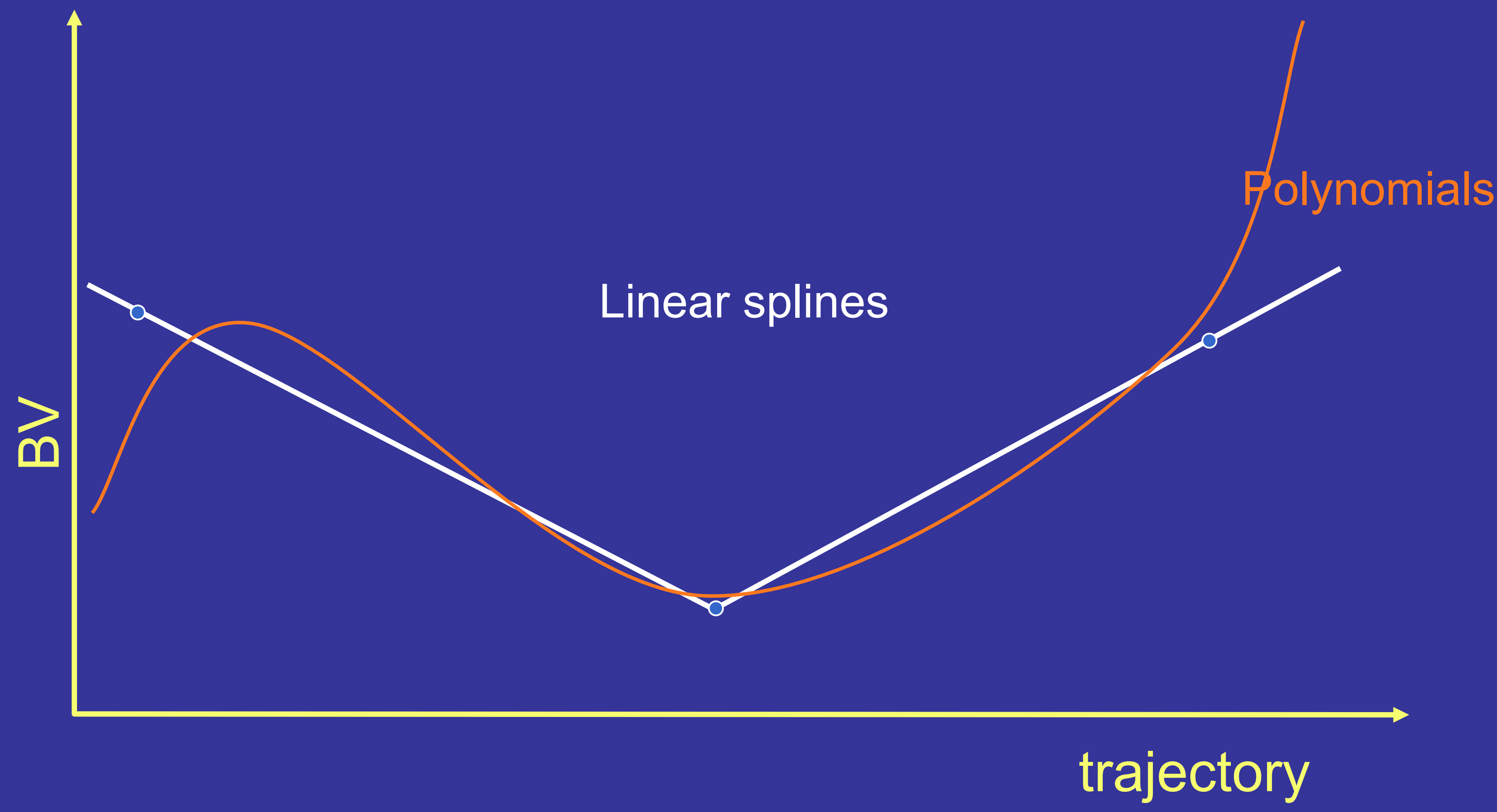
Properties of random regression models using linear splines

Ignacy Misztal

University of Georgia, USA

Breeding values in random regression models



Covariables in RRM

- Polynomials
  - Artifacts
  - Poor computing properties without diagonalization
  - Cryptic parameters
  - Poor modeling of localized variation
- Linear splines
  - Good computing properties
  - Parameters on multiple-trait scale
  - Properties unknown
  - Not clear how to choose knots

Objectives

- Present properties of random regression models using linear splines
- Provide guidelines for selecting the number and position of knots

Why computations with linear splines are more robust?

General random regression model:

$y_{ijk} = \ldots + \Phi(t)a_j + \ldots.$

Polynomials: $\Phi = [\varphi_1 \ \varphi_2 \ .. \ \varphi_n]$

all coefficients nonzero

Splines: $\Phi = [0..0 \ \ 1\text{-}t \ \ t \ \ 0 \ ..0]$

only 2 coefficients nonzero

Sparser equations with splines!

(Co)variances & correlations

With 2 knots:

$a(t)=(1-t)a_1+ta_2 \quad 0 \le t \le 1 \quad var(a)=G$

$var[(a(t)]=(1-t)^2 g_{11}+t(1-t)g_{12} + t^2 * g_{22}$

$cov[a(0),a(t)] = (1-t)g_{11}+tg_{12}$

If $g_{11}=g_{22}=1$, $g_{12}=\rho$

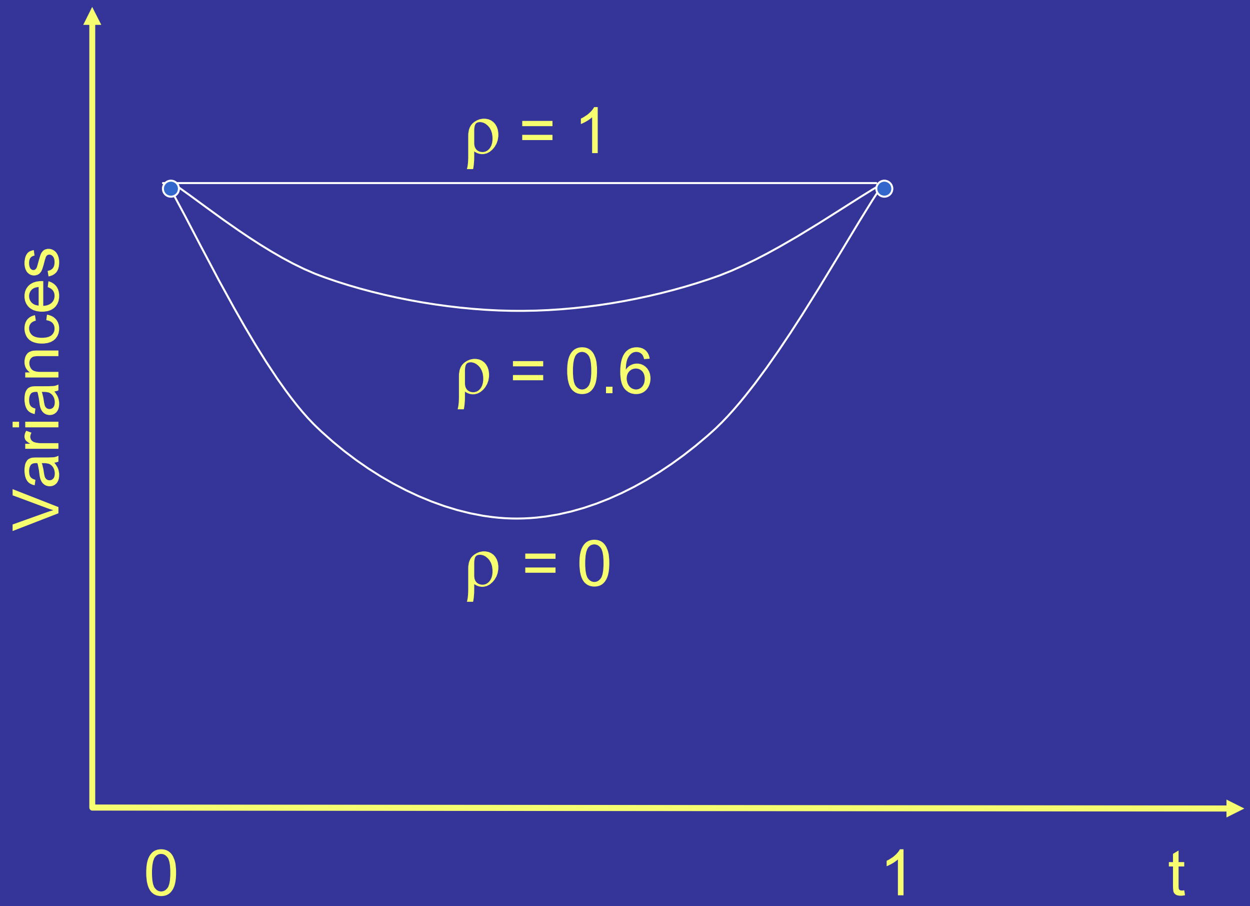
$Var[a(t)]= =(1-t)^2+t(1-t)\rho + t^2$

$Var[a(0)]=1$; $var[a(1)]=1$; $var[a(0.5)]=0.5+0.5\rho$

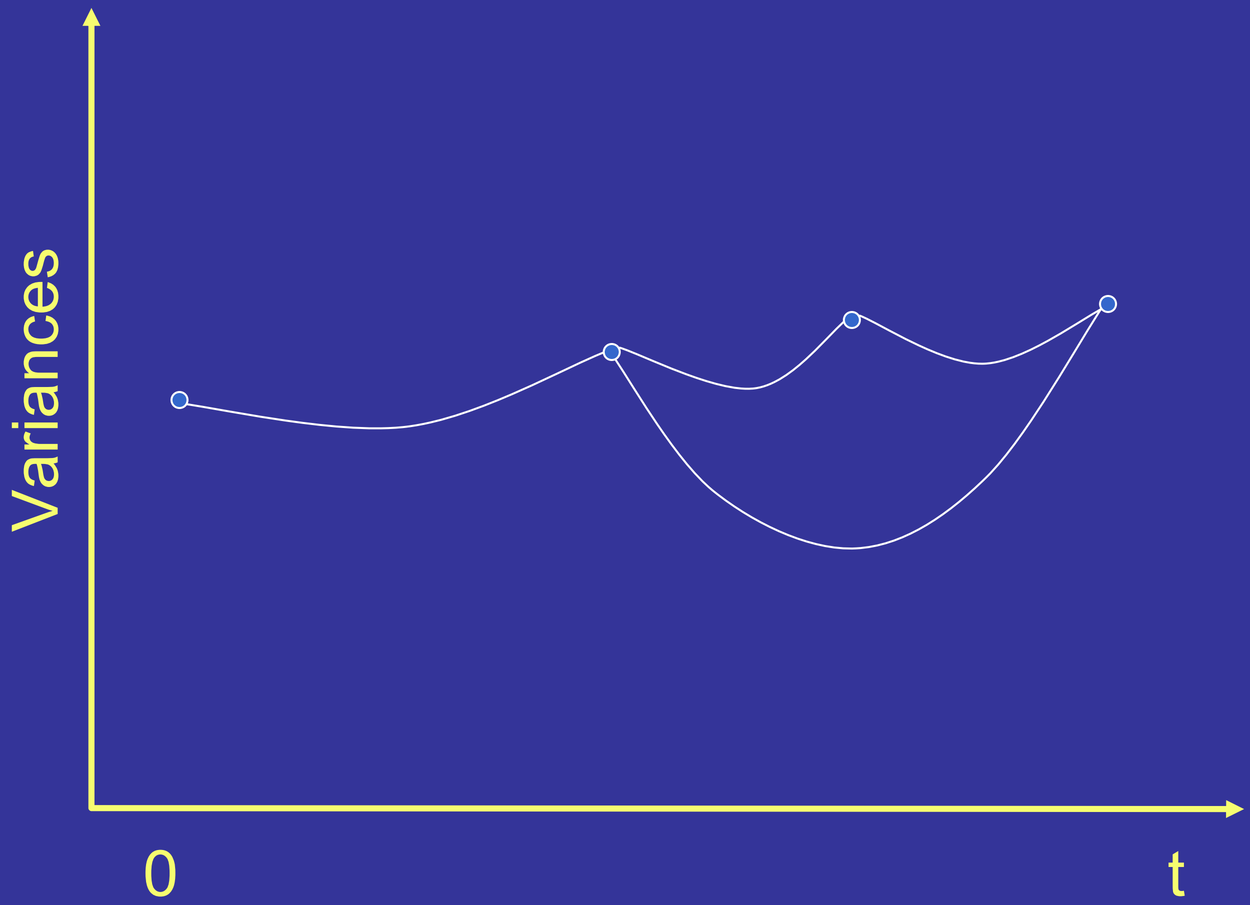
$Cov[a(0),a(1)]=\rho$

$Cov[a(0),a(t)]=1-t(1-\rho)$

Variances with linear splines



Variances with different number of knots



Modification to smoothen variances

Replace: $\Phi(t) = [\ldots\ 1-t \quad t \quad \ldots]$ with

$\Phi(t) = [\ldots\ (1-t)^q \quad t^q \quad \ldots]$ $0 \leq q \leq 1$

Then

$var[(a(t)]=(1-t)^{2q}+t^q\,(1-t)^q + \rho t^{2q}$

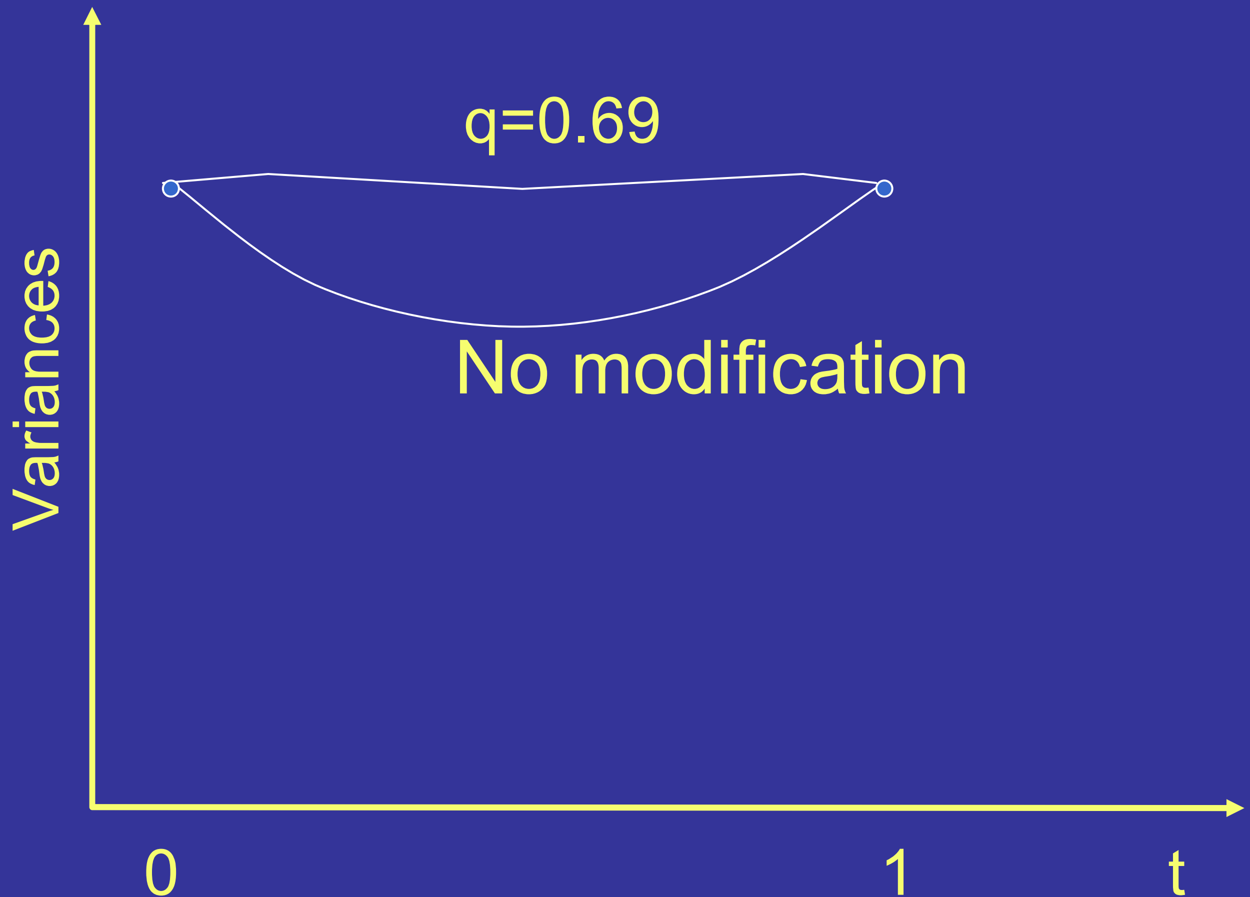
$cov[a(0),a(t)] = (1-t)^q+\rho t^q$

q can be set so that var(a(0.5))=1

$q=\log[2(1+\rho)]/[(2*\log(2)]$

$\rho=0.5 \rightarrow q=0.69$

Variances after corrections (ρ=0.5)



Simulation study

Data simulation

5 knots (t=1,2,3,4,5)
$corr(a_1,a_5)$= 0.0 to 0.9

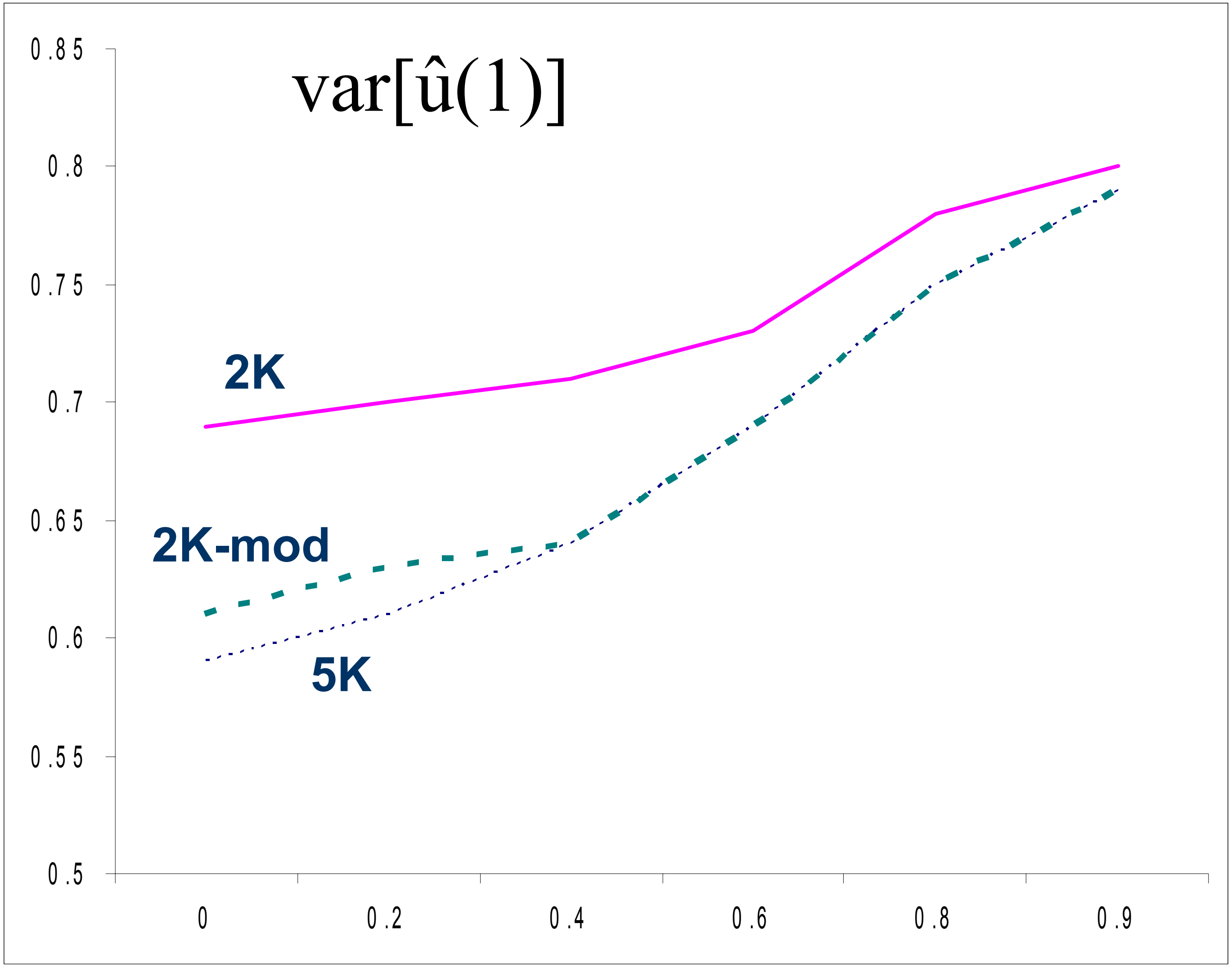
Models for analyses

5 knots (5K)
2 knots (2K)
2 knots with modification (2K-mod)

Runs

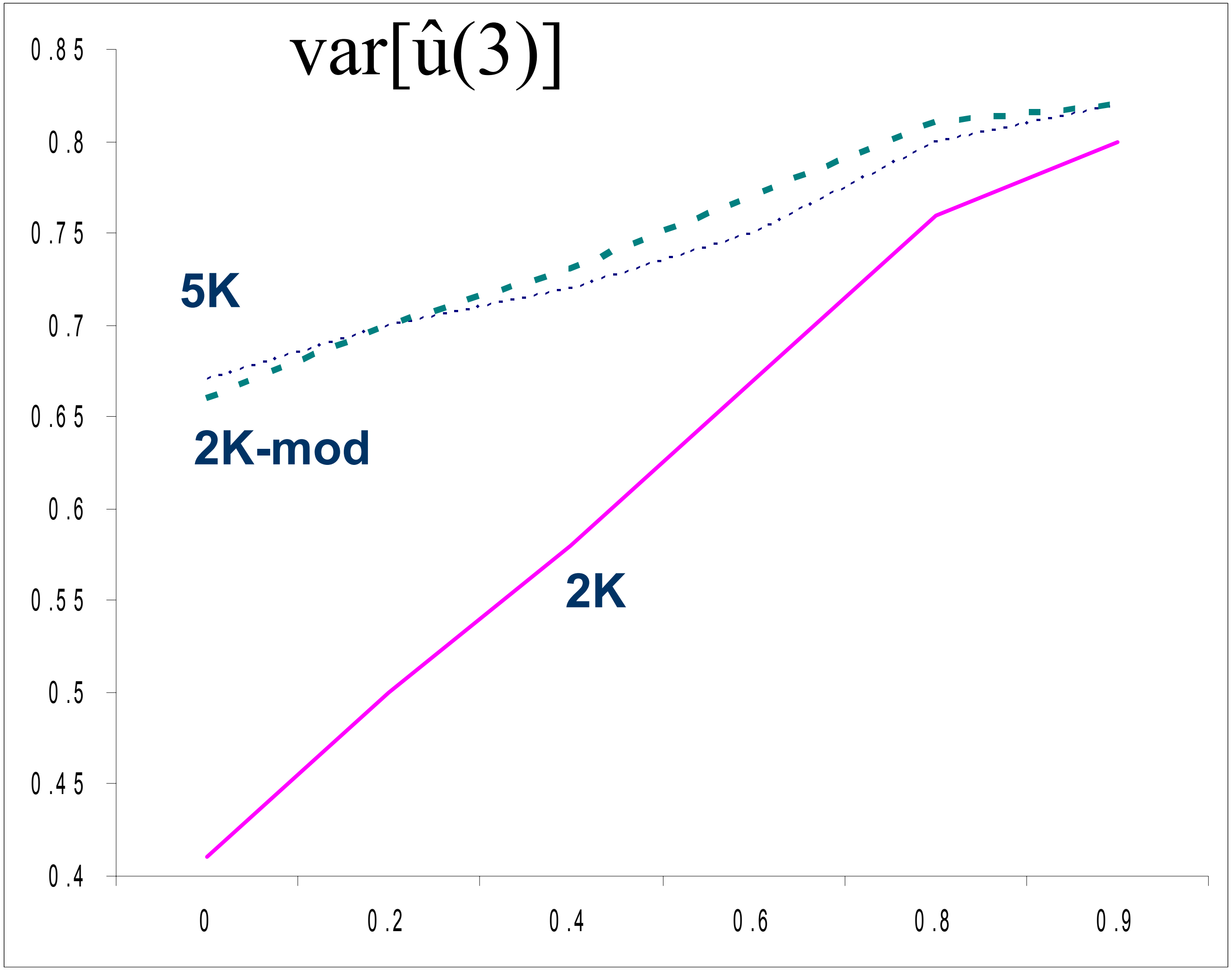
1000 sires with 10 observations each
10 replicates

Variance of prediction at one extreme

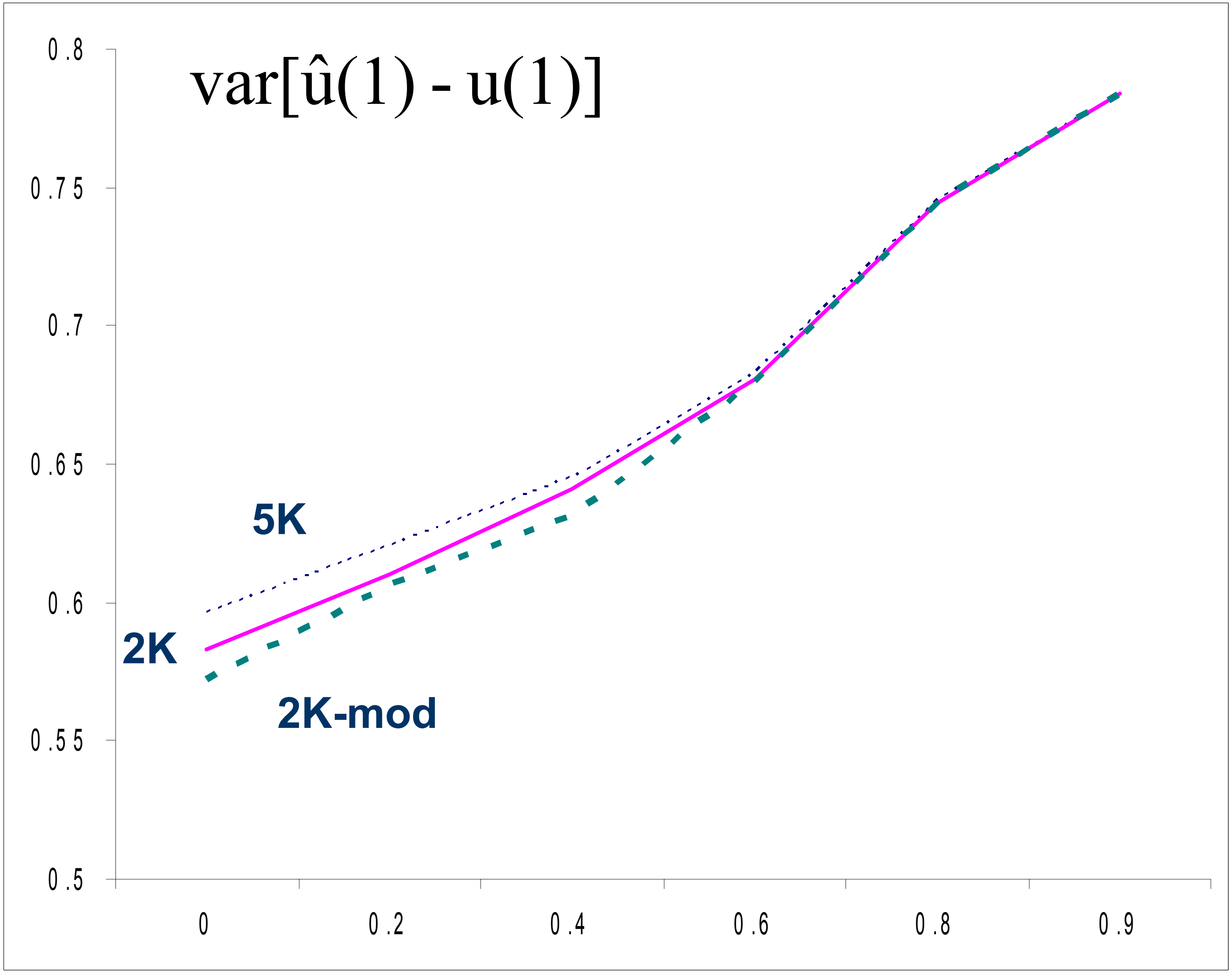


Corr(a_1,a_5)

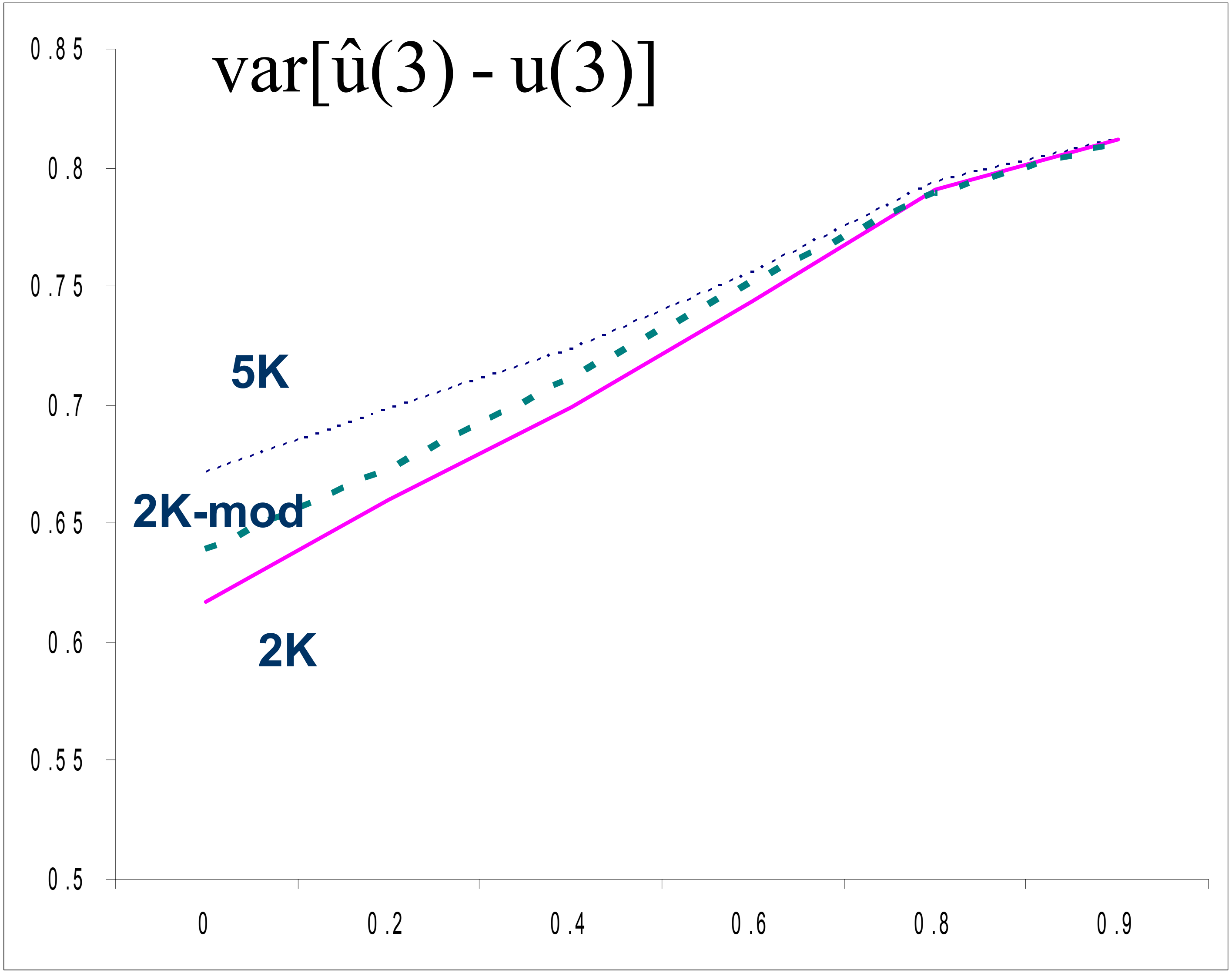
Variance of prediction in the middle



Accuracy of prediction at one extreme



Accuracy of prediction in the middle



Rules for selection of knots

- **Select knots so that correlations corresponding to adjacent knots:**

  ≈0.8 for regular linear splines

  ≈0.6 for splines with covariables modified

- **Example: dairy**
  - knots at 10, 70, 250 and 350 d

- **Under/over predictions average out over the trajectory**

Conclusions

- **Computing costs with linear splines low because of sparcity of LHS of mixed model equations**

- **Some problems with convexity/inflation/deflation**

- **Problems greatly decreased with a simple modification**

- **RRM with linear splines simple and robust**