=0

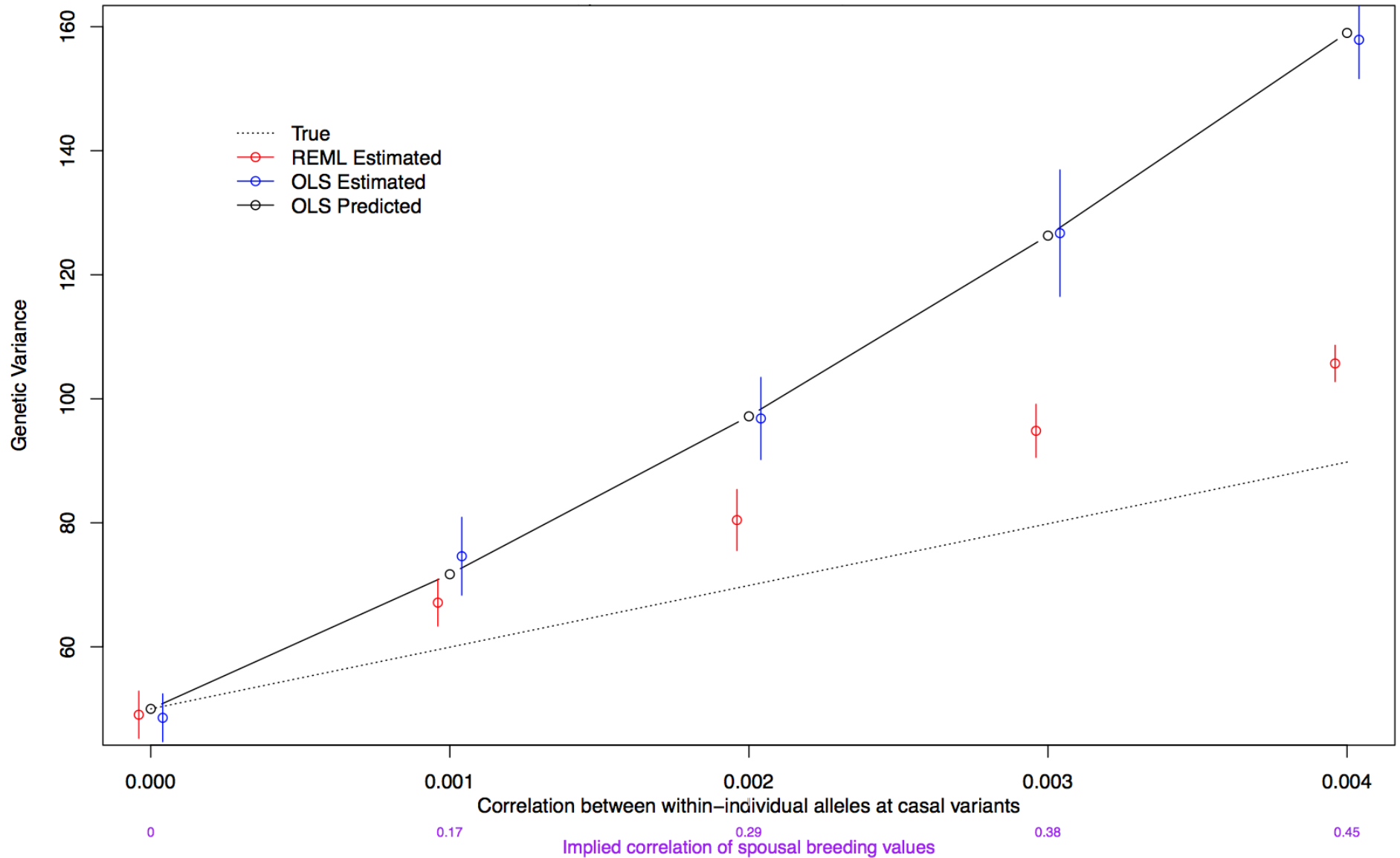


Using Simulated Genes and Phenotypes: REML/FIML (e.g., GCTA) Estimates CVs=100, other markers=0



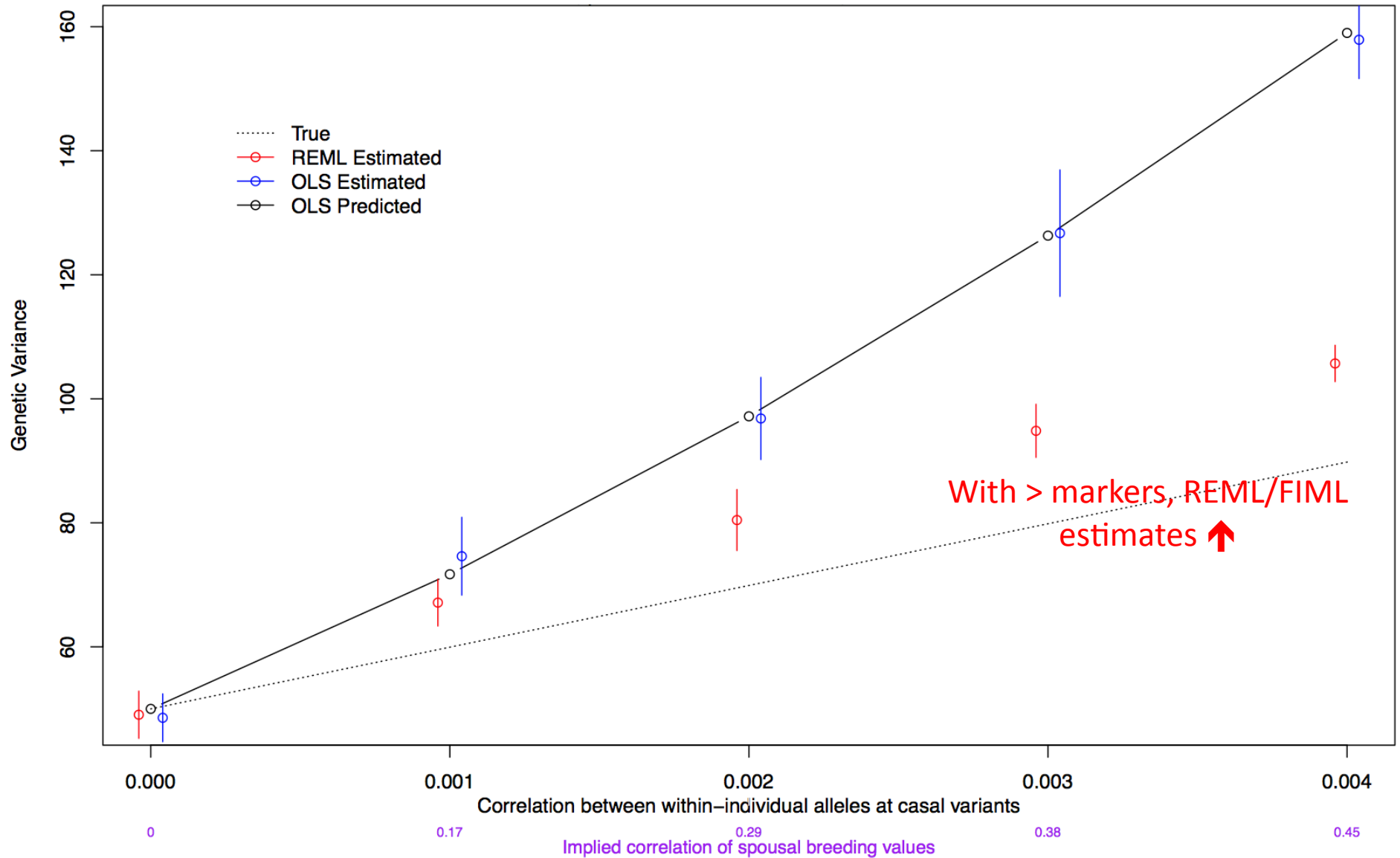
Using Simulated Genes and Phenotypes: REML/FIML (e.g., GCTA) Estimates

CVs=100, other markers=10k



Using Simulated Genes and Phenotypes: REML/FIML (e.g., GCTA) Estimates

CVs=100, other markers=10k



Differences Between OLS and REML/FIML

- While we can predict biases using OLS, we don't know how to predict biases in REML/FIML, nor why the latter estimates depend on the number of markers included in the analysis
- REML/FIML methods model within-person genetic effects and estimate environmental variance directly, while OLS does not
- REML/FIML assume that breeding values and environmental values each follow a normal distribution while OLS does not. Under AM, are breeding value distributions leptokurtic (fat tailed)?

Acknowledgments

Rasool Tahmasbi

Matt Jones

Greg Carey

Lindon Eaves

Matt Keller

