

# Causal Modeling-Based Approach for Testing and Improving Credit Decisions Over Time

Dr. Gerald Fahner, FICO  
August 24, 2011



Confidential. This presentation is provided for the recipient only and cannot be reproduced or shared without Fair Isaac Corporation's express consent.

© 2009 Fair Isaac Corporation.

# Broad Approaches and Considerations for Improving Credit Decisions Through Data



- » Champion-challenger testing:
  - » Comparing groups randomly assigned to alternative decision rules
- » Decision modeling and optimization:
  - » Predicting subjects' potential responses to alternative treatments (“What would happen to Nick if...?”)
  - » Means estimating causal effects of credit decisions [1]
  - » Often approached as observational studies
  - » Matching on propensity scores is a promising technique [2]
    - » Transparent - lets you appreciate amount of overlap in covariate distributions between groups receiving alternative treatments
    - » Important - because regression estimates of causal effects are not robust if overlap is small [3]
  - » What if there's little overlap?
    - Limit predictions of causal effects to local overlap regions
    - Improve testing practices to mitigate this problem in the future

# Agenda



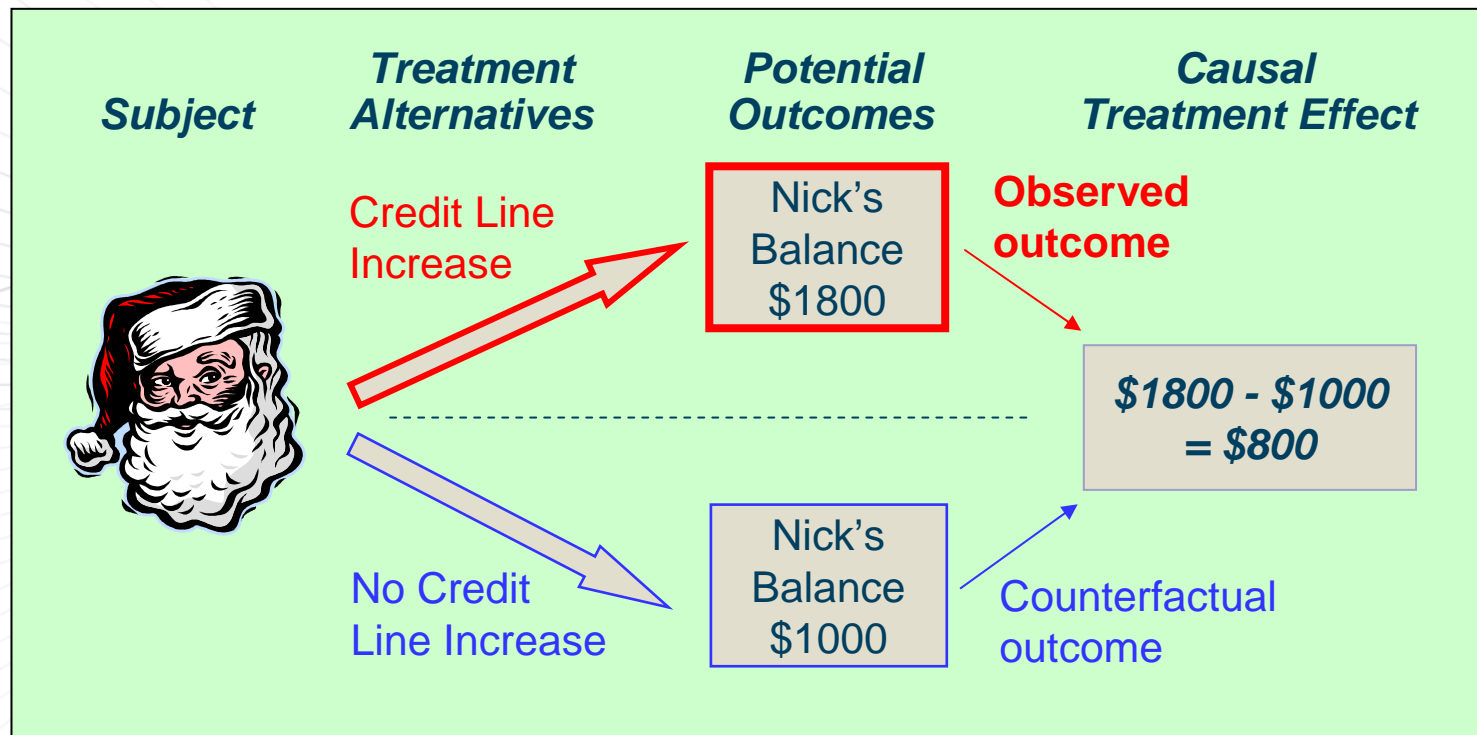
- » Rubin Causal Model and Overlap Issues
- » Thoughts on Improving Test Design
- » Test-and-Improve Cycle Simulations
- » Discussion

# Rubin Causal Model

Simplest Case of 2 Treatment Alternatives



A causal effect is a comparison of potential outcomes for alternative treatments defined on the same subject



**Complication:**  
Only one of the potential outcomes is observed

# Challenges for Estimating Causal Effects of Credit Decisions



- » Selection bias in business-as-usual data:
  - » “Treated” subjects differ from “control” subjects in systematic ways.
    - Need to adjust for subject differences to estimate treatment effects.
  - » Possible pitfalls:
    1. If treatment groups don’t overlap\* in their attribute distributions.
      - Estimation results depend on extrapolation.
    2. If treatment selection depends on side information not available for analysis. → Estimation results could be biased.

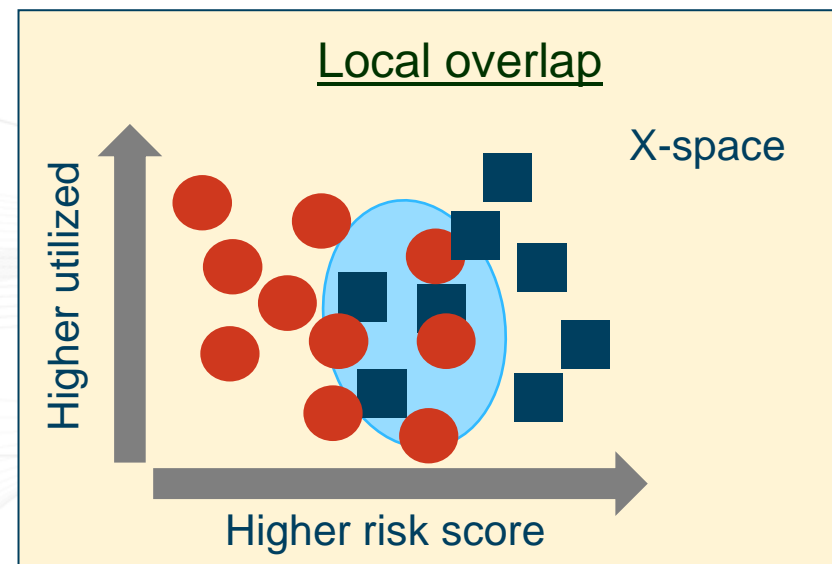
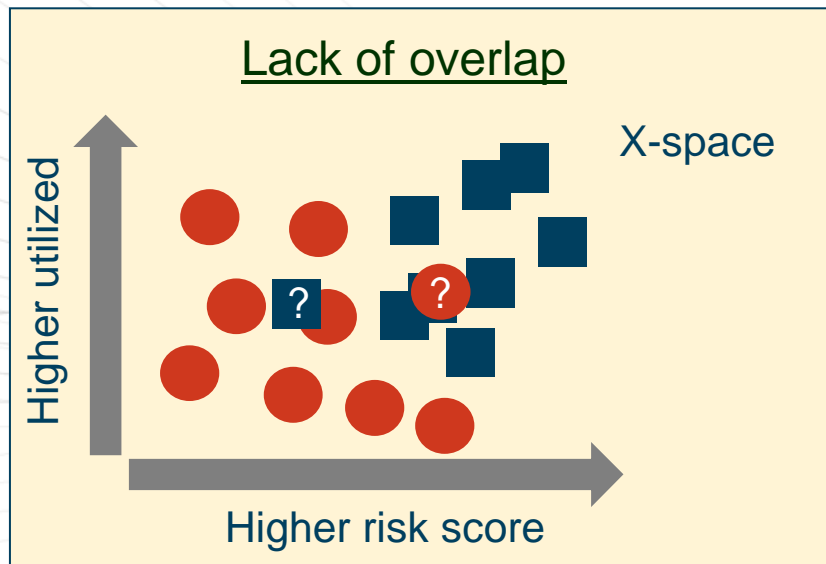
We will assume that treatment selection is solely based on observables which are also available for response analysis. (“Unconfoundedness” assumption). This can be ascertained for rule-based, automated decision systems.

\*Overlap in high dimensions can be understood by developing propensity scores. These model probabilities  $Pr\{T = t | X\}$  of being assigned to treatment alternatives  $t$ , given attribute values  $X$ .

# Objective Data-driven Causal Inference? - A Question of Overlap!

Treatment of subjects:

- No Credit Line Increase
- Credit Line Increase



Counterfactual estimation is problematic.

- Depends on extrapolation.
- Sensitive to model specification.
- Strong functional assumptions needed.
- Substantial domain expertise required.
- Subjective.

Counterfactual estimation within local overlap

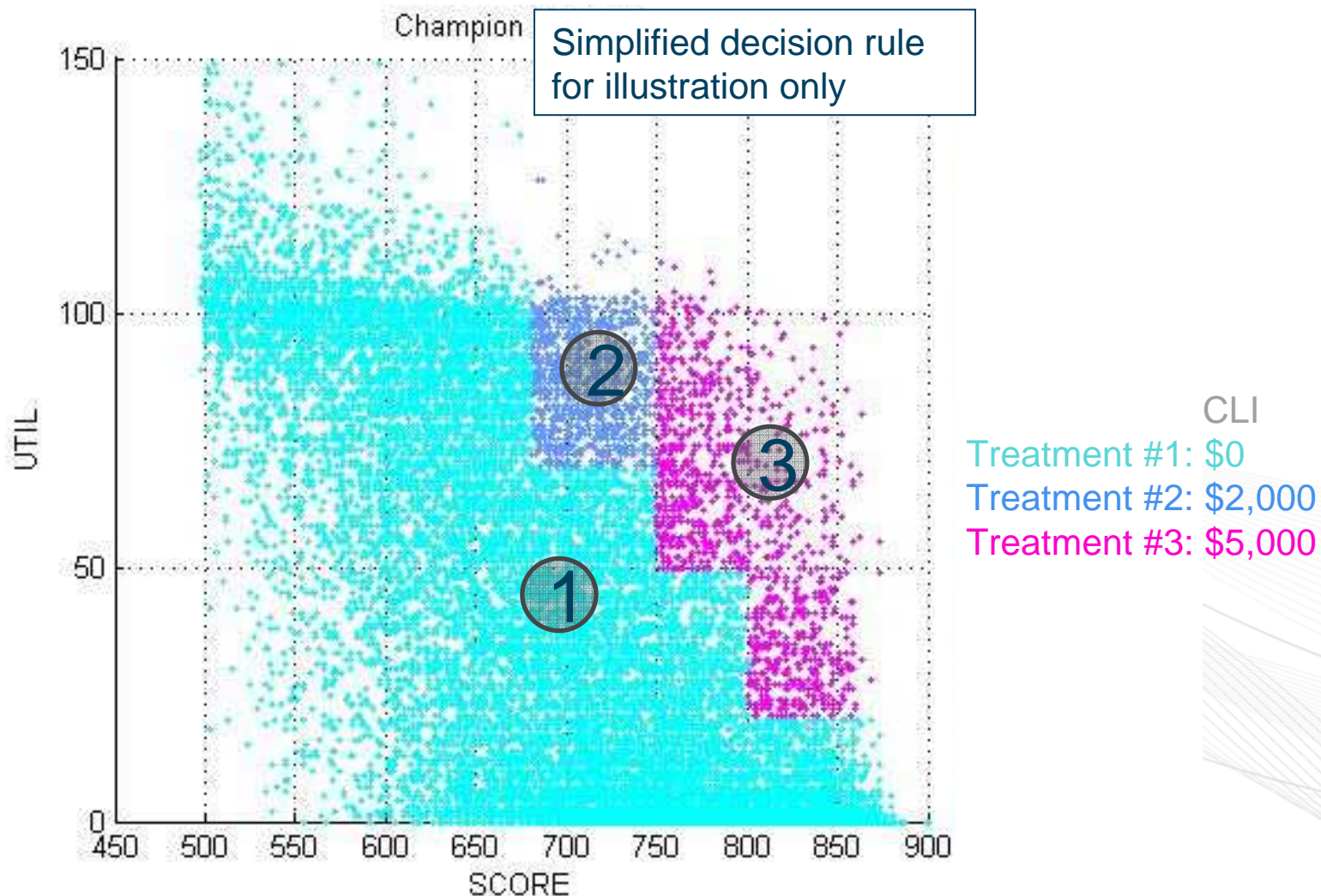
- regions requires interpolation only.
- Can use flexible modern regression techniques making minimal functional assumptions (e.g. GAMs), as long as we restrict predictions to overlap regions.
- Domain expertise less critical (but still helps).

# Agenda



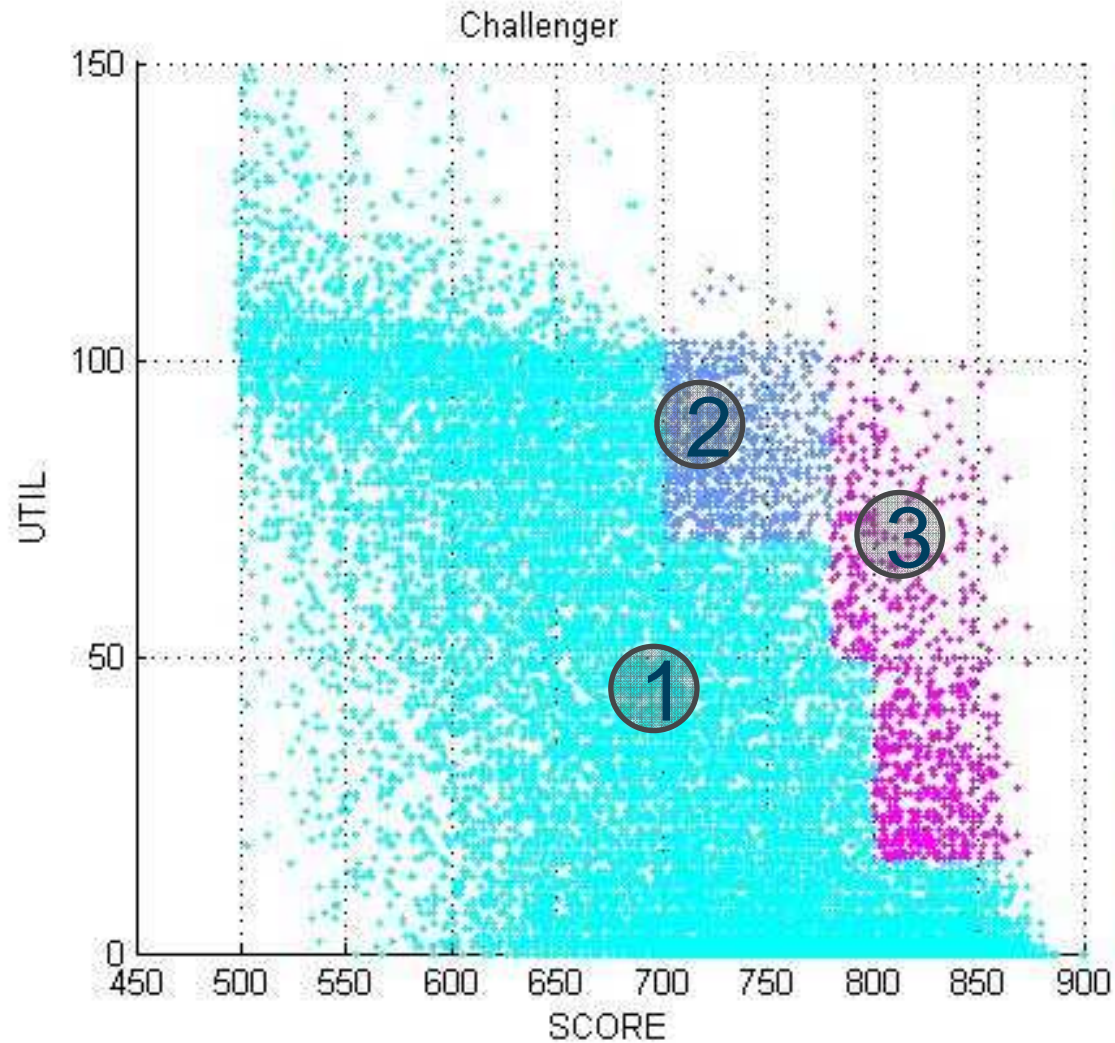
- » Rubin Causal Model and Overlap Issues
- » Thoughts on Improving Test Design
- » Test-and-Improve Cycle Simulations
- » Discussion

# Deterministic Decision Rule for Credit Line Increase Generates No Overlap



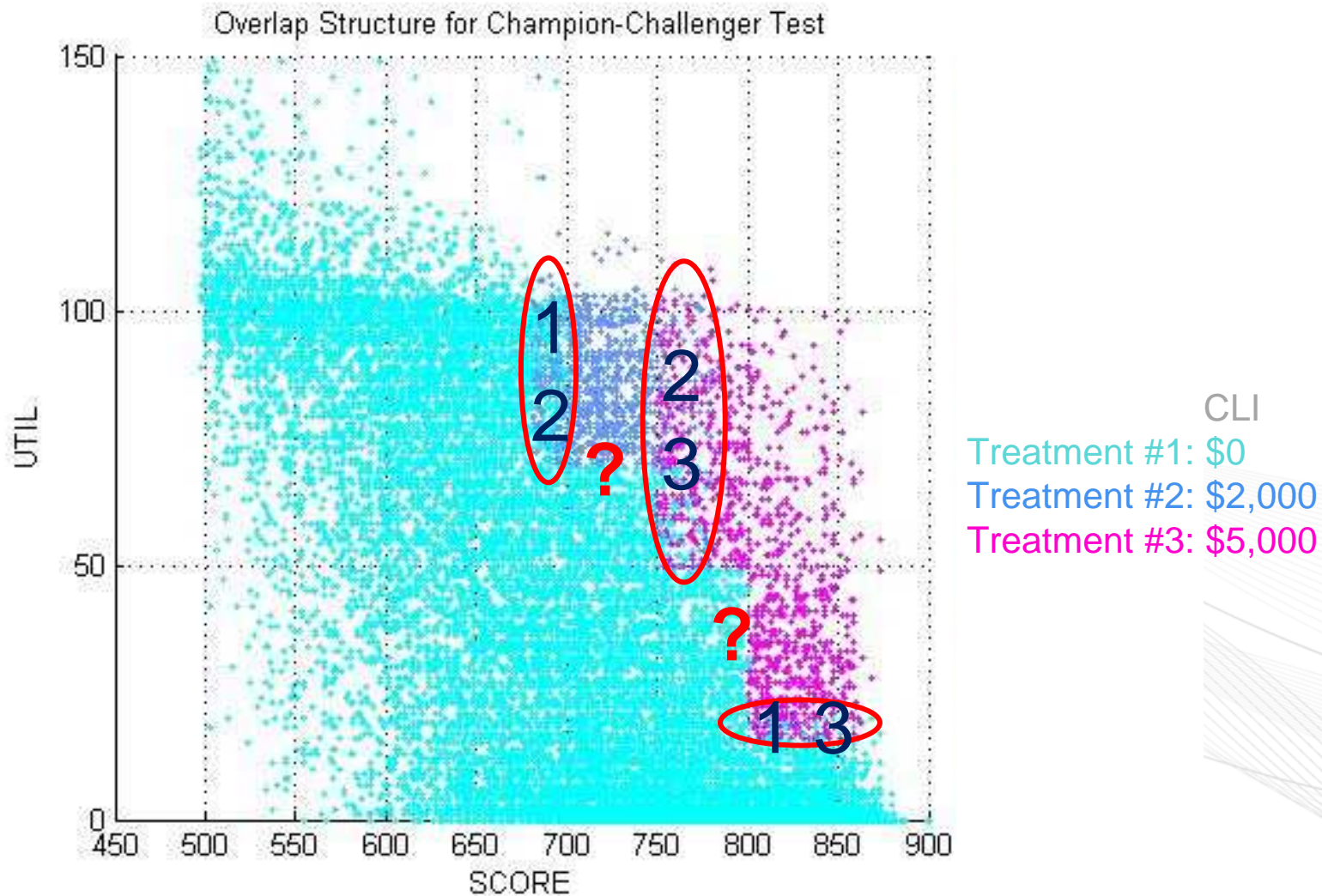
Decision rule is a tree with 4 split levels, 7 leaf nodes.  
SCORE splits at {680, 750, 800}. UTIL splits at {20, 50, 70}

# Challenger Rule Using Somewhat Different Split Values



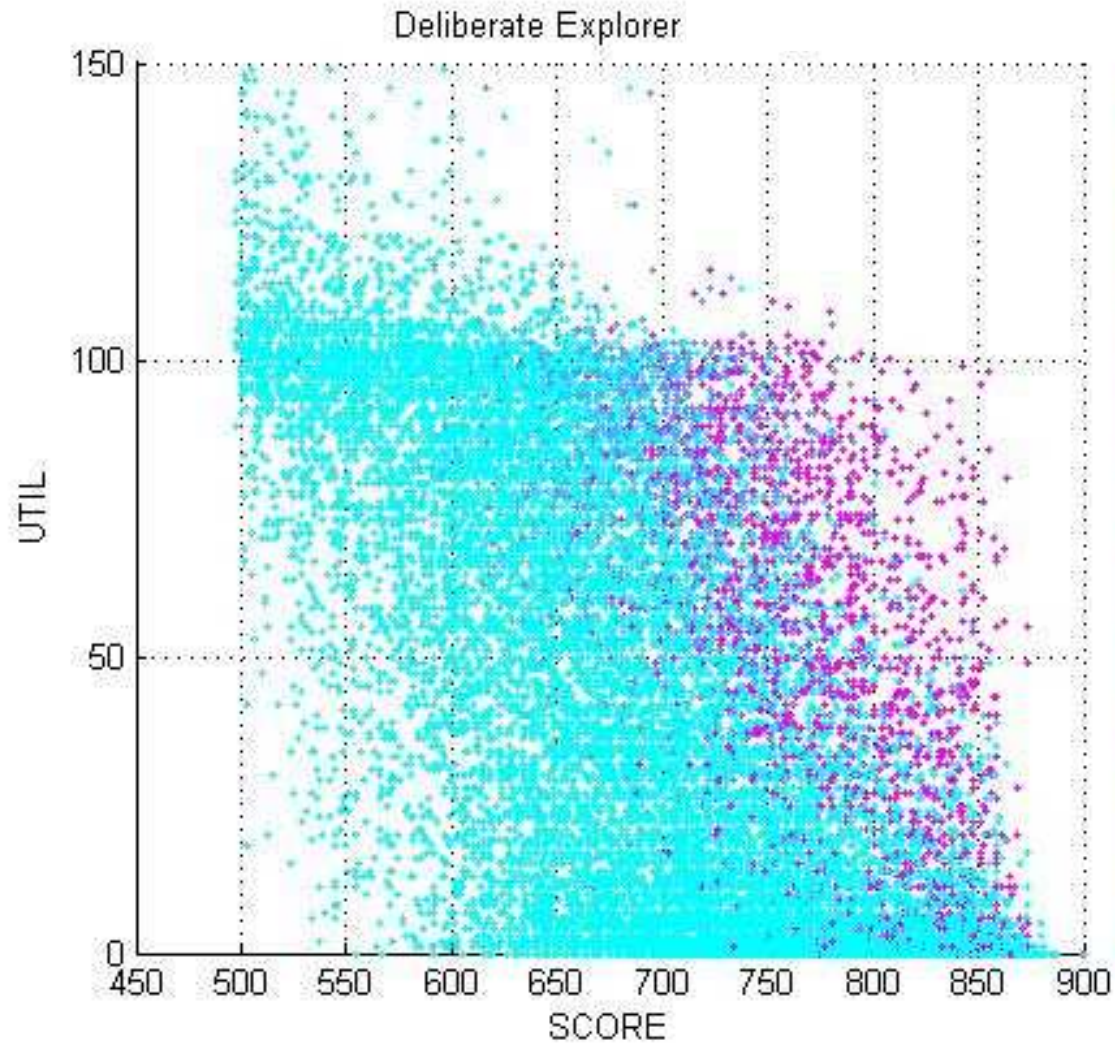
CLI  
Treatment #1: \$0  
Treatment #2: \$2,000  
Treatment #3: \$5,000

# Champion-Challenger Test Generates Limited Local Overlap Regions



Overlap regions are irregular, happenstance.  
Potentially interesting areas for establishing “swap sets” may not be covered.

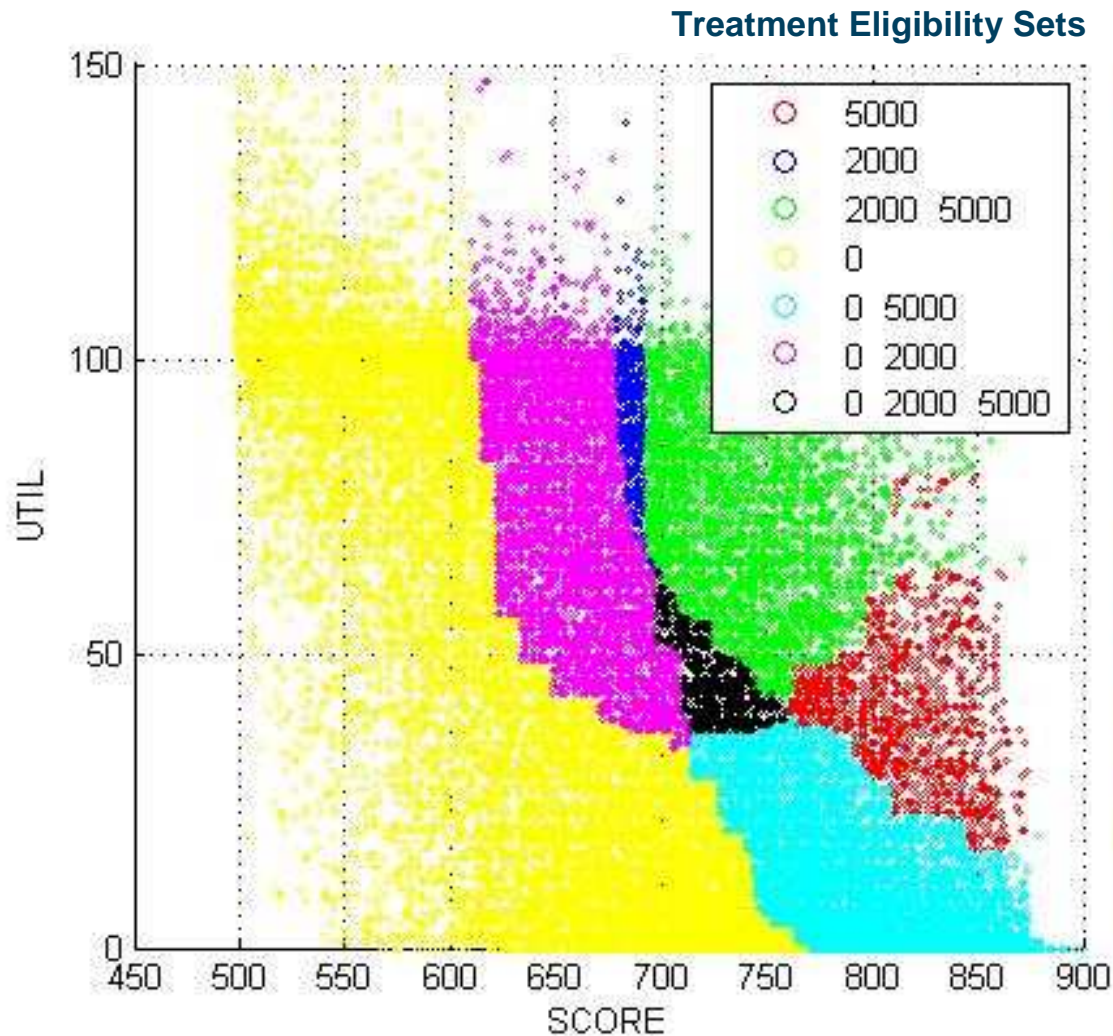
# Proposal for Boundary-Hugging Test Designs



CLI  
Treatment #1: \$0  
Treatment #2: \$2,000  
Treatment #3: \$5,000

Subjects closer to decision boundaries are more likely randomly "flipped" to neighboring treatments.

# Overlap Structure and Treatment Eligibility Sets



For optimization consider restricting eligible treatments for a subject at  $X = (S,U)$  to treatments whose propensities at  $X$  exceed some threshold

Visualization based on estimated propensity scores and thresholding:  
E.g. if  $\Pr\{\text{CLI} = \$0 \mid S,U\} > 0.1$  and  $\Pr\{\text{CLI} = \$2000 \mid S,U\} > 0.1 \rightarrow \text{Color}(S,U) = \text{'magenta'}$

# Agenda



- » Rubin Causal Model and Overlap Issues
- » Thoughts on Improving Test Design
- » **Test-and-Improve Cycle Simulations**
- » Discussion

# Simulating a Hypothetical Credit Card Portfolio

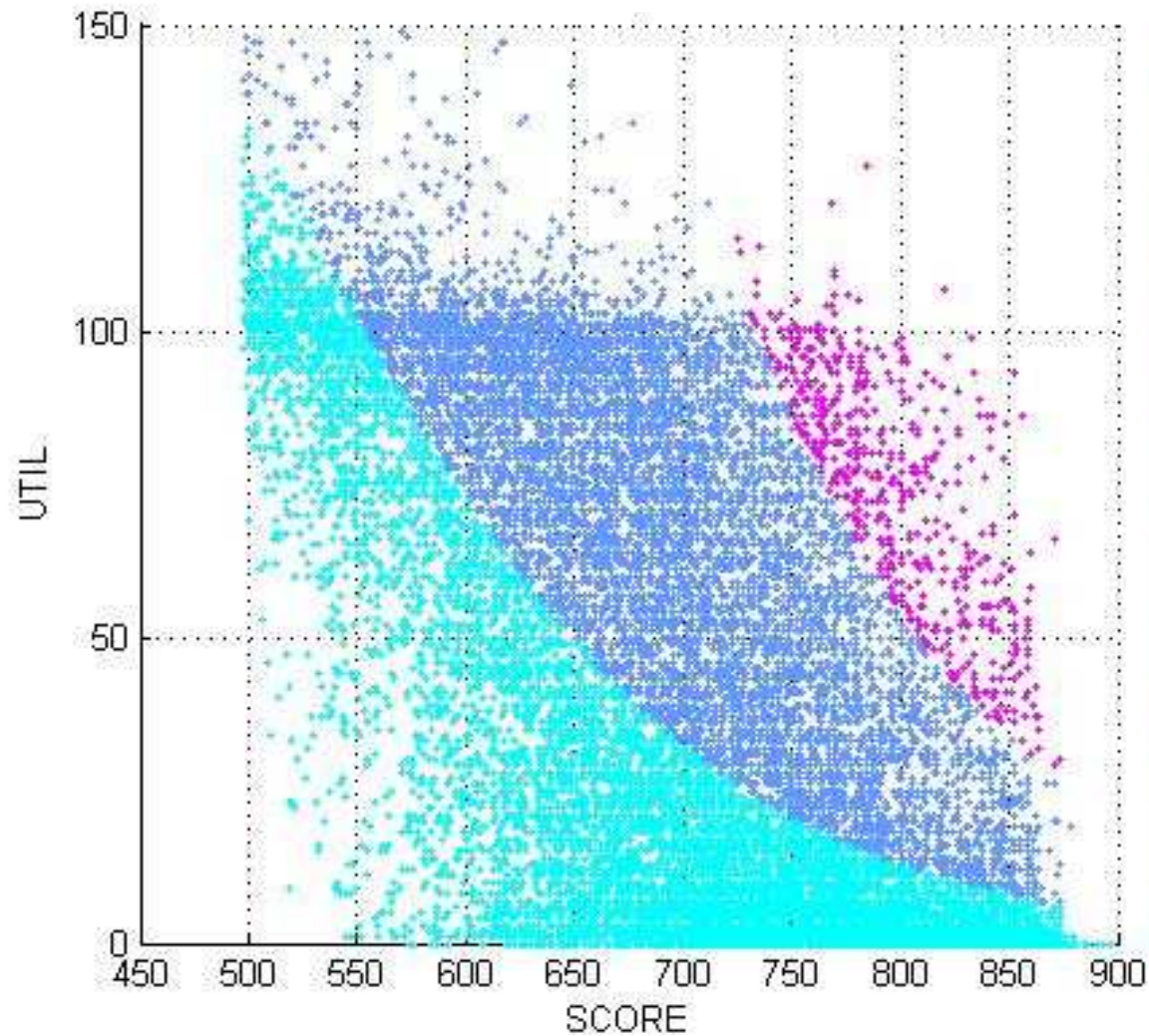


» Posit model\* of the form:

$$\begin{aligned} F(X) &= \text{Joint density of attributes } X = (\text{Score}, \text{Util}) && (\text{Copula simulation}) \\ Y(X, T) &= M(T, X) + \text{noise} && (\text{"Observed" outcomes: Revenue, Loss}) \\ T(X) &= \text{Rule for assigning CLI treatment} && (\text{Boundary-hugging tests}) \end{aligned}$$

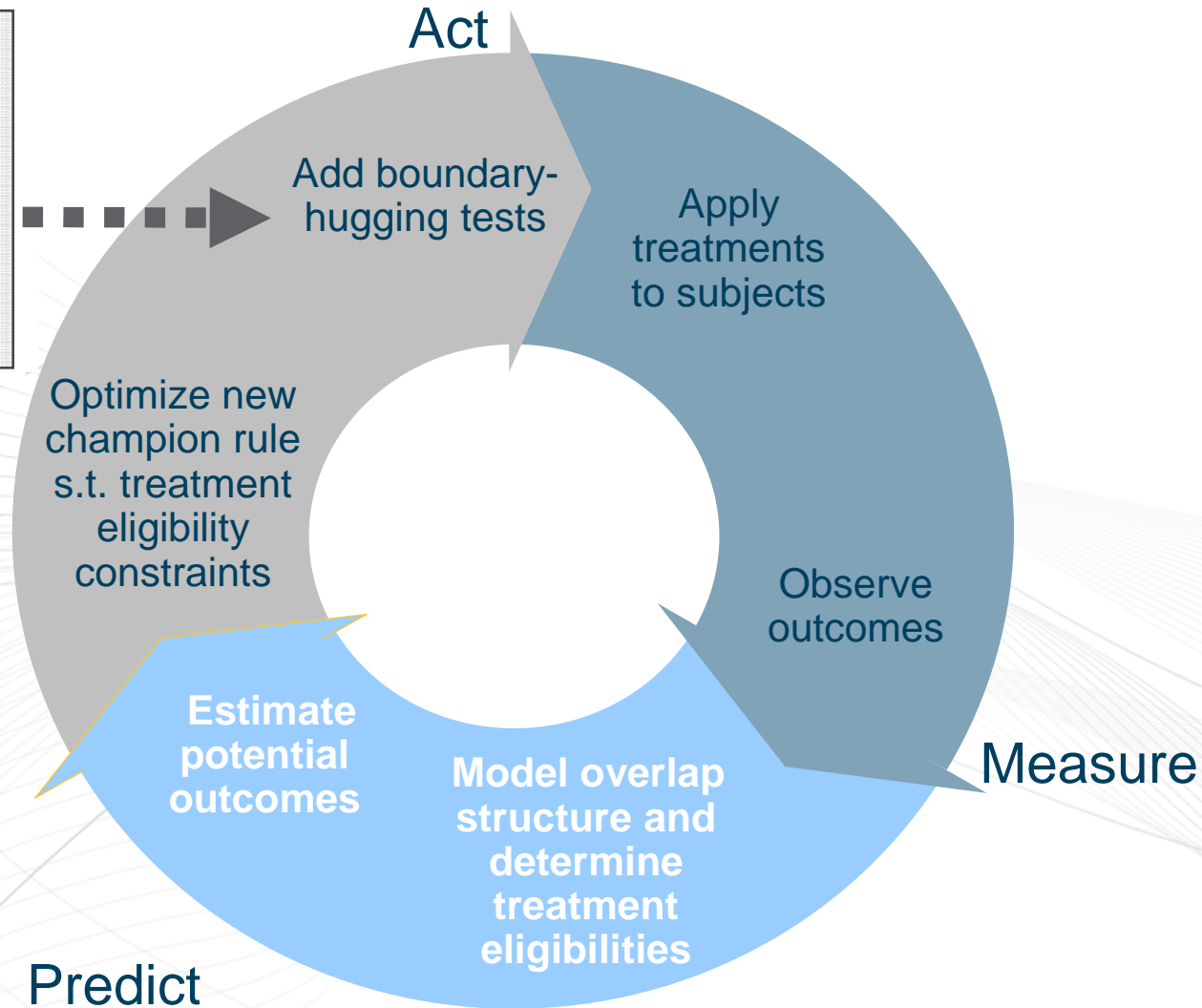
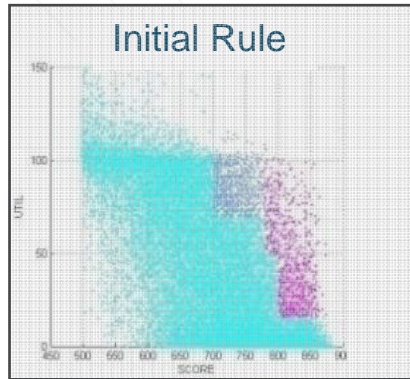
\*For illustration only. No attempt to model comprehensively all important attributes or to describe a particular portfolio.

# Profit-Maximizing Rule from Posited Model



CLI  
Treatment #1: \$0  
Treatment #2: \$2,000  
Treatment #3: \$5,000

# Test-and-Improve Cycle



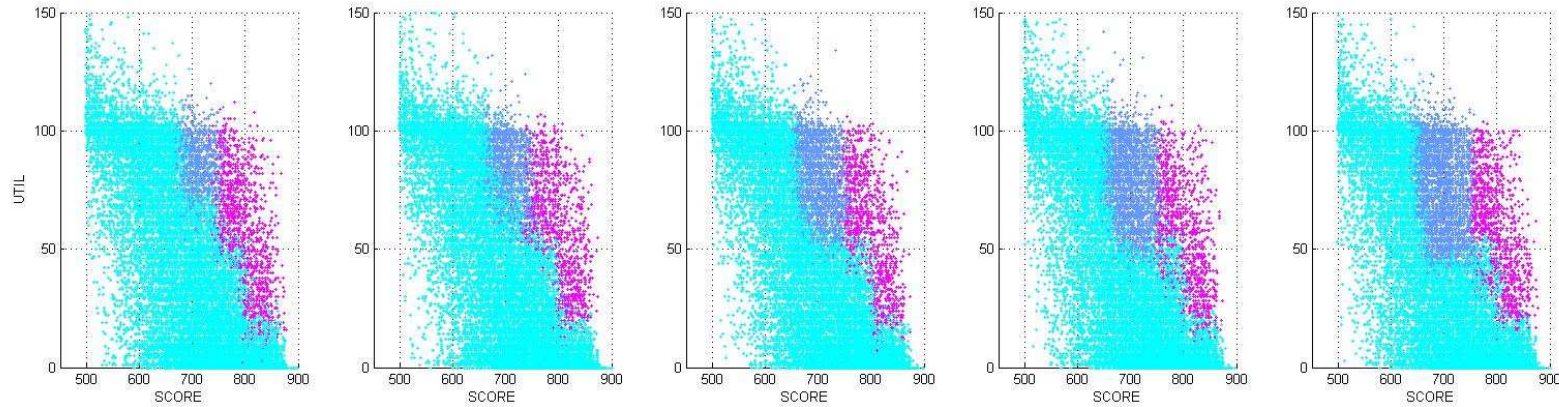
## Predict

(Propensity scores and treatment response functions are estimated by stochastic gradient boosting [4], making minimal functional assumptions).

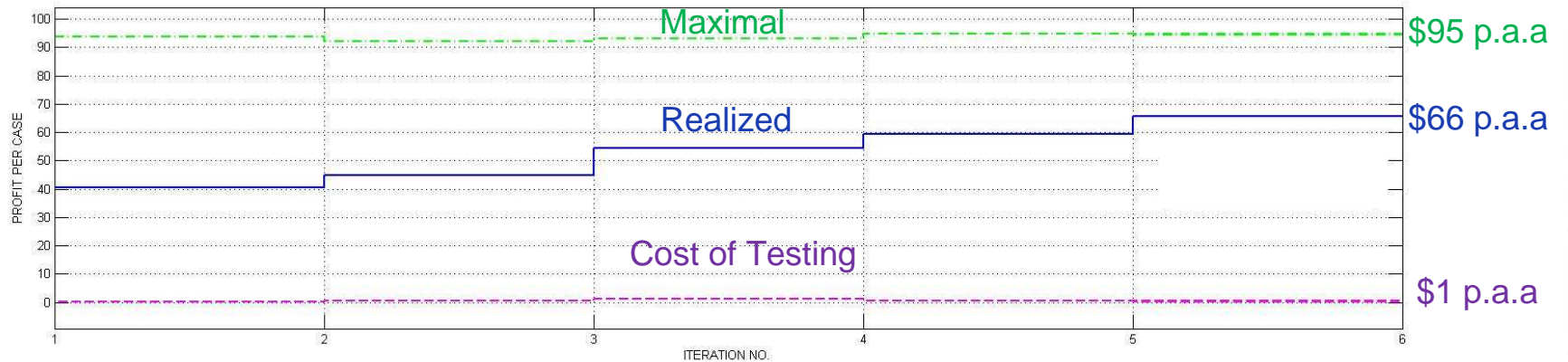
# “Timid Explorer” Test Design Has Negligible Testing Cost. Improves Slowly Towards Optimum



Rule Evolution over 5 cycles



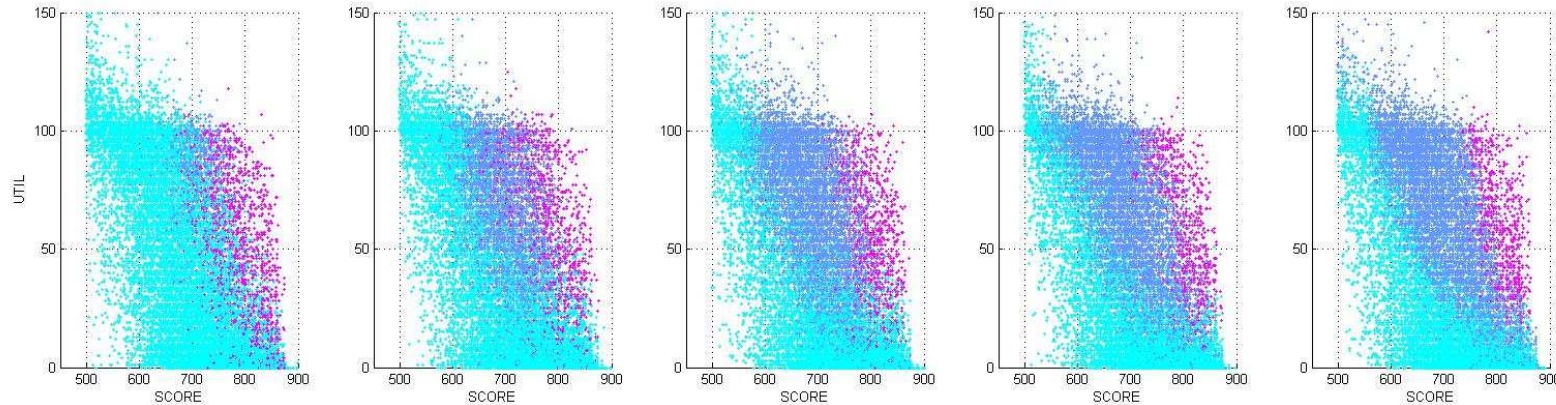
Profit over Time



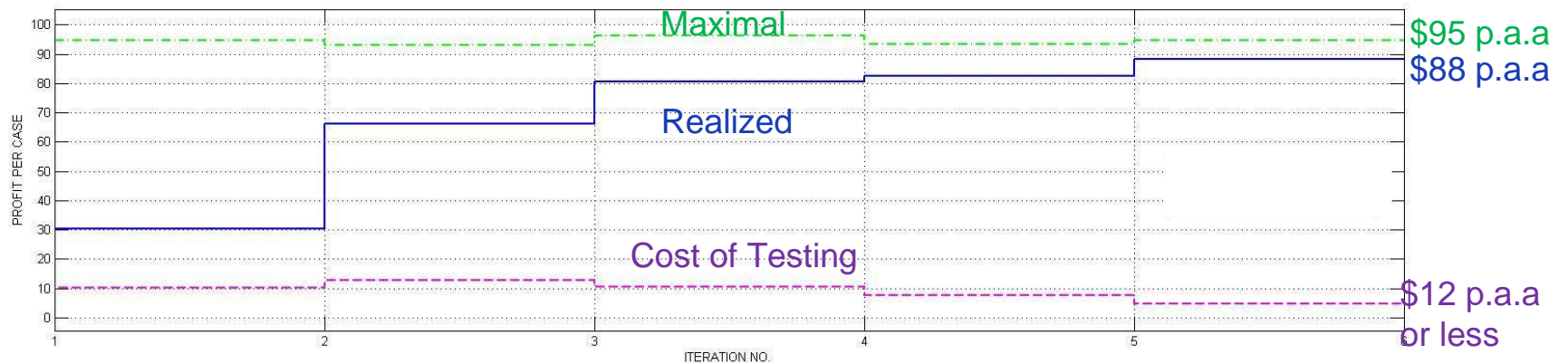
# “Deliberate Explorer” Test Design Improves Much Faster. Cost of Testing is Moderate



Rule Evolution over 5 cycles



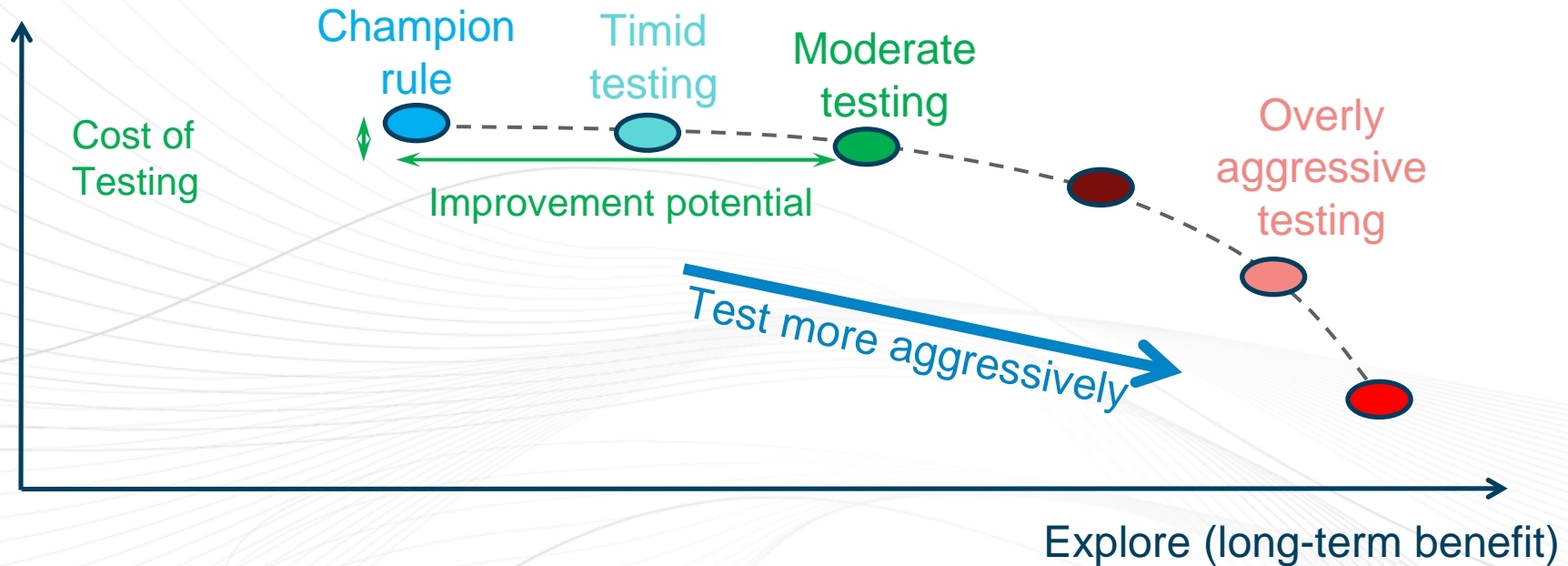
Profit over Time



Aggressiveness of testing was deliberately reduced over the cycles according to a schedule. Intuitively, less exploration is required when getting close to the optimum. In reality unknown optimum will likely evolve. Should never stop testing.

# Exploration-Exploitation Tradeoff

Exploit (short-term benefit)



Experiences from simulation studies and actual project work point at the opportunity to increase overlap substantially while keeping costs for testing in check.

- » Overlap is key for learning to improve credit decisions from data.
- » Propensity scores help us to understand overlap situation. Where overlap lacks, beware of risky extrapolation, ramp up testing
- » Champion-challenger tests aren't designed with overlap in mind, potentially leaving improvement opportunities unexplored
  - Boundary-hugging test designs appear as a powerful alternative
  - Simulations provide insight into the merits of alternative experimental designs and the exploration-exploitation tradeoff

Where on the exploration-exploitation curve do you operate?

# References



- [1] “Estimating Causal Effects of Credit Decisions Using Propensity Score Methodologies”, by G. Fahner. *Edinburgh Scoring Conference Proceedings*, 2009. <http://www.crc.man.ed.ac.uk/conference/archive/2009/presentations/Paper-36-Presentation.pdf>
- [2] “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, by P. R. Rosenbaum and D. B. Rubin. *Biometrika*, Vol. 70, No. 1. (Apr., 1983), pp. 41-55.
- [3] “*Data Analysis Using Regression and Multilevel/Hierarchical Models*”, by A. Gelman and J. Hill. Cambridge University Press, 2007.
- [4] “Stochastic Gradient Boosting”, by J. Friedman. *Computational Statistics and Data Analysis*, 38, 367, 2002.