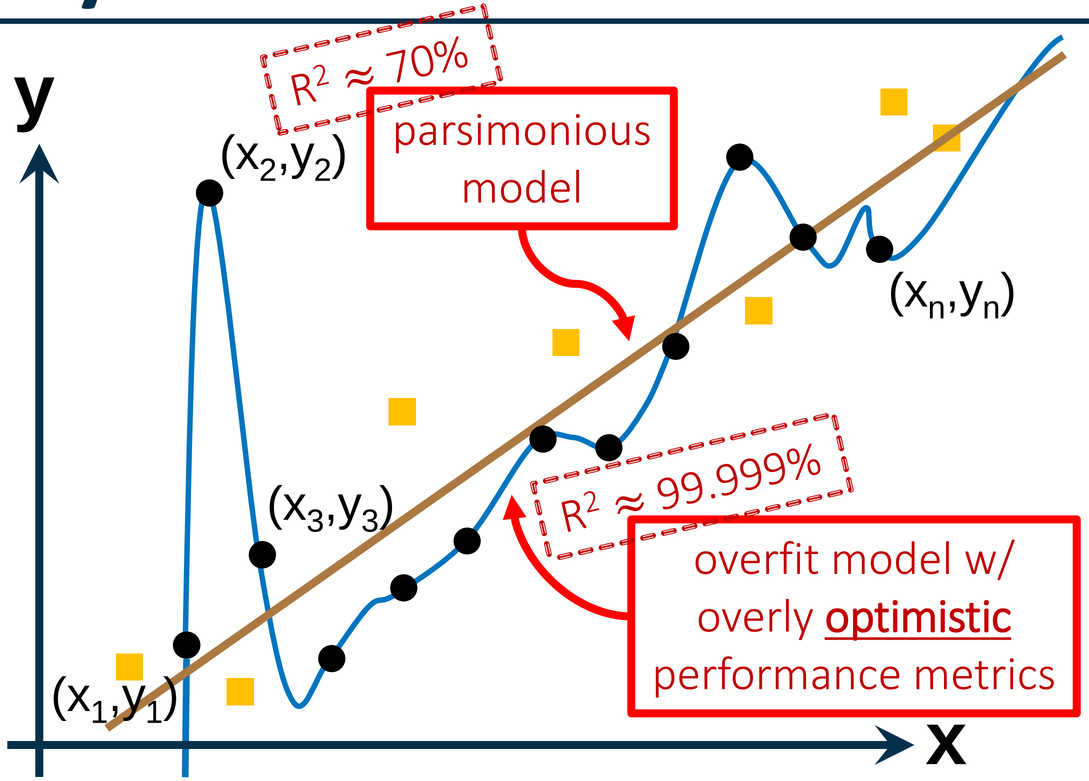
A Framework for Simple and Efficient Bootstrap Validation in SAS®, with Examples

Isaiah Lankham
University of California
Office of the President
Oakland, CA

Matthew Slaughter
Kaiser Permanente
Center for Health Research
Portland, OR

Invited Breakout Session • Statistics • Submission No. 1034

Why We Validate Predictive Models



x	y
x_1	y_1
x_2	y_2
⋮	⋮
x_n	y_n

How We Validate Predictive Models

external validation

original dataset + validation dataset

full data means no loss of statistical power

internal validation

original dataset

partial data for model development + **partial** data for model validation — e.g., split-sample and cross validation

full data for model development + **resampled** data for model validation — e.g., bootstrap validation

Call to Action!

- **Takeaway**: Use internal validation for predictive modeling. Bootstrap validation allows us to internally validate models without loss of statistical power.
- **Tradeoff**: Computing resources.
- **Read**: No. 1034 in SGF 2021 Proceedings
- **Replicate**: Examples available on GitHub

Appendix: **Bootstrap Validation Framework**