



CMU 15-889e

Offline Batch RL:

Quantifying the Quality of

the Estimated Value

Function or Policy 2

Emma Brunskill

Fall 2015

Class Feedback: Key Themes

Areas of Strength

- Exercises in class
- Doing algorithms & proofs on board
- Examples and comparisons between algorithms

Areas for Improvement

- More real life examples
- More comparisons across algorithms
- Spend a bit more time clarifying proof notation

Vast majority said pace is just right



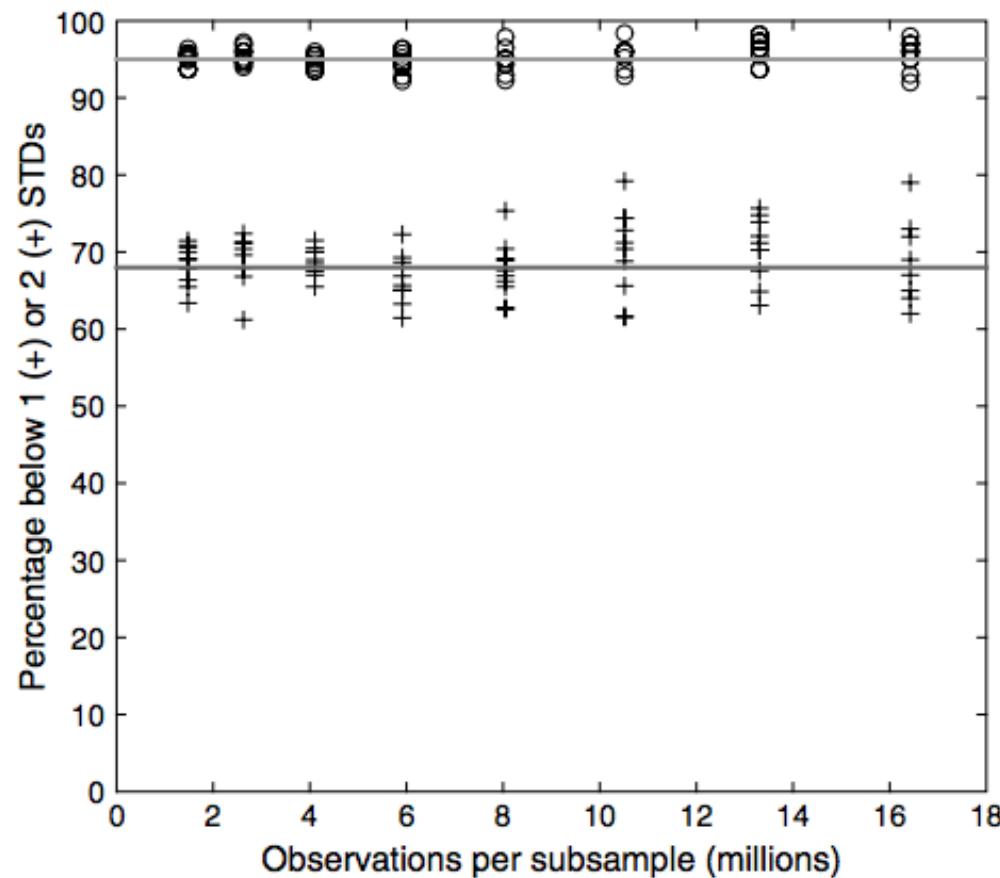
Quantifying the Quality of the Estimated V/Q or Policy

1. Bound estimation error

- Focus is on error due to finite samples
- Does not address approximation error
- If know model class, or have an extremely expressive model class, estimation error may dominate generalization error



Figure 3 The Percentage of the AVF Estimates that Fall Within One ("+" and Two ("O") Standard Deviations from the Value Calculated Based on the Full Data Set



Note. Each "+" and "O" represents a random partition of the full data to subsamples. The discount factor is $\alpha = 0.98$.

Figure 4 The Differences (Marked by “+”) Between the AVF Estimates (in Dollars, and Averaged Over All States) Based on the Calibration Sample and the Validation Sample for the Policy Identified Through an Optimization Process

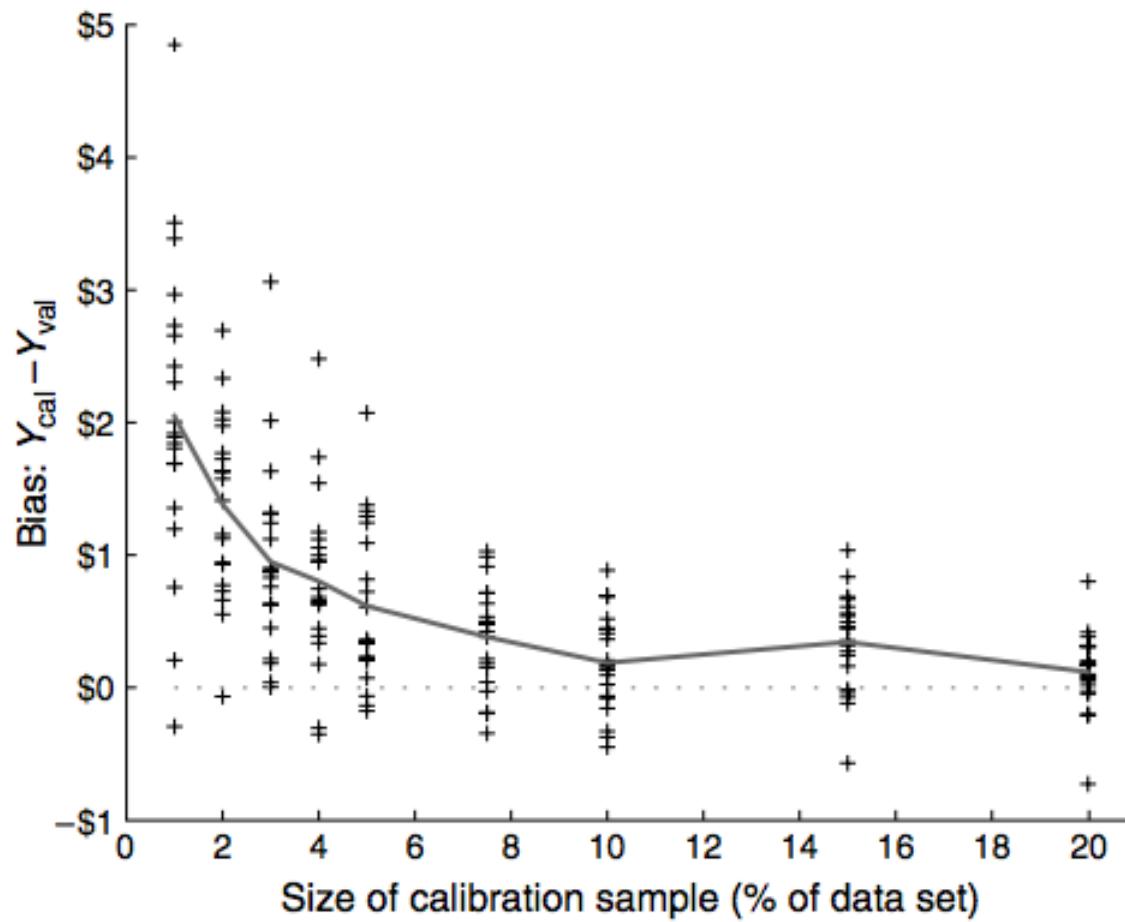
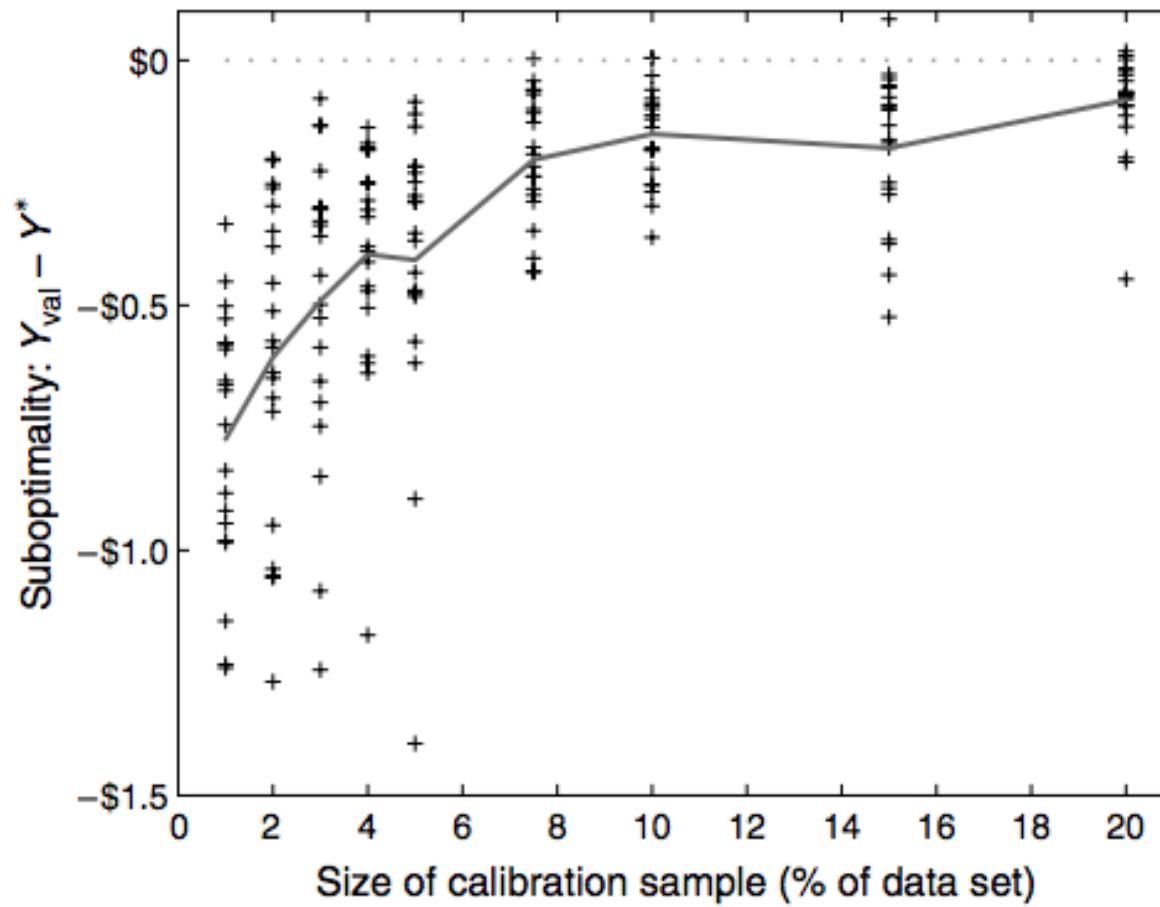


Figure 5 The Differences (Marked by “+”) Between the AVF Estimates (in Dollars) of the Optimal Policy Based on the Calibration Sample and the AVF of the Optimal Policy Found by Optimizing on the Validation Sample



Bounding Estimation Error: Key Points

- Be able to reproduce simulation lemma proof
- Understand key idea about bias and variance for a single policy (how to calculate, what's being approximated)
- Understand problem for control setting and one way to get around
- Know that these approaches ignore approximation error

