Assessing Variable Selection Uncertainty in Linear Models

Aldo Solari

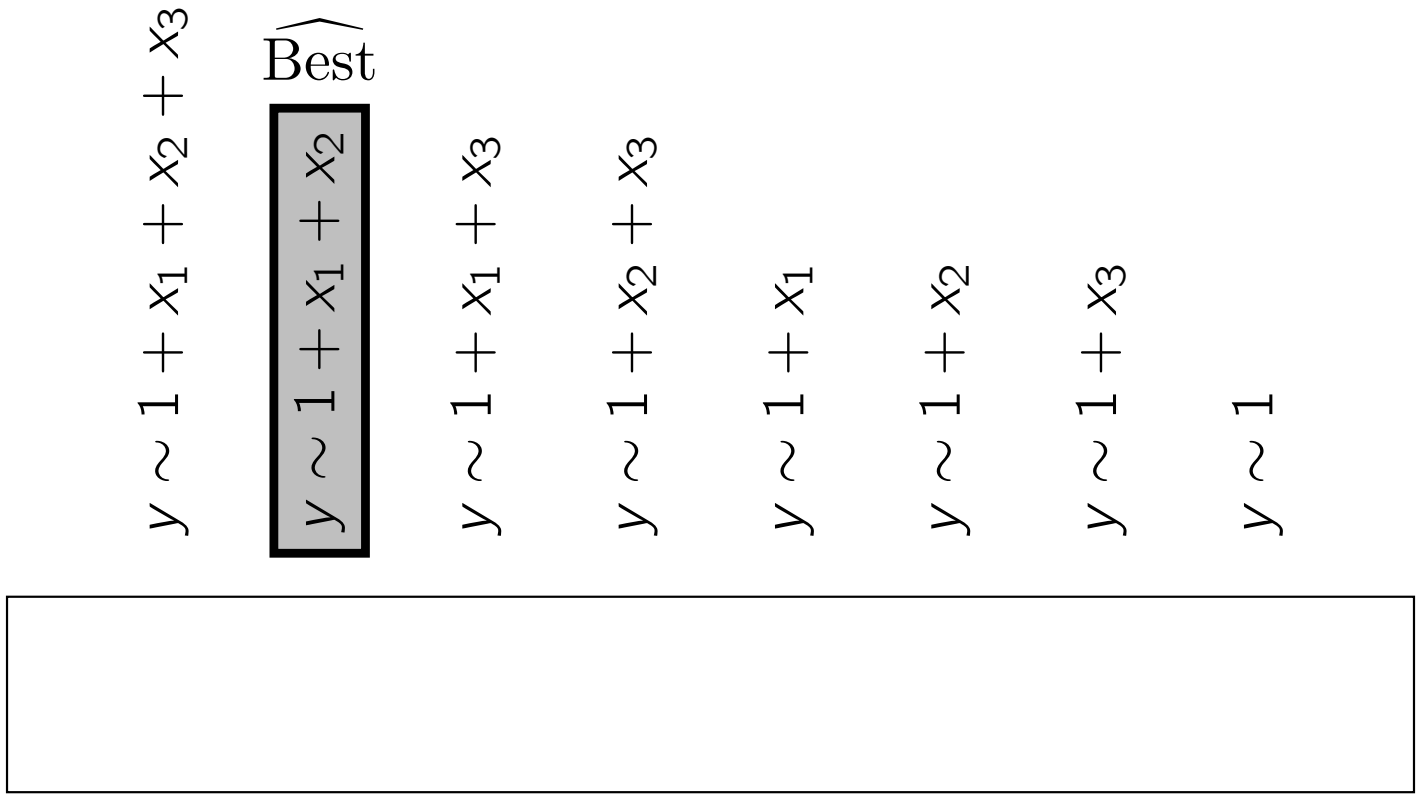
University of Milano-Bicocca, Italy

Joint work with Ningning Xu and Jelle Goeman

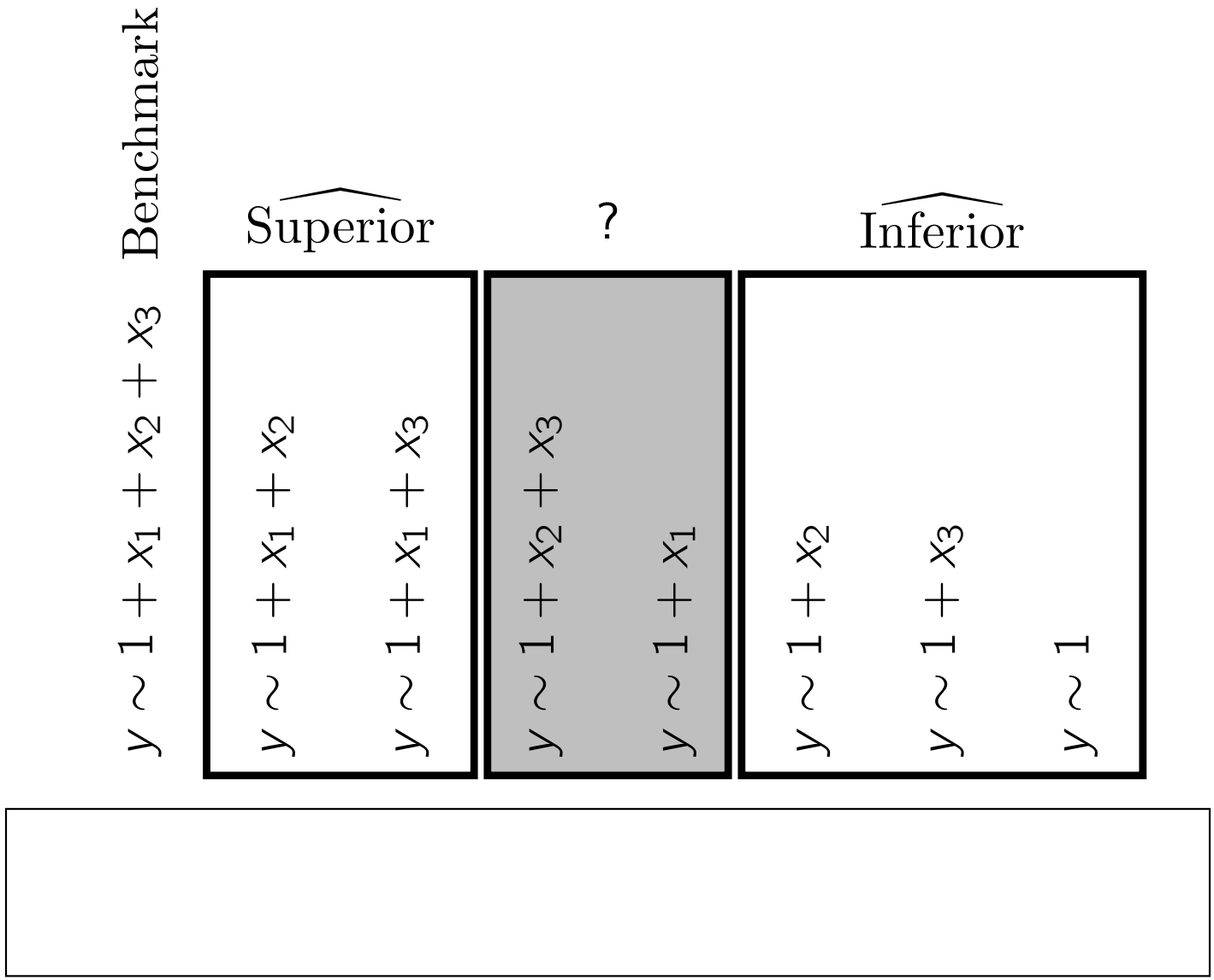
10th Conference on Multiple Comparison Procedures
Riverside, California - June 22, 2017

Outline

Variable selection = point estimation



Many-to-one comparisons = uncertainty



Linear model

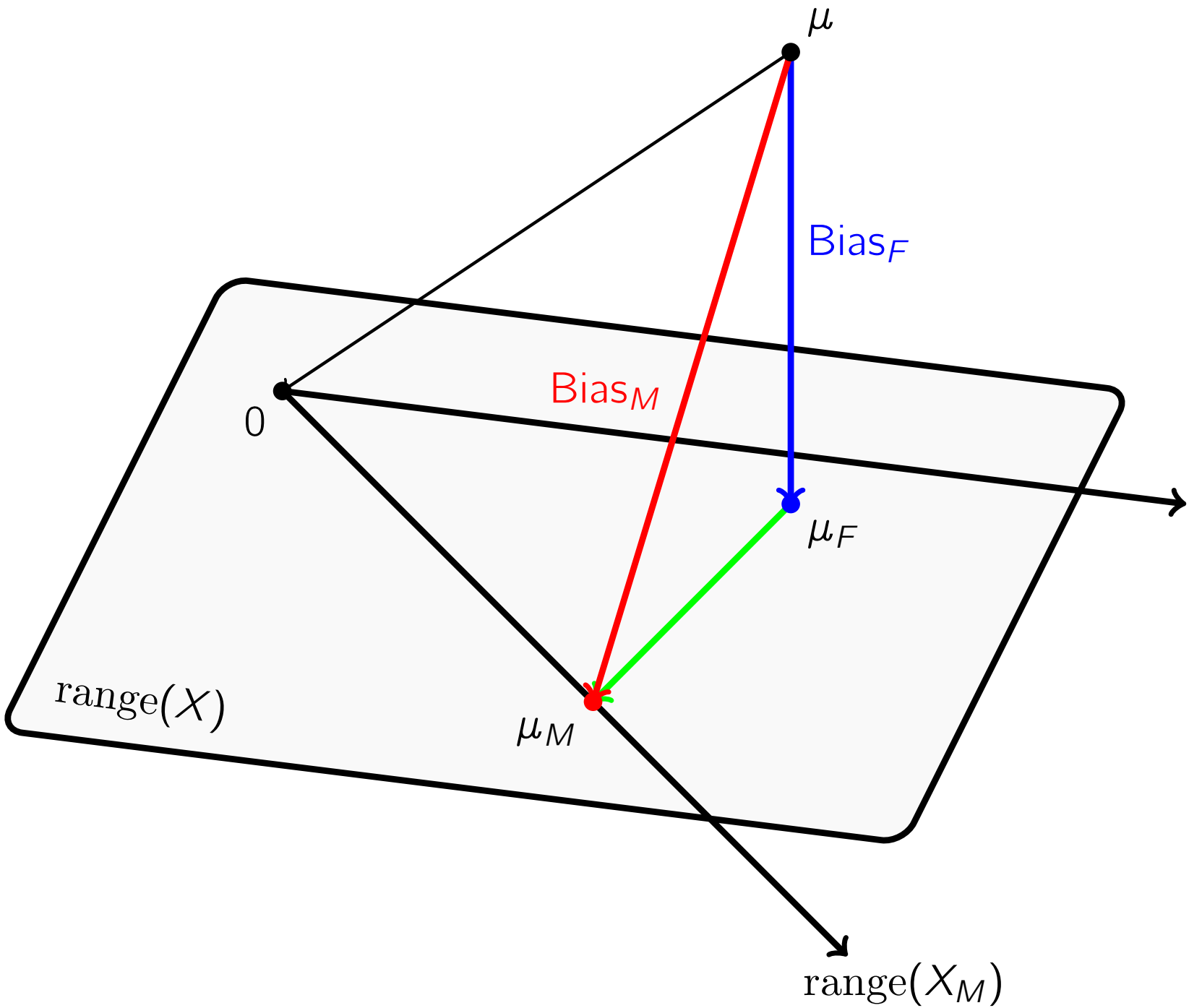
- $y \sim \mathcal{N}(\mu, \sigma^2 I_n)$
- $X \in \mathbb{R}^{n \times p}$: design matrix

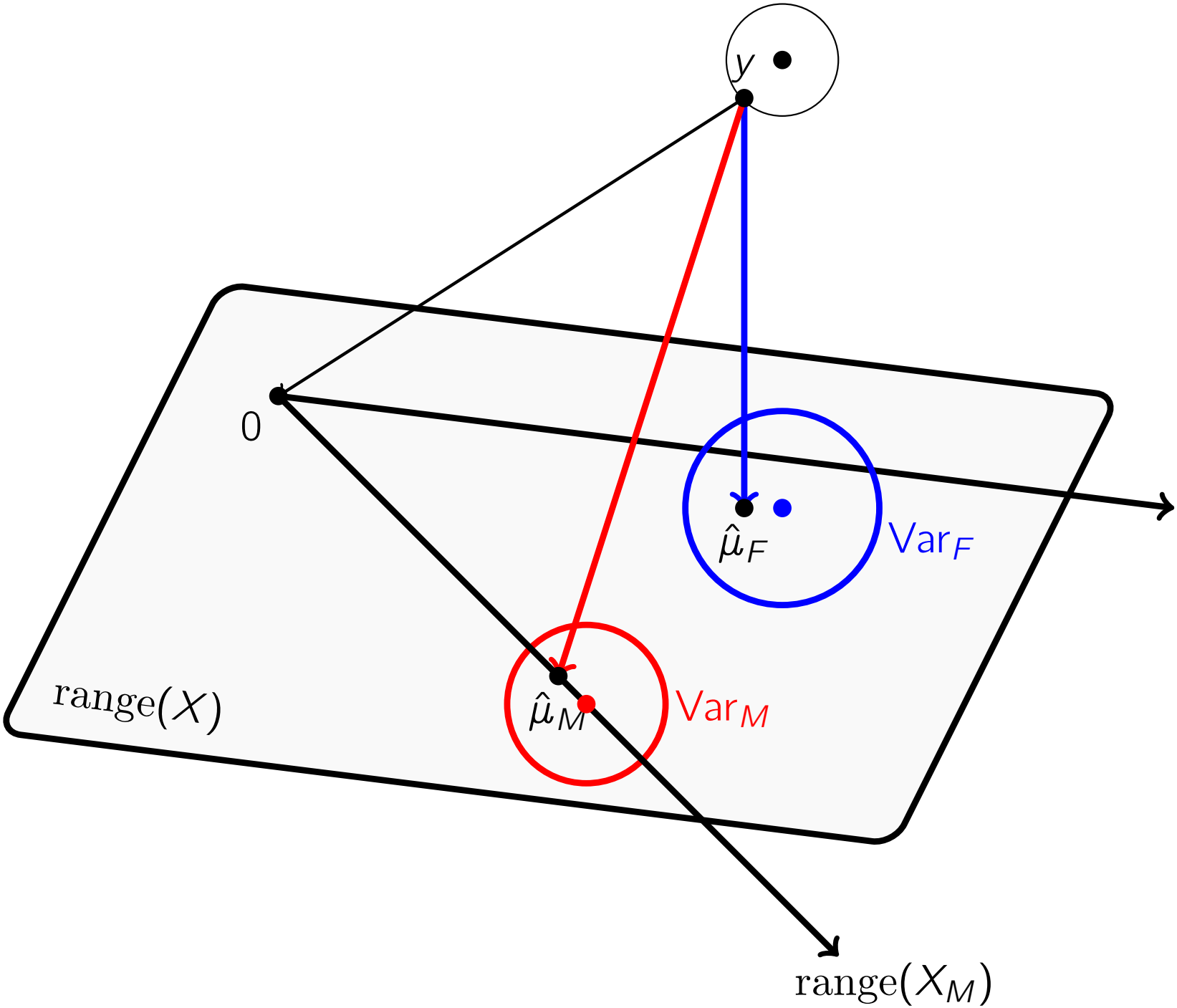
First-order misspecification

- $\mu \in \mathrm{range}(X)$: full model is correct (unbiased)
- $\mu \notin \mathrm{range}(X)$: first-order misspecification

Models

- $M \subseteq F = \{1, \ldots, p\}$ with $\#M = m$
- $\hat{\mu}_M \sim \mathcal{N}(\mu_M, \sigma^2 P_M)$
  with $\hat{\mu}_M = P_M y$, $\mu_M = P_M \mu$ and $P_M = X_M (X_M^\mathsf{T} X_M)^{-1} X_M^\mathsf{T}$


μ
Bias_F
Bias_M
0
μ_F
$\text{range}(X)$
μ_M
$\text{range}(X_M)$


y
0
$\hat{\mu}_F$
Var_F
$\hat{\mu}_M$
Var_M
$\mathrm{range}(X)$
$\mathrm{range}(X_M)$

Outline

To explain or to predict? (Shmueli, 2010)

Explanatory modeling

- Obtain the most accurate representation of the underlying theory
- Avoid/minimize Bias
- Omitted-variable bias compromises interpretation

Predictive modeling

- Generate good predictions of new y
- minimize $\text{Bias}^2 + \text{Variance}$
- A biased model can predict better than an unbiased one

Predictive modeling

Mean Squared Error

$$\frac{\mathrm{MSE}_M}{\sigma^2} = \lambda_M + m \qquad \text{where } \lambda_M = \frac{\|\mu_M - \mu\|^2}{\sigma^2}$$

Relative efficiency

$$\frac{\mathrm{MSE}_M}{\mathrm{MSE}_F} = \frac{\lambda_M + m}{\lambda_F + p} > 1 \quad \text{iff} \quad \lambda_M^F = \frac{\|\mu_M - \mu_F\|^2}{\sigma^2} > p - m$$

Inferior and superior models

- $\mathcal{I} = \{M \subseteq F : \lambda_M^F > p - m\}$
- $\mathcal{S} = \{M \subseteq F : \lambda_M^F \leq p - m\}$

Hypothesis testing

One true hypothesis

$M \in \mathcal{I}$ or $M \in \mathcal{S}$

Testing for superiority

- $M \in \mathcal{I}$ against $M \in \mathcal{S}$
- If $M \in \mathcal{I}$ rejected, then $M \in \hat{\mathcal{S}}_\alpha$

Testing for inferiority

- $M \in \mathcal{S}$ against $M \in \mathcal{I}$
- If $M \in \mathcal{S}$ rejected, then $M \in \hat{\mathcal{I}}_\alpha$

Uncertainty

If both $M \in \mathcal{S}$ and $M \in \mathcal{I}$ not rejected, then $M \in \hat{\mathcal{U}}_\alpha$

Confidence sets

$1-\alpha$ **confidence of no type I errors**

$$\mathrm{P}(\{\hat{\mathcal{S}}_\alpha \cap \mathcal{I} = \emptyset\} \cap \{\hat{\mathcal{I}}_\alpha \cap \mathcal{S} = \emptyset\}) \geq 1-\alpha$$

Familywise error control

The probability of at least one type I error in testing the family of 2^{p+1} null hypotheses $\{(M \in \mathcal{I}, M \in \mathcal{S}), M \subseteq F\}$ should be at most α

Uncertainty set

$\hat{\mathcal{U}}_\alpha =$ models that are not in $\hat{\mathcal{S}}_\alpha$ or $\hat{\mathcal{I}}_\alpha$

$\mathcal{F}$ test statistic

$$T_M^F = \frac{\|\hat{\mu}_M - \hat{\mu}_F\|^2}{\hat{\sigma}_F^2} \quad \text{with } \hat{\sigma}_F^2 = \frac{\|\hat{\mu}_F - y\|^2}{n-p}$$

Correct full model assumption

$$T_M^F \sim (p-m)\mathcal{F}'_{p-m,n-p}(\lambda_M^F)$$

First-order misspecification

$$T_M^F \sim (p-m)\mathcal{F}''_{p-m,n-p}(\lambda_M^F, \lambda_F)$$

No testing for superiority

$$? \overset{\text{st}}{\leq} \mathcal{F}''_{p-m,n-p}(\lambda_M^F, \lambda_F) \overset{\text{st}}{\leq} \mathcal{F}'_{p-m,n-p}(\lambda_M^F)$$

Scheffé's adjustment

Maximum test statistic

$$T^F_\varnothing = \max_{M \subseteq F} T^F_M \sim p\mathcal{F}''_{p,n-p}(\lambda^F_\varnothing, \lambda_F)$$

Confidence set

$$\hat{\mathcal{I}}_\alpha = \{M \subseteq F : T^F_M > pf'^{1-\alpha}_{p,n-p}(p)\}$$

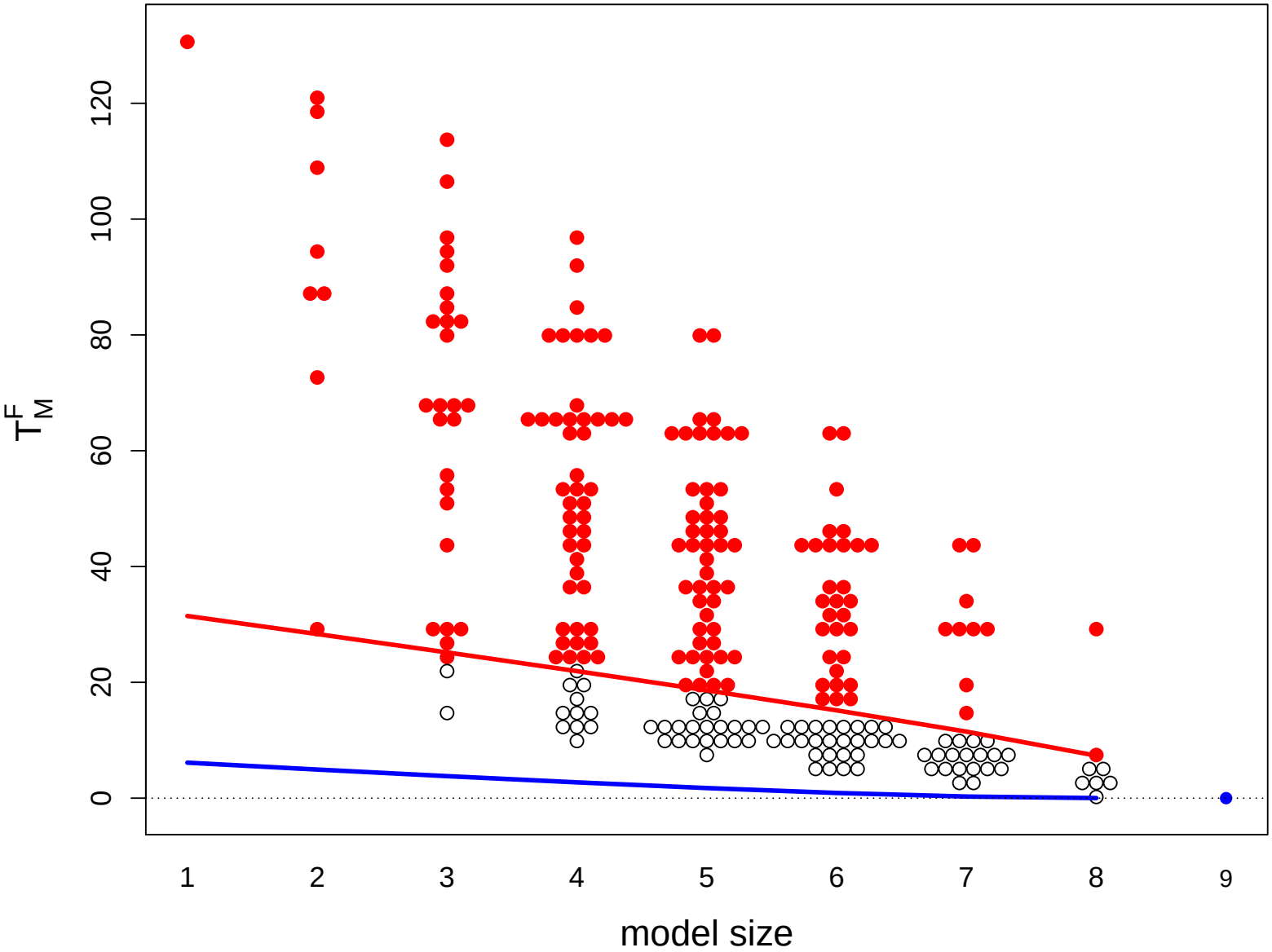
Outline

Prostate cancer data: $n = 67, p = 8$

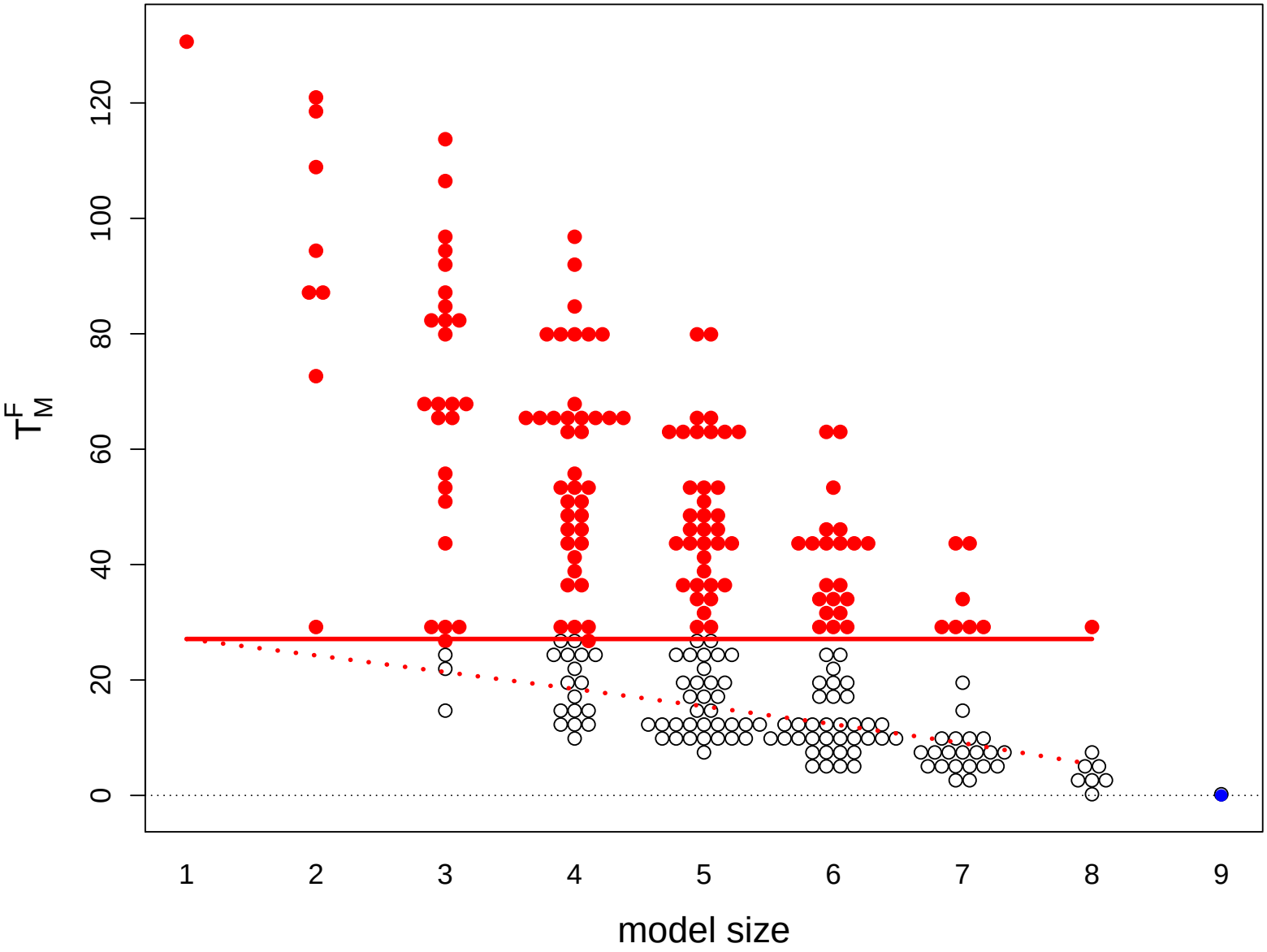
	C_P	BIC	LASSO	FS
lcavol	•	•	•	•
lweight	•	•	•	•
age	•			
lbph	•		•	
svi	•		•	•
lcp	•			
gleason				
pgg45	•			

C_P	best subsets with min C_P/AIC
BIC	best subsets with min BIC
LASSO	10-fold CV with 1-SE rule (Hastie et al. 2009)
FS	forward stop rule on LAR path at 10% FDR (G'Sell et al. 2016)

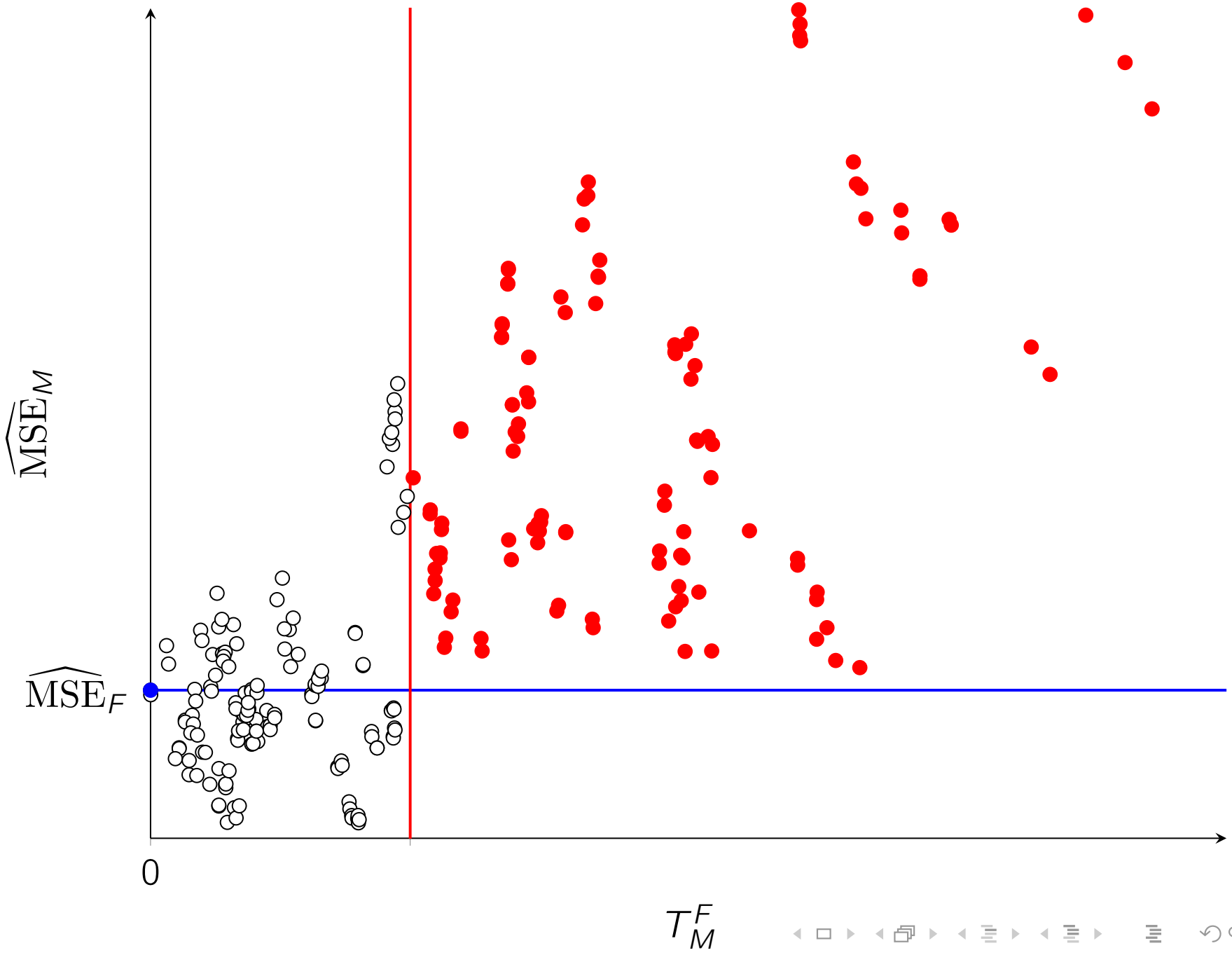
Raw rejections



Uncertainty $u_{5\%} = 54.1\ \%$



Predictions



Explanatory modeling

Correct (unbiased) and wrong (biased) models

- $\mathcal{C} = \{M \in \mathcal{M} : \lambda_M = 0\}$
- $\mathcal{W} = \{M \in \mathcal{M} : \lambda_M > 0\}$

Confidence for correct models?

- Null $M \in \mathcal{W}$ against point alternative $M \in \mathcal{C}$ implies α power
- Confidence for wrong models only: $\mathrm{P}(\hat{\mathcal{W}}_\alpha \cap \mathcal{C} = \emptyset) \geq 1 - \alpha$

More power

- $\mathcal{C} \subseteq \mathcal{S}$
- $\mathcal{W} \supseteq \mathcal{I}$ implies a more powerful confidence set $\hat{\mathcal{W}}_\alpha \supseteq \hat{\mathcal{I}}_\alpha$

Adequate Models

Adequate and non-adequate models

- $\mathcal{A} = \{M \in \mathcal{M} : \ \lambda_M^F = 0\}$
- $\mathcal{B} = \{M \in \mathcal{M} : \ \lambda_M^F > 0\}$

with $\mathcal{C} \subseteq \mathcal{A} \subseteq \mathcal{S} \quad \text{and} \quad \mathcal{I} \subseteq \mathcal{B} \subseteq \mathcal{W}$

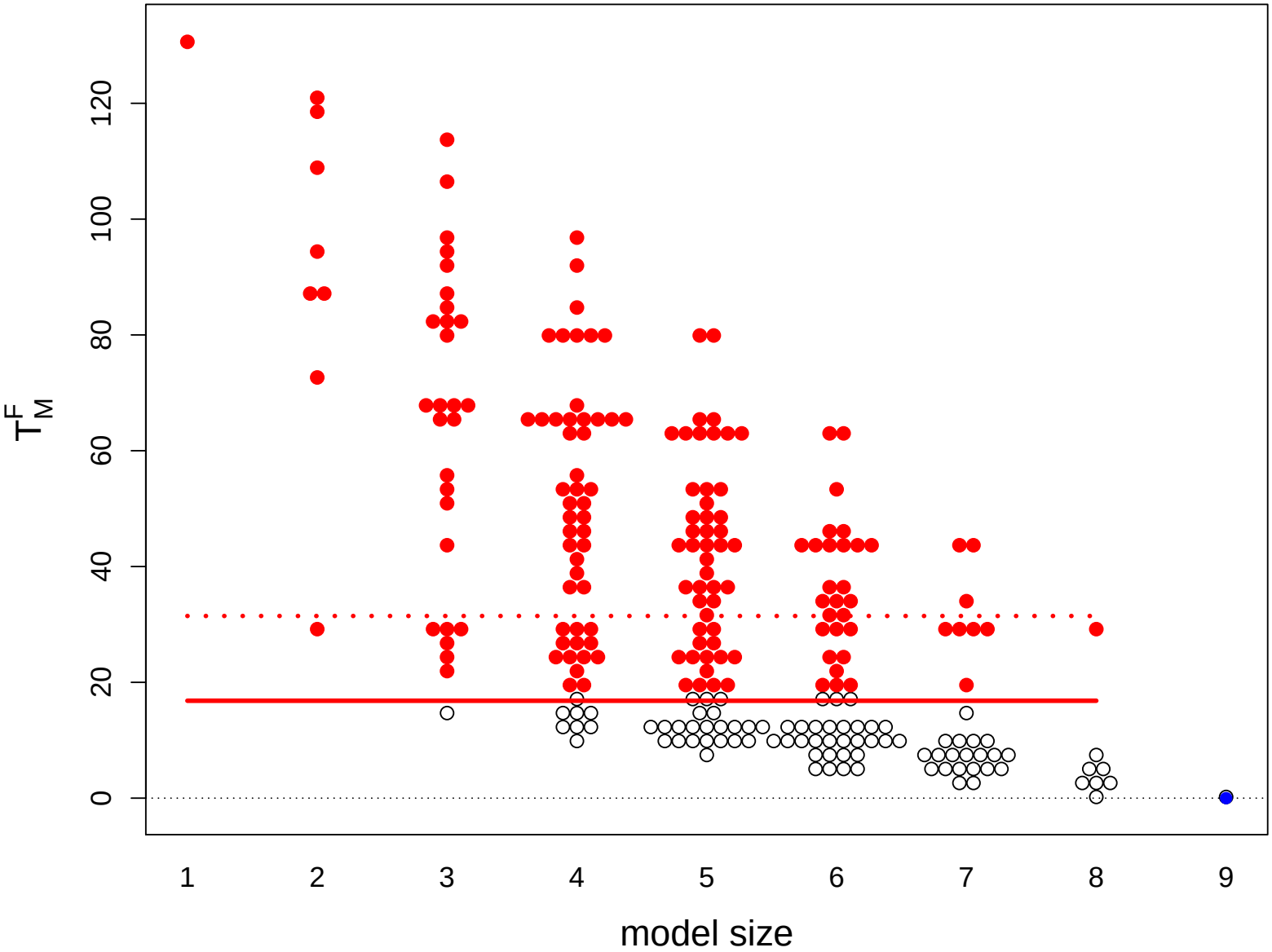
Mallows (1973)

Assumption "correct full model" and Scheffé's adjustment

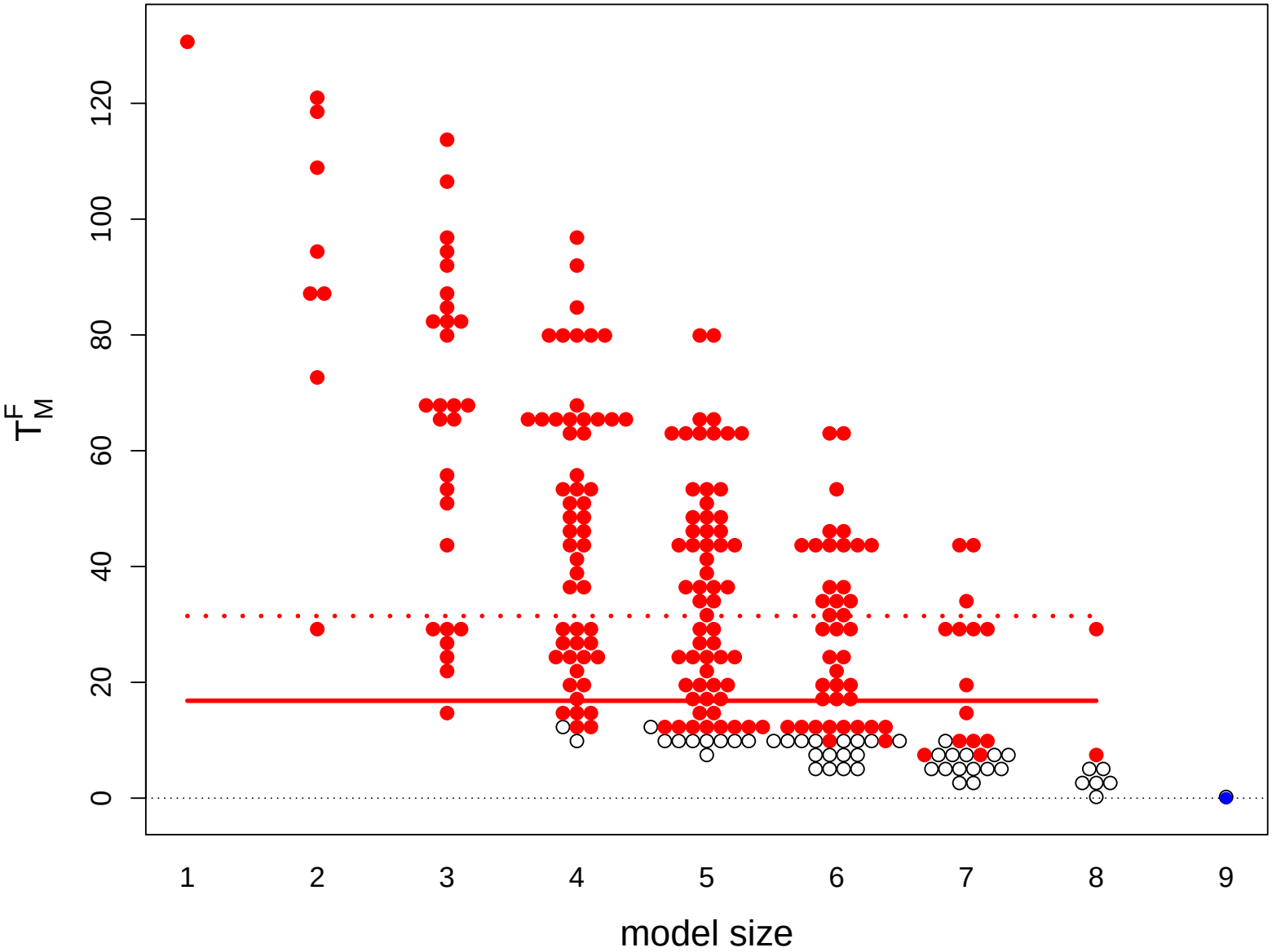
Spjøtvoll (1977)

Assumption "correct full model" and closed testing adjustment

Scheffé's adjustment: $u_{5\%} = 32.5\ \%$



Closed testing adjustment: $u_{5\%} = 18.8\ \%$



Summary

Training set: $n = 67$

Inference	*Adjustment*	*Size*	*Uncertain*	$u_{5\%}$
Inferior	Scheffé	117	138	54.1 %
Non-adequate	Scheffé	172	83	32.5 %
Non-adequate	Closed testing	207	48	18.8 %

Training + test: $n = 97$

Inference	*Adjustment*	*Size*	*Uncertain*	$u_{5\%}$
Inferior	Scheffé	136	119	46.6 %
Non-adequate	Scheffé	179	76	29.8 %
Non-adequate	Closed testing	223	32	12.5 %

Outline

Discussion

Measuring uncertainty

is a statistician's task

Uncertainty in variable selection

High even for 'small' problems,
especially in the presence of collinearity

High-dimensional data

$p \gg n$: strong assumptions needed

Bibliography


Mallows (1973)
Some comments on C_P
Technometrics, 15:661-765


Spjøtvoll (1977)
Alternatives to plotting C_P in multiple regression
Biometrika, 64:1–8


Shmueli (2010)
To explain or to predict?
Statistical Science, 25:289–310