Estimating Genotype- and Environment-Specific Heritabilities

Jode W. Edwards
USDA ARS

Motivation

Comstock and Moll (1963) state:

"we know from experience that the plot error variance is variable from one experiment to another, … and there is nothing that compels the variances of the GE interaction effects to be homogeneous."

Implications

- Heritability, or more accurately **repeatability,** of observed data may vary by
  - Environment
  - Cultivar
- Does this matter?
  - As early as 1937, Cochran outlined the superiority of **weighted means** with variable repeatability

Weighted means

- Weighted means are always superior if weights are known (known variances)
  - If weights are unknown, they must be estimated
  - Estimation of weights introduces error
- Weighted versus unweighted means is a choice between 2 evils:
  - Ignoring heterogeneity (use pooled estimator, unweighted analysis)
  - Ignoring error in estimating weights (use individual estimators, weighted analysis)
- Is there a compromise?

Outline

- Demonstration of a compromise between pooled estimators and individual estimators (very simple model) of variances
- Example results from cultivar trial (full Bayesian model)

A Model of Heterogeneity

- m cultivars are grown in n environments (balanced 1-way classification)
  - $y_{ij} = \theta_i + \varepsilon_{ij}$
- Both means and variances are random variables specific to each cultivar
  - Cultivar means: $\theta_i \sim \mathrm{N}(\mu, \sigma_g^2)$
  - Cultivar variances: $\ln(\sigma_i^2) \sim \mathrm{N}(\alpha, \sigma_\alpha^2)$
- $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_i^2)$ (normally distributed error)

Source: Leonard, 1975, Technometrics 17:95

How to estimate σ_i^2?

- Pooled estimator:

  $\sigma_i^2 = (m(n\text{-}1))^{-1}\Sigma_{ij}(x_{ij}\text{-}\ x_{i\cdot})^2$

- Cultivar specific variance:

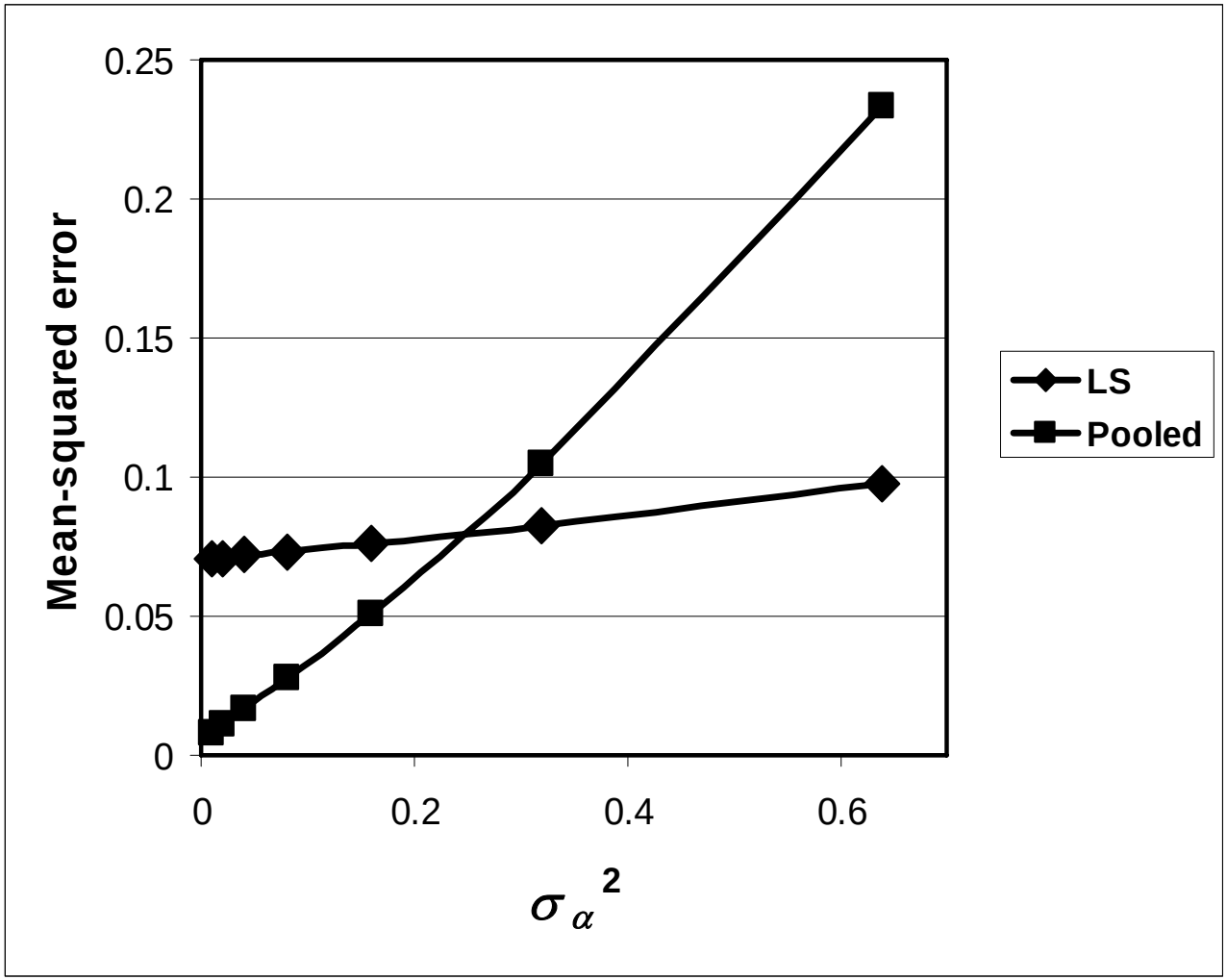
  $\sigma_i^2 = (n\text{-}1)^{-1}\Sigma_j(x_{ij}\text{-}\ x_{i\cdot})^2$

- The better estimator depends on
  - Heterogeneity = σ_α^2
  - Number of environments (replicates)

Simulation

- Parameters:
  - $\sigma_g^2 = 1.0$
  - $m = 100$ cultivars
  - $n = 8$ environments
  - Average error variance = 1.0
- $\mathrm{Var}(\ln(\sigma_i^2)) = \sigma_\alpha^2$: 0.01..0.64
- Run 1000 simulation replicates
- Estimate standard errors of variance estimators (using square roots)
  - $e.g., m^{-1}\Sigma_i(\sigma_i\text{-}s_i)^2$

Simulation results



Compromise Estimator

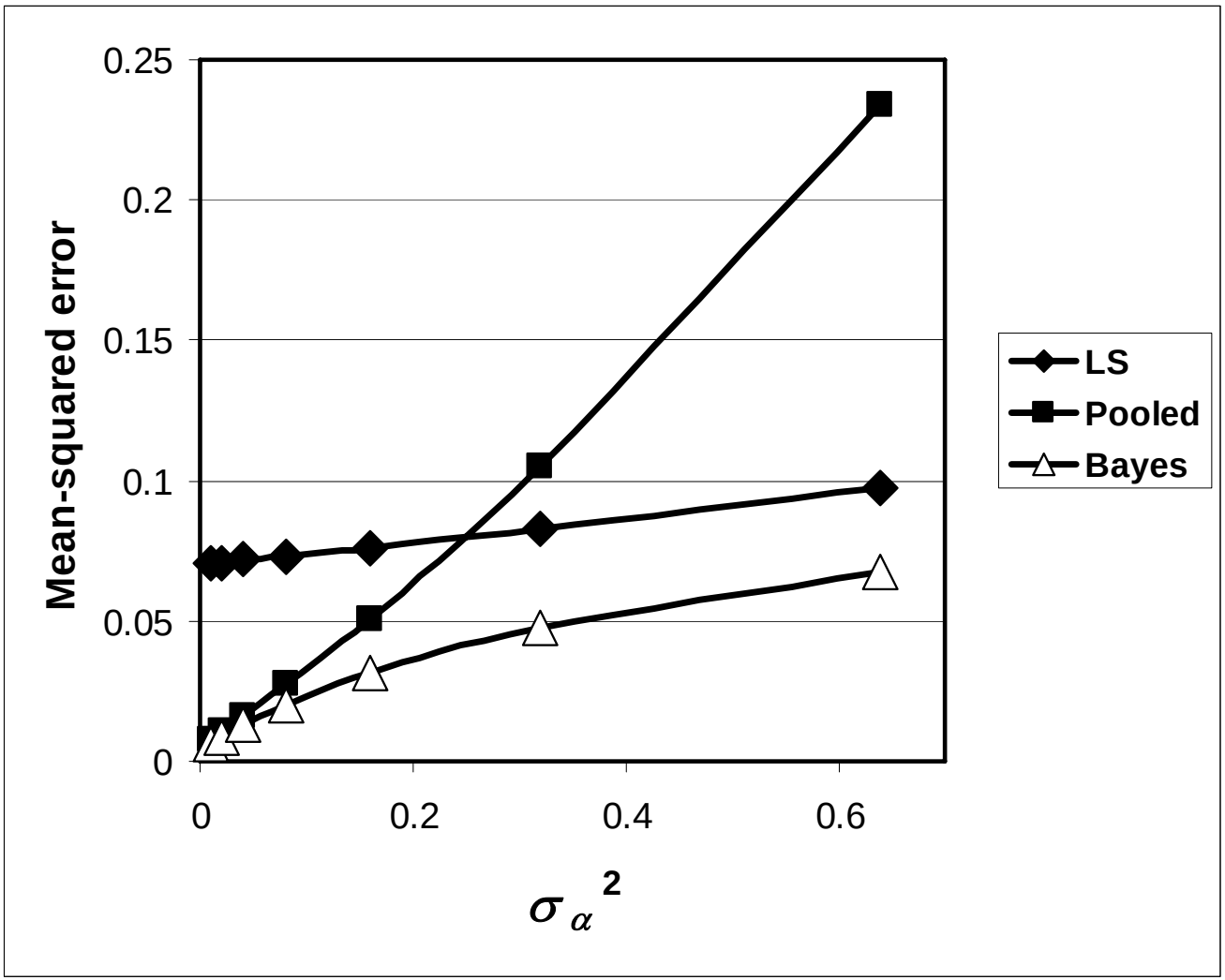
- What about a compromise between the individual estimator and the pooled estimator?

$$\ln\left(\sigma_i^2\right) = \frac{\frac{1}{2} n_i}{\frac{1}{2} n_i + \frac{1}{\sigma_\alpha^2}} \left(\ln\left(s_i^2\right)\right) + \frac{\frac{1}{\sigma_\alpha^2}}{\frac{1}{2} n_i + \frac{1}{\sigma_\alpha^2}} \tilde{\alpha}$$

where $s_i^2 =$ cultivar-specific variance

and $\tilde{\alpha} =$ pooled estimator

Simulation results



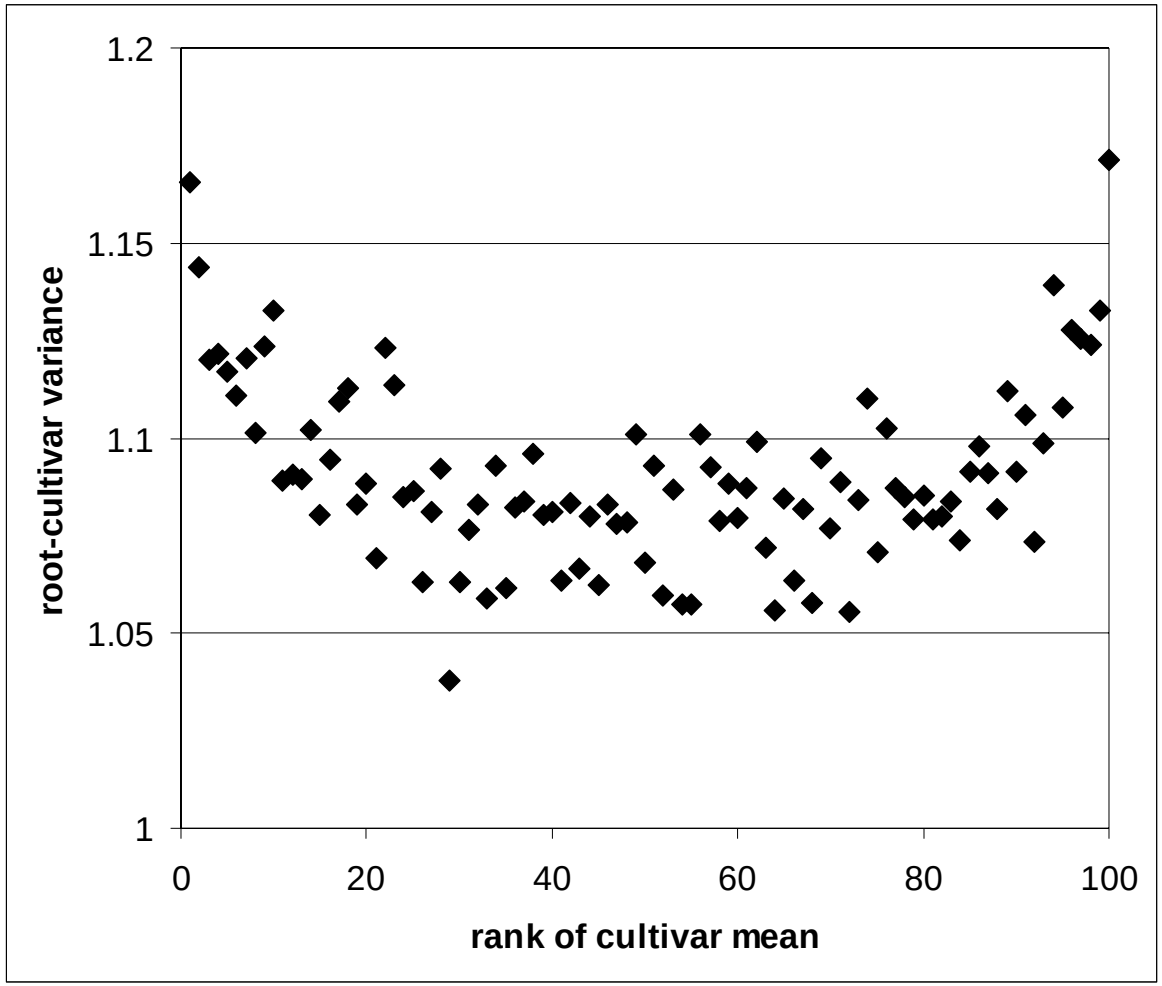
Means Estimation

The ith cultivar mean estimator is:

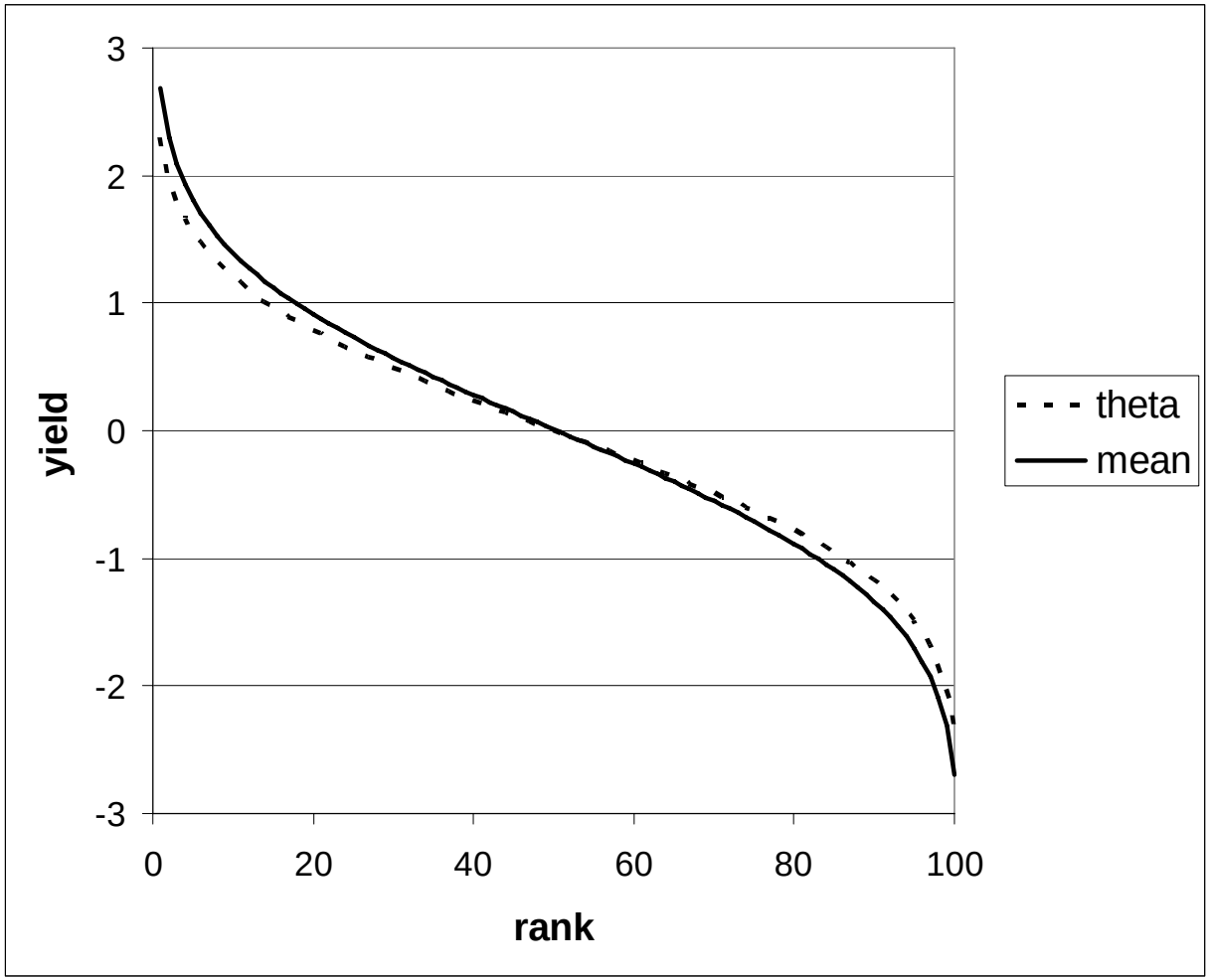
$$\theta_i = \frac{n_i / \sigma_i^2}{n_i / \sigma_i^2 + 1 / \sigma_g^2} \overline{x}_i + \frac{1 / \sigma_g^2}{n_i / \sigma_i^2 + 1 / \sigma_g^2} \tilde{\theta}$$

- Estimators of means are compromise estimators between
  - Grand mean
  - Mean of cultivar i

Selection on means



Selection on means



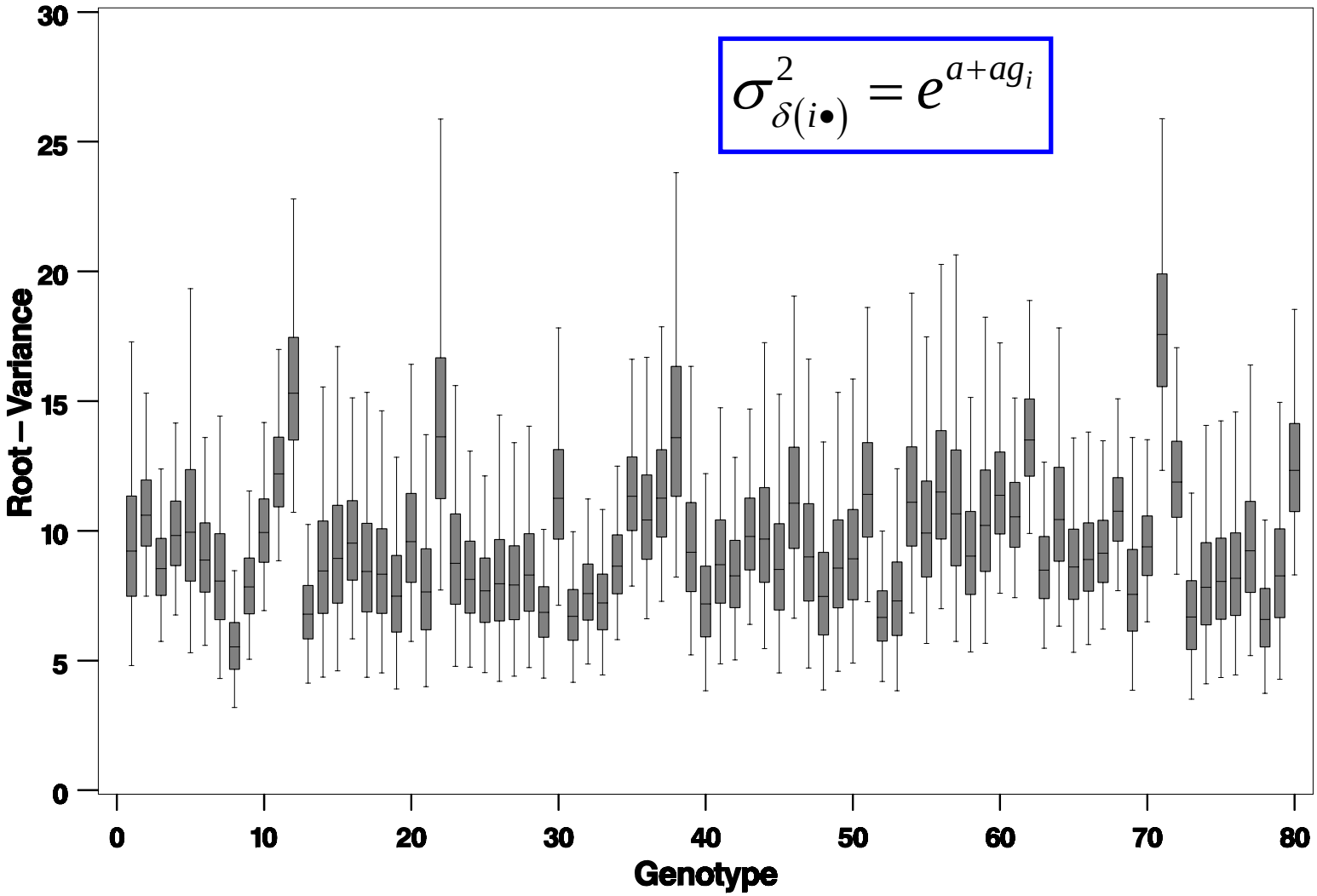
Simulation Summary

- When conclusions are drawn about an estimator (e.g., a mean) conditional on its rank, beware
- In the face of heterogeneity, the best estimator is a compromise between the pooled variance and a cultivar-specific variance

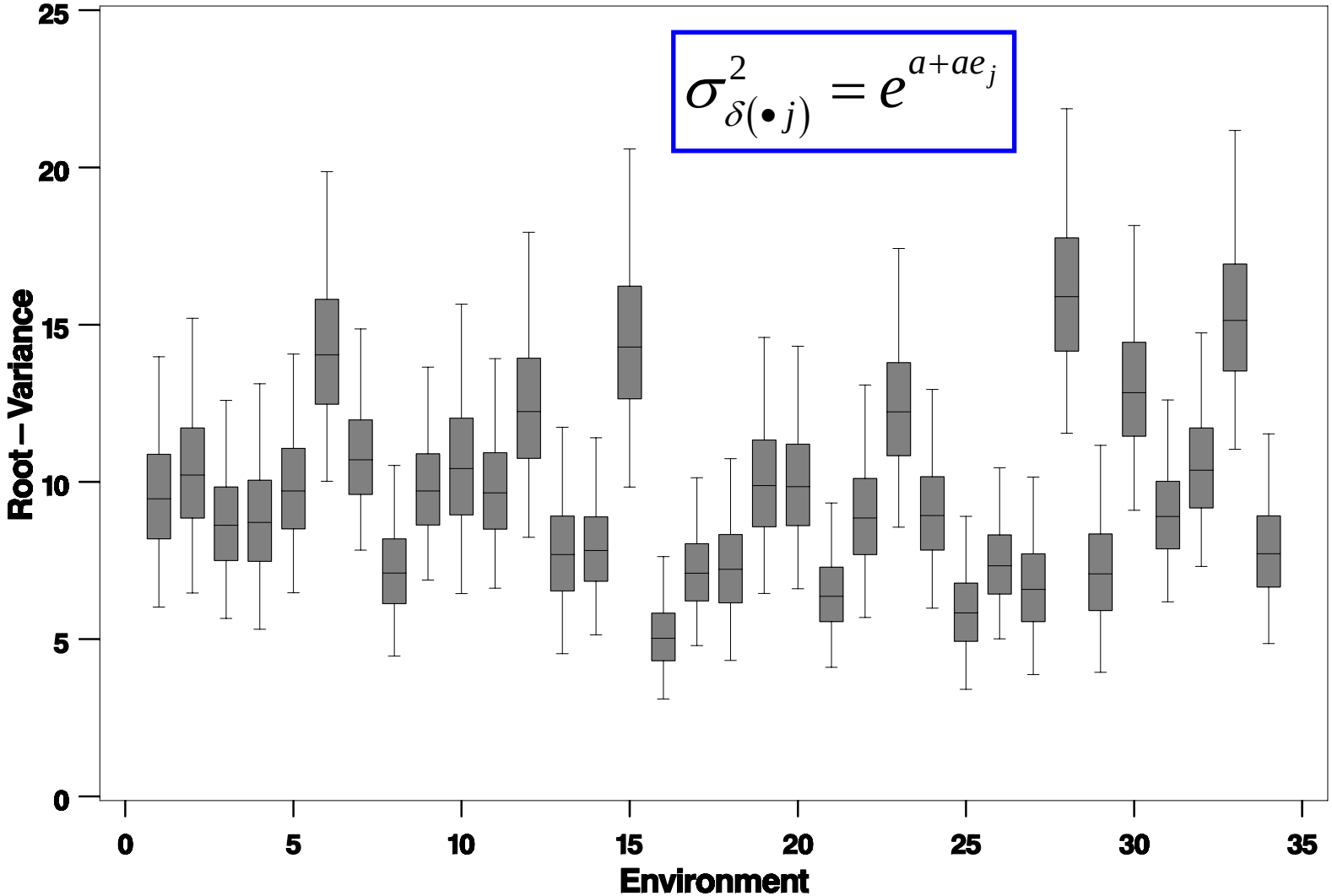
Real Data Example

- Iowa State Oat variety trial, 1997-2003
- 5 locations per year (34 environments)
- 40 genotypes per year; 80 genotypes total
- 3 replicates per location in a randomized complete block design
- Modeled heterogeneity of error and g x e variances among environments and cultivars
- See Edwards and Jannink, 2006, Crop Science for details

Marginal G x E Variances



Marginal G x E Variances



Results

Line	Hetero-geneous	Homo-geneous	Root G x E
IL95-1241	17.1	17.2	8.3
WIX8179-2	13.8	16.0	9.2
Rodeo	9.4	8.0	8.5
Blaze	9.4	9.4	8.5
SD96024	9.4	9.9	8.6
P976A6-9-6	9.3	12.2	11.5

Conclusions

- The best solution to weighted versus unweighted means is a compromise based on shrinkage estimation
- Greatest advantage of modeling heterogeneity of variance may be in combination with **selection**

Future Research

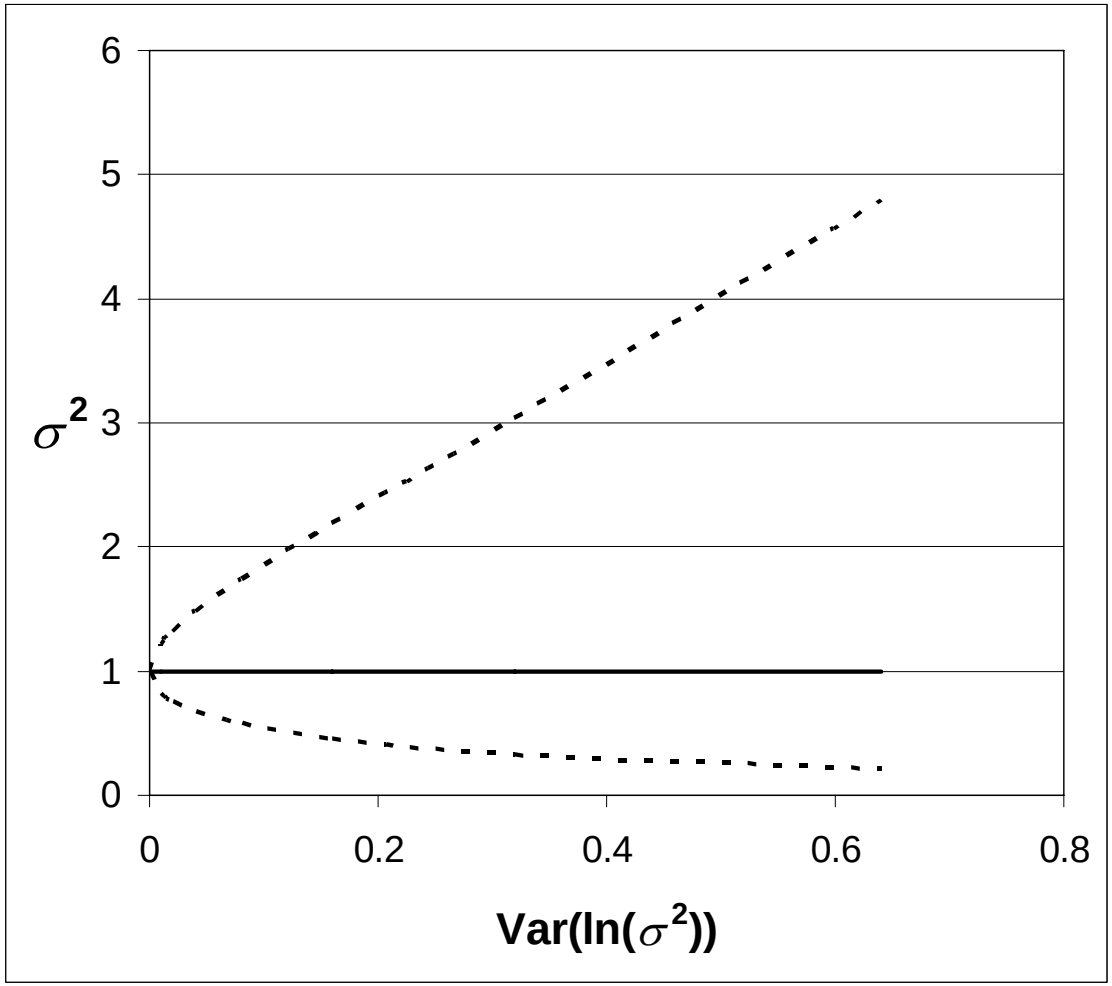
- Model cross-validation – currently underway
- Analysis of corn data
- Compare alternative models of heterogeneity
- Simulation to quantify impact of heterogeneity on gain from selection

Acknowledgements

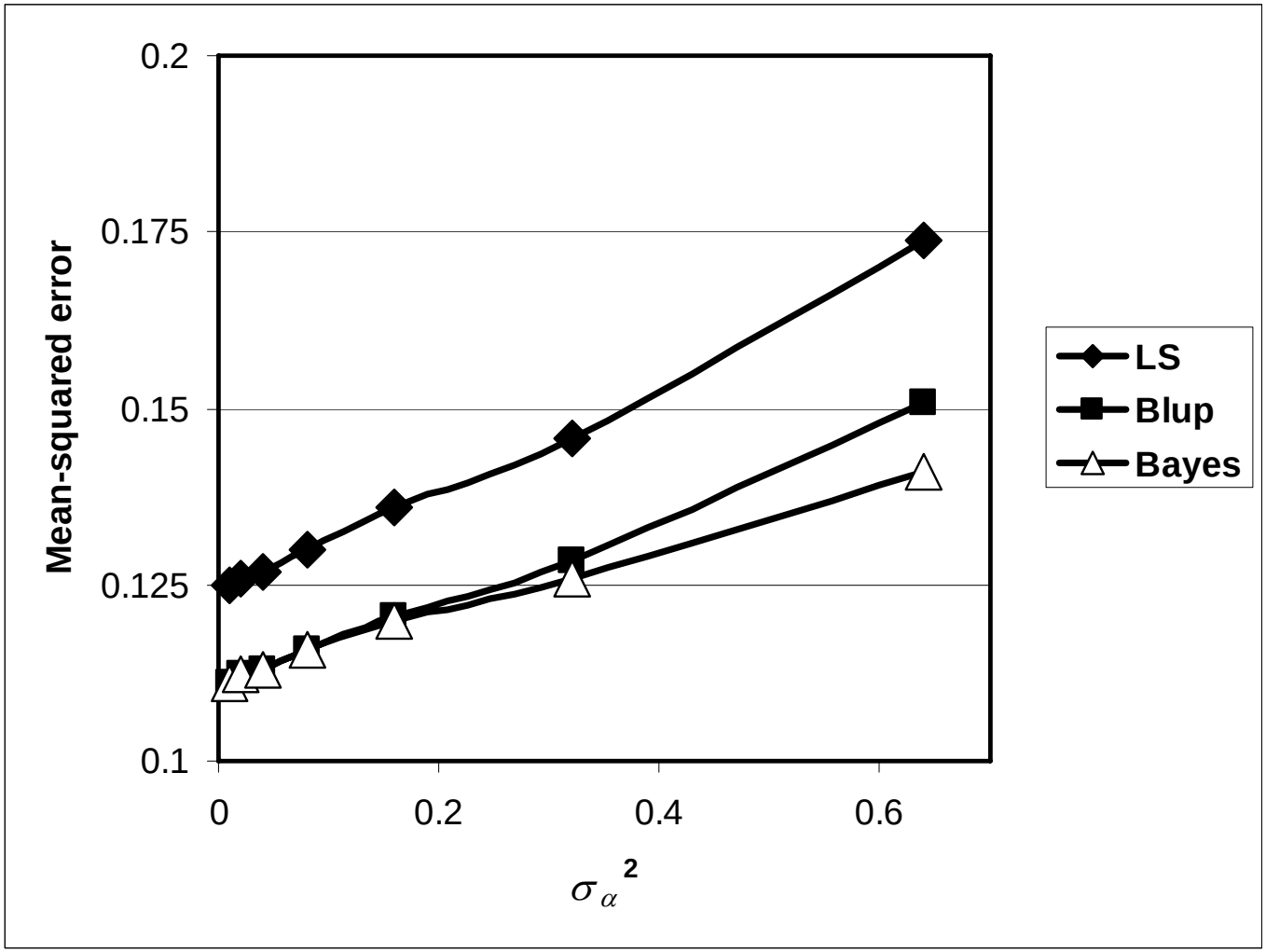
- Faculty
  - Jean-Luc Jannink
  - Alicia Carriquiry
- Students
  - Massiel Orellana
  - Yoon-Sup So

Thank you!

Distribution of Variances



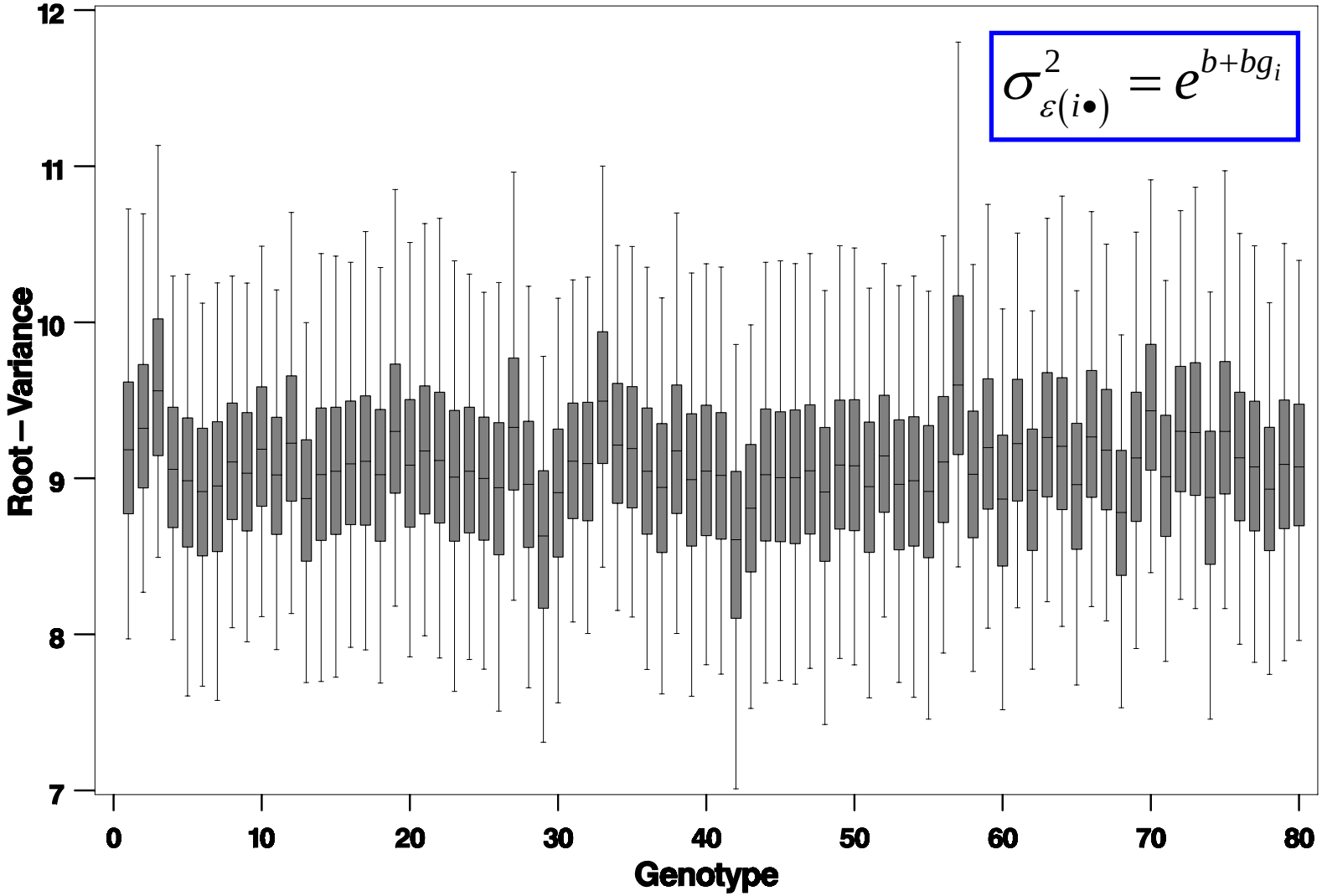
Precision of Means



Implications

- On average, modest improvement in precision of estimating means
- Better estimates of precision on individual cultivars
  - Eliminate “1-year wonders”?
- The biggest advantage may be when selecting among means (more to come)

Marginal Error Variances



Marginal Error Variances