Risk Inflation of Sequential Tests

Bob Stine and Dean Foster
Dept of Statistics, Wharton School
University of Pennsylvania

JSM, August 4, 2013

Manuscript and slides available at
www-stat.wharton.upenn.edu/~stine

Wharton
Department of Statistics

Thank you, NSF! (#1106743)

Plan

- Example of sequential tests
  - Streaming feature selection
- Alpha investing
  - Test a possibly infinite sequence of hypotheses
  - Flexible control of expected false positives
- Risk analysis
  - Exact analysis via Bellman equations
  - Feasible set of possible risks
  - Comparison of procedures

Streaming Feature Selection

- Canonical problem
  - Pick predictors for regression

    $\hat{y} = b_0 + b_1\, x_1 + \ldots\ ???\ \ldots + b_k\, x_k$
- Streaming selection
  - Have current model
  - External source offers new candidate z
  - Decide whether to add z to model
- Novelties
  - Choice of z may depend on prior outcomes

    Construction of interactions, transformations
  - Can be done very fast

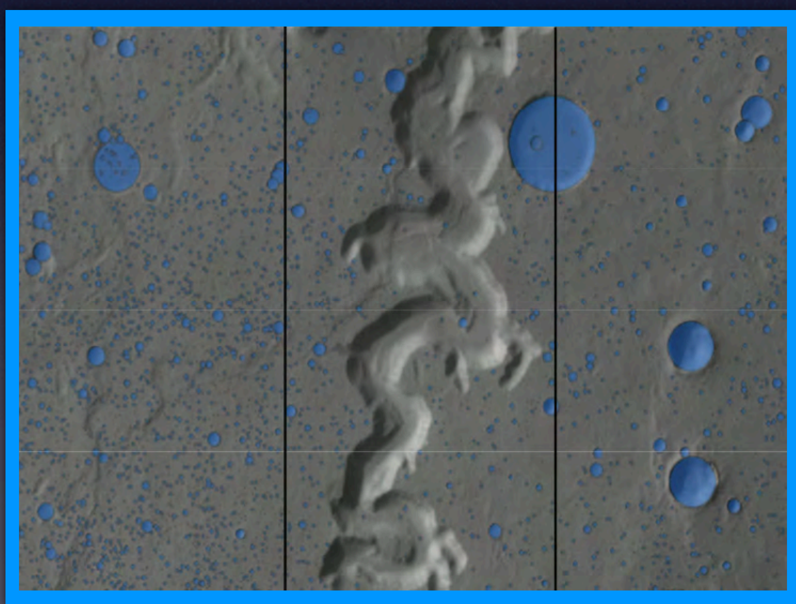
    VIF regression
  - Does not require all possible x_j at start

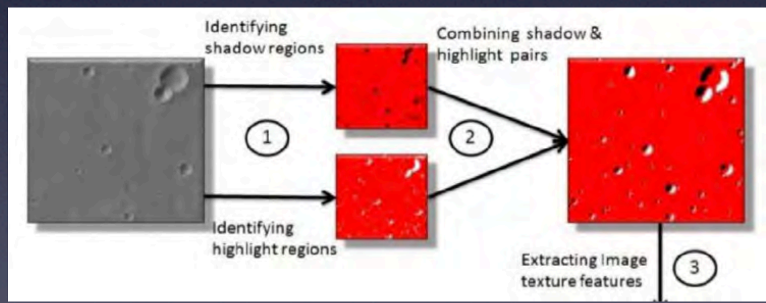
    Image processing, database query. Collection may be infinite…

Example

- Image processing
  - Crater recognition

    WU et al (2013, IEEE Trans Pattern Analysis)
  - Build features sequentially from image
  - Too complex, slow to construct every feature





Figures from Wu et al (2013)

Question for Streaming

- How to control variable selection?
  - Full domain of predictors not available
- Cross-validation
  - Sacrifices data for fitting, estimation
  - Need repeated CV to reduce variation
- Alpha investing
  - Designed for sequential testing
  - Proven to control expected false discovery ratio
- What about risk?
  - Does control of mFDR at typical rates (e.g. 5%) produce estimates with small risk?

~~Bonferroni~~ ~~FDR~~ ~~AIC~~

mFDR

Wharton
Department of Statistics

Alpha-Investing

- Test sequence of hypotheses $H_1, H_2, \ldots$
  - Rejecting H_j provides power for subsequent tests
  - Provable control of expected false discovery rate
- Alpha wealth
  - Initial allowance $W_0=\omega$ for Type I error
  - Invest some wealth $0 \leq \alpha_i \leq W_{i-1}$ in test of H_i
  - Compute p-value p_i of test of H_i
  - Gain wealth for subsequent tests if reject

$$W_i = W_{i-1} - \alpha_i + \omega\, I\{p_i \leq \alpha_i\}$$

earn ω if reject

some set $\omega=0.05$

- Comments
  - Investing: Spend α_i to test, but can gain ω
  - Flexible: Variety of rules for picking α_i

Wharton
Department of Statistics

Foster and Stine, 2008, JRSS-B

Risk

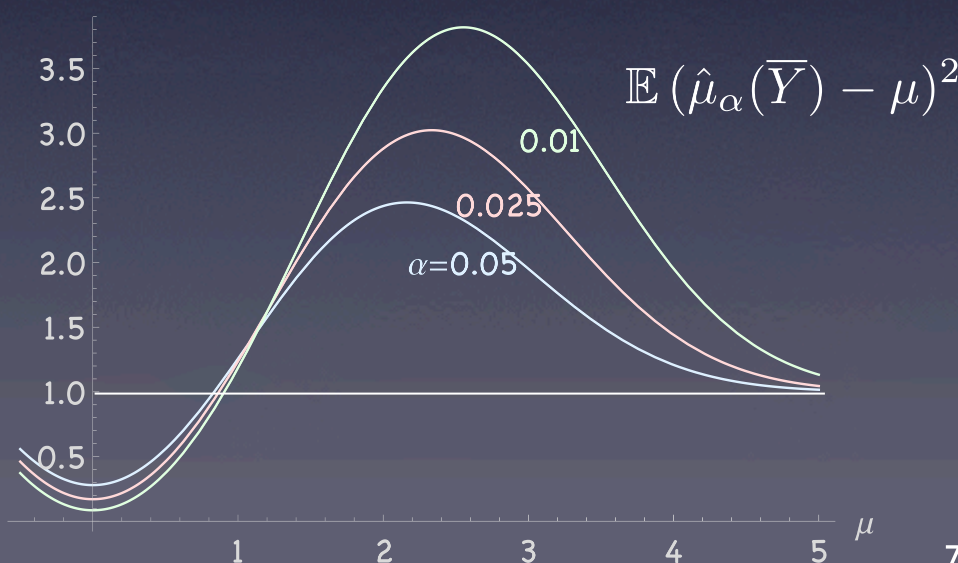
- Idealized problem
  - H_j: $\mu_j = 0$ vs $\mu_j \neq 0$
  - Observe means
  - Independent

$$\overline{Y}_1, \overline{Y}_2, \ldots \overline{Y}_p, \quad \overline{Y}_j \sim N(\mu_j, 1)$$

- Estimator
  - Testimator
  - Hard thresholding

$$\hat{\mu}_\alpha(\overline{Y}) = \begin{cases} \overline{Y} & z_\alpha^2 \leq \overline{Y}^2 \\ 0 & \text{otherwise} \end{cases}$$

- Risk of testimator
  - Well known in conventional testing
  - Unknown for sequential



Wharton
Department of Statistics

Cumulative Sequential Risk

- Cumulative risk

$$R(\hat{\boldsymbol{\mu}}_\alpha, \boldsymbol{\mu}) = \mathbb{E} \sum_{j=1}^{p} \Big(\hat{\mu}_\alpha(\overline{Y}_j) - \mu_j\Big)^2$$

- Recursion for risk

$$\begin{aligned} R(\hat{\mu}(\alpha(\cdot), W_0, \omega), \mu_{1:p}) &= R(\hat{\mu}_{\alpha(W_0)}, \mu_1) + \mathbb{E} \textstyle\sum_{j=2}^{p} R\Big(\hat{\mu}_{\alpha(W_{j-1})}, \mu_j\Big) \\ &= R(\hat{\mu}_{\alpha_1}, \mu_1) \\ &\quad + r_{\mu_1}(\alpha_1)\ R\Big(\hat{\mu}(\alpha(\cdot), W_0 - \alpha_1 + \omega, \omega), \mu_{2:p}\Big) \\ &\quad + (1 - r_{\mu_1}(\alpha_1))\ R\Big(\hat{\mu}(\alpha(\cdot), W_0 - \alpha_1, \omega), \mu_{2:p}\Big) \end{aligned}$$

- Worst case mean process

$$\mu_1 = \arg\max_m \ \Big\{ \ R(\hat{\mu}_{\alpha_1}, m) + r_m(\alpha_1)\ \max_{\mu_{2:p}} R(\hat{\mu}(\alpha, W_0 - \alpha_1 + \omega, \omega), \mu_{2:p}) + (1 - r_m(\alpha_1)) \max_{\mu_{2:p}}\ R(\hat{\mu}(\alpha, W_0 - \alpha_1, \omega), \mu_{2:p}) \Big\}.$$

initial wealth W_0

payout if reject ω

$$r_\mu(\alpha) = P(\text{reject}) = \Phi(\mu - z_\alpha) + \Phi(-\mu - z_\alpha)$$

Computation

- Bellman equations
  - See paper (on-line)
- State dependent
  - Carry all characteristics of estimator
- Size of state space
  - Number tests × States of Est 1 × States of Est 2
- Implications
  - Oracles are nice (no state space)
  - Investing procedure depends only on wealth
  - Wealth tracked on discrete grid

Wealth Function

- Minimize state dependence
  - Necessary in Bellman recursions
- Write spending rule as function of wealth

  $$\alpha_j(\text{history}) = \alpha(W_{j-1})$$

  - Sacrifice rejection history
- Two examples
  - Geometric: Spend fraction of available wealth

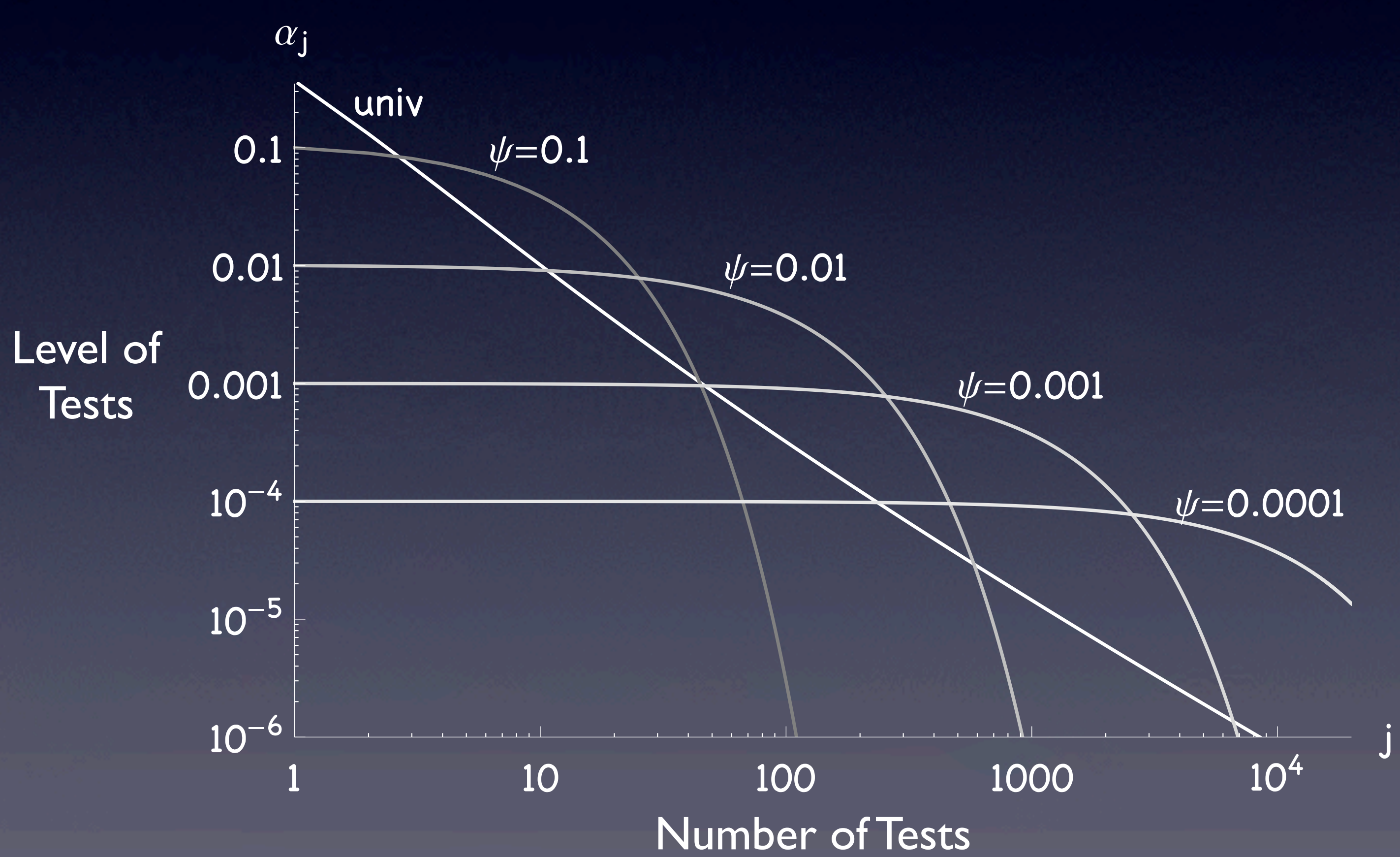
    $$\alpha_g(w) = \psi\, w \qquad 0 < \psi < 1$$

  - Universal
    Spend diminishing fraction of wealth

    $$\alpha_u(w) = w - \frac{1}{\log_2(1 + 2^{1/w})}$$

Universal Rule

- Spends almost as much as each geometric when that geometric rule is most powerful



Wharton
Department of Statistics

Oracle

- Compare risk of realizable estimator to that obtained by an oracle.

  Easier to compute since oracle has no wealth constraint

- Risk inflation oracle
  - Knows whether $\mu_j^2 < 1$
  - Risk is $\min(\mu_j^2, 1)$

  $$\tilde{\mu}(\overline{Y}) = \begin{cases} 0 & \mu^2 < 1 \\ \overline{Y} & \text{otherwise} \end{cases}$$

  Either all bias for small means or all variance for large means

- Bounds for conventional estimation

$$2\log p - o(\log p) \leq \sup_{\mu} \frac{1 + R(\hat{\boldsymbol{\mu}}_\alpha, \boldsymbol{\mu})}{1 + \inf_\eta R(\hat{\boldsymbol{\mu}}_\eta, \boldsymbol{\mu})} \leq 2\log p + 1 \; .$$

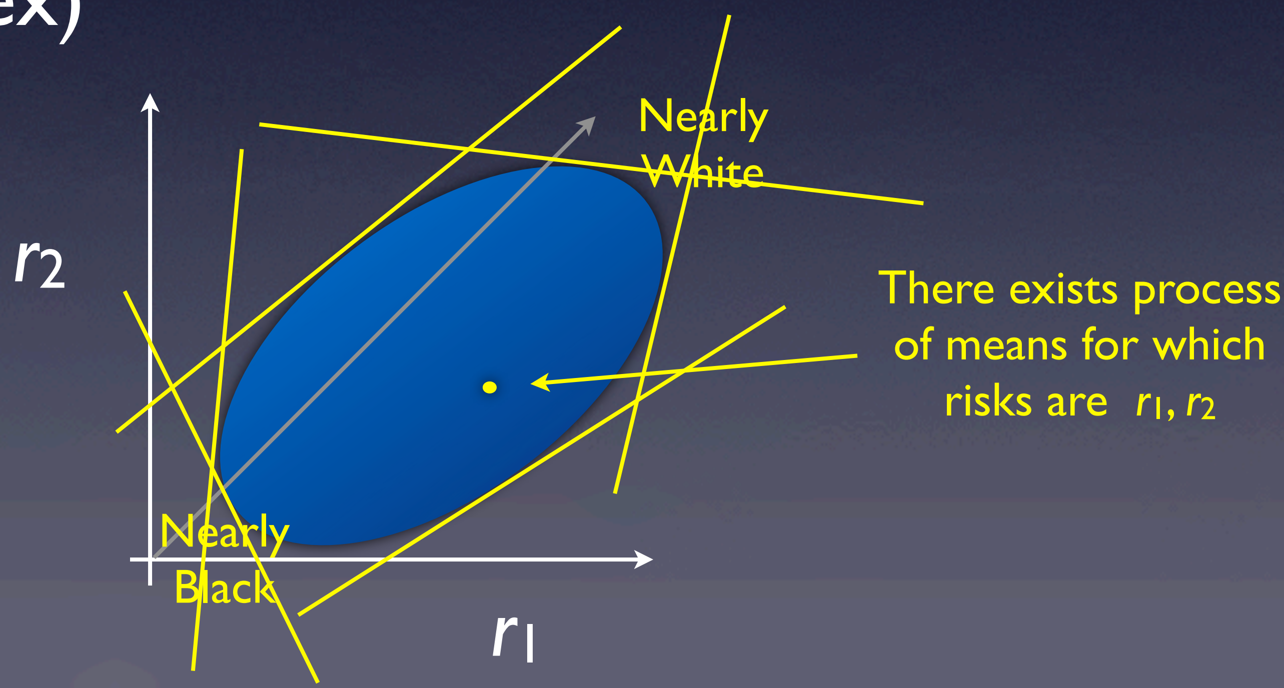
Wharton
Department of Statistics

Feasible Risk Set

- Definition
  - Arbitrary stochastic process $\{\mu_1, \mu_2, \ldots \mu_n\}$
  - Two "estimators"

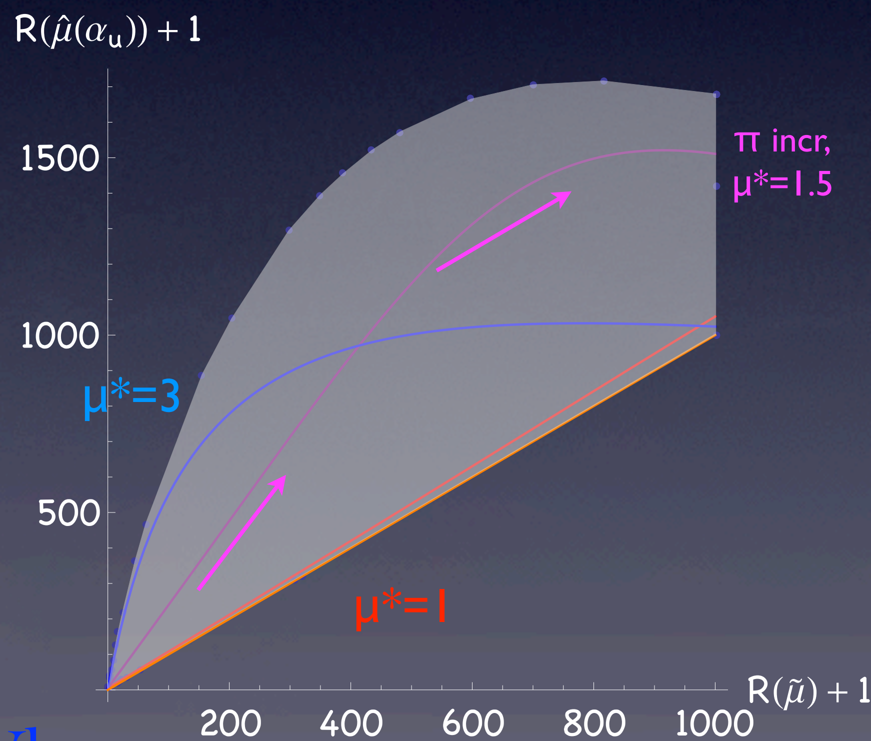
$$\mathcal{R}_p(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2) = \{(r_1, r_2) : \exists \boldsymbol{\mu} \text{ s.t. } r_1 = \mathbb{E}_{\boldsymbol{\mu}} R(\hat{\boldsymbol{\mu}}_1, \boldsymbol{\mu}),\ r_2 = \mathbb{E}_{\boldsymbol{\mu}} R(\hat{\boldsymbol{\mu}}_2, \boldsymbol{\mu})\}$$

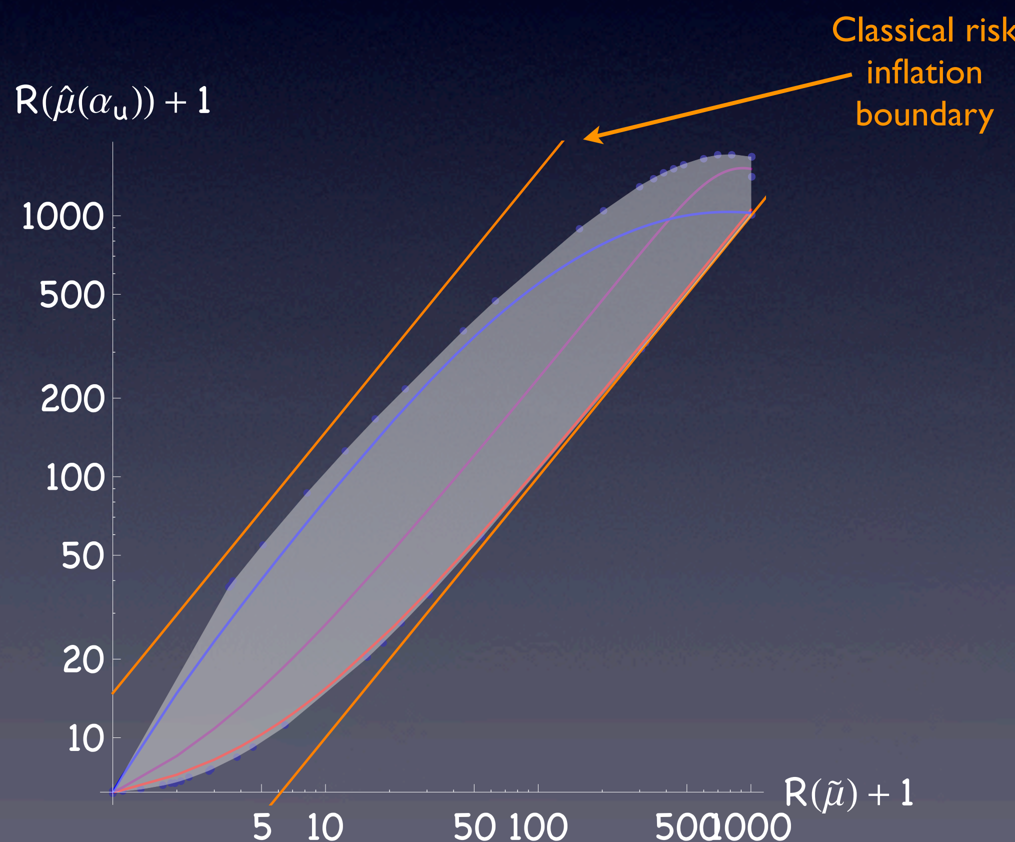
- Graph (convex)



Feasible Risk Set

- Risks of universal and oracle, *p=1000*
- Paths associated with simple models $\mu_j = \begin{cases} \mu^* & \text{w.p. } \pi \\ 0 & \text{otherwise} \end{cases}$
- Logs emphasize nearly black models

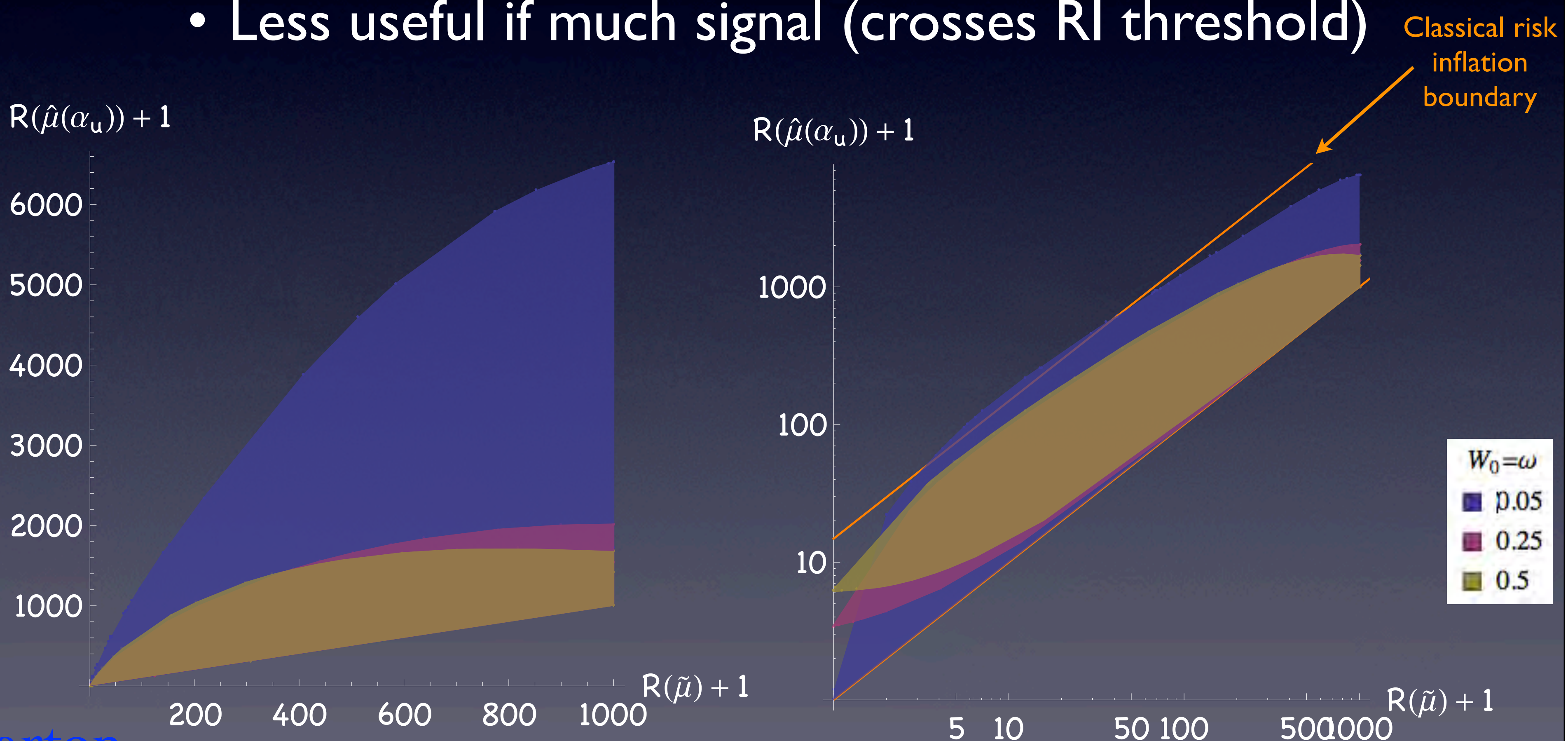




Wharton
Department of Statistics

Feasible Set, varying W_0

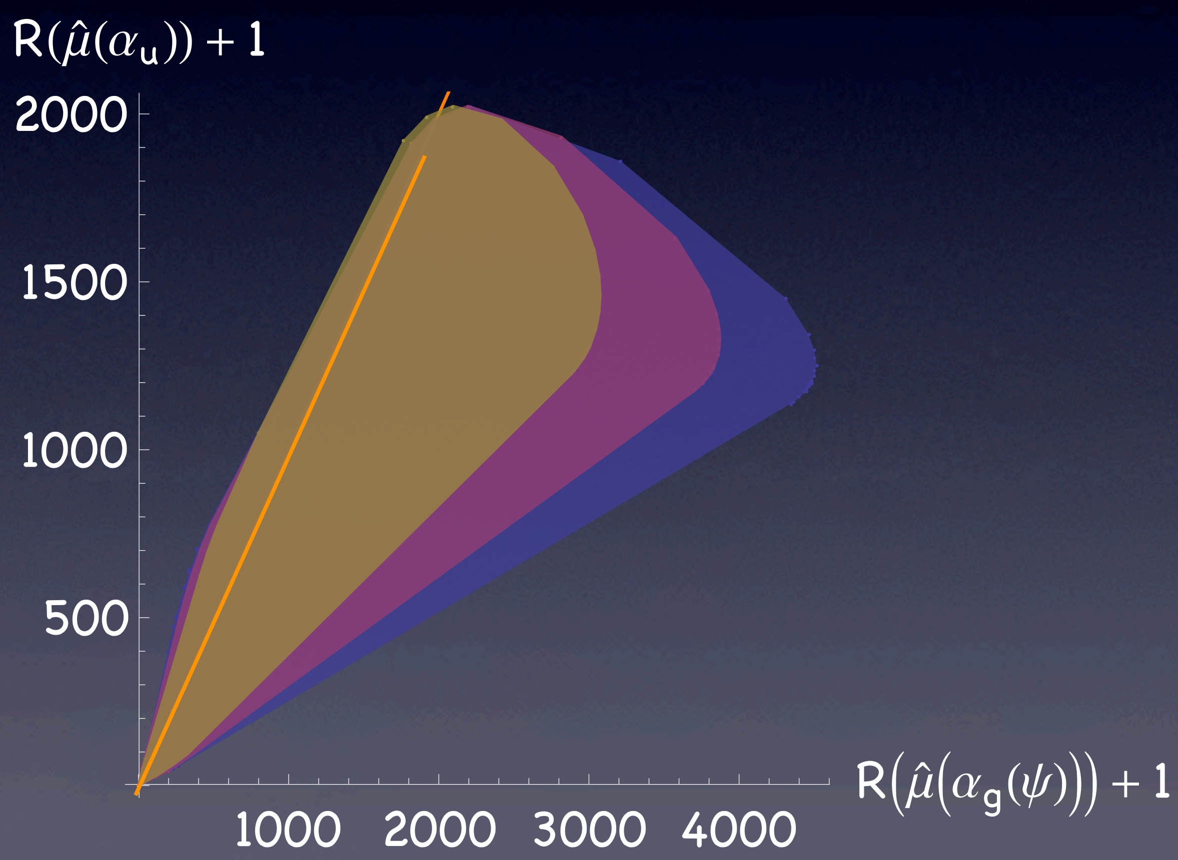
- Impact of higher wealth W_0 and payout ω
  - W_0 = ω
  - Small W_0,ω great for nearly black process
  - Less useful if much signal (crosses RI threshold)



Wharton
Department of Statistics

Universal vs Geometric

- Direct comparison favors universal
  - Only small geometric rates to be competitive
  - Geometric has higher worst case risk



Wharton
Department of Statistics

Wrap-Up

- Streaming selection with alpha investing
  - Fast variable selection with provable control
- Universal spending rule
  - Competitive with best geometric rules
  - Better overall with larger than expected W_0
- Feasible set
  - Computational exact risk inflation
- Conjectures
  - Approximate boundaries using 2-point models
  - Shape of the feasible set at origin: non-analytic?
  - Proofs of the universality of investing rule

Wharton
Department of Statistics

Thanks for coming!