Closing the gap between risk estimation and decision-making: Efficient management of trade-related invasive species risk

Michael Springborn
Department of Environmental Science and Policy
University of California, Davis

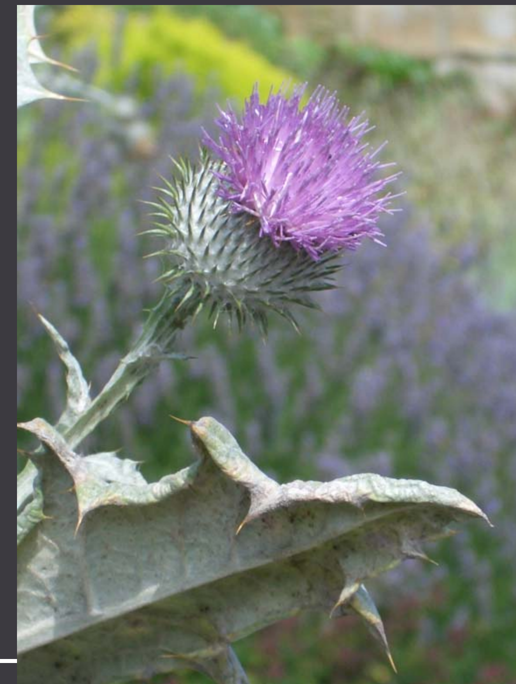
Robert Lieli
Department of Economics,
University of Texas, Austin

July, 2009

Motivation I: Methodological

- Typical approach in decision making under uncertainty (maximize payoffs/minimize losses):
  - Segregated two step process
    - classical model parameter estimation (probability of bad outcome)
    - decision making based on expected net benefits
  - Introduces an unnecessary intermediate objective (e.g. minimize sum of the squared residuals)

Motivation II: Damages from invasive alien species



Scolymus hispanicus
(Common Golden Thistle)

Onopordum acanthium
(Cotton Thistle)
detrimental to livestock foraging

Sagittaria montevidensis
(California Arrowhead)
Invades rice and irrigation supply

Sorghum halepense
(Johnson grass)
poisonous to equines (hydrogen cyanide)

Model

- Decision: accept or ban proposed biological imports
- Objective: maximize expected net benefits of imports
- Prediction: propensity of a proposed import (species) to naturalize and cause harm
- Predictive attributes:
    - » species history
    - » biogeography
    - » undesirable traits
- Information: training data set of existing naturalized species (attributes and invasive status).

Comparison of methods

1. (Australian) Weed Risk Assessment (WRA) model (Pheloung et al. 1999)
   - subjective, non-statistical
2. Maximum likelihood estimation
3. Bayesian decision-theoretic approach
4. Maximum utility estimation

Overview

- Add statistical foundations to the risk assessment and decision process
  - Current practice: subjective Weed Risk Assessment system
  - Maximum likelihood (ML) estimation (Caley et al. 2006; Hughes and Madden 2003)
  - Treatment of endogenously stratified sample
- Decision theoretic foundations: account for the real economic cost of decision making errors (vs. generic statistical error)
- First side by side comparison of:
  - classical, 2-step methods: (1) ML and (2) Bayesian
  - novel, integrated/single step method: (3) maximum utility (MU) estimation

Information and objective

$S_N=\{(X_1,Y_1),\ldots,(X_N,Y_N)\}$: Observed training data set

- X_n: predictive covariates/attributes for species *n*
- Y_n: observed invasive/non-invasive status for *n*

For a proposed species to import, *N*+1:

$$\max_{a_{N+1}} E[U(a_{N+1}, Y_{N+1}, X_{N+1} | X_{N+1})]$$

action outcome attributes

$$U(a,Y,X) = \begin{cases} u_{1,1}(X) & \text{if } a = 1 \text{ and } Y = 1 \\ u_{1,-1}(X) & \text{if } a = 1 \text{ and } Y = -1 \\ u_{-1,1}(X) & \text{if } a = -1 \text{ and } Y = 1 \\ u_{-1,-1}(X) & \text{if } a = -1 \text{ and } Y = -1 \end{cases}$$

ban (a = 1); accept (a = −1); invasive (Y = 1); non-invasive (Y = −1)

Decision structure

Objective:

$$\max_{a_{N+1}} E^Y\left[U(a_{N+1}, Y_{N+1}, X_{N+1} | X_{N+1})\right]$$

Uncertainty:

$$p(x;\theta) = P(Y_{N+1} = 1 \mid X_{N+1} = x)$$

e.g. logit/probit link function

Restate objective:

$$\max_a \{p(x;\theta)u_{a,1}(x) + [1 - p(x;\theta)]u_{a,-1}(x)\}$$

Optimal action:

Take action ban (a=1) iff:*

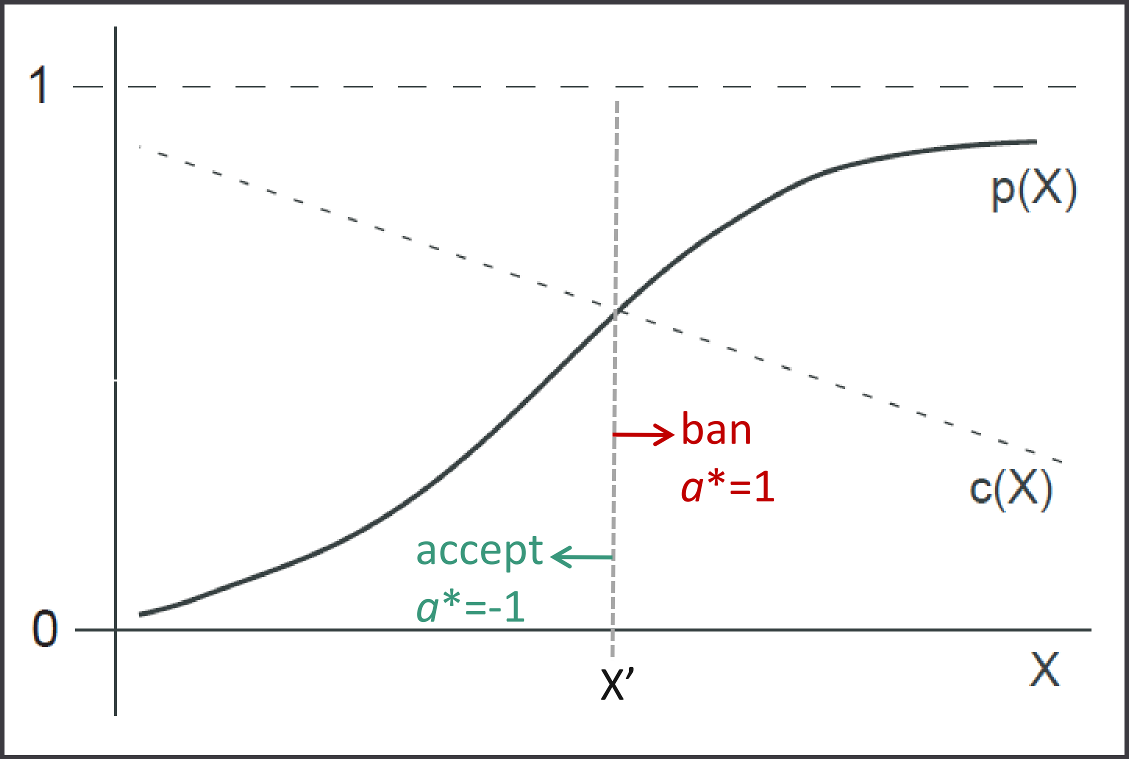
$$p(x;\theta) > \frac{u_{-1,-1}(x) - u_{1,-1}(x)}{[u_{-1,-1}(x) - u_{1,-1}(x)] + [u_{1,1}(x) - u_{-1,1}(x)]} \equiv c(x)$$

Decision structure

Optimal action:

Take action ban (a*=1) iff:

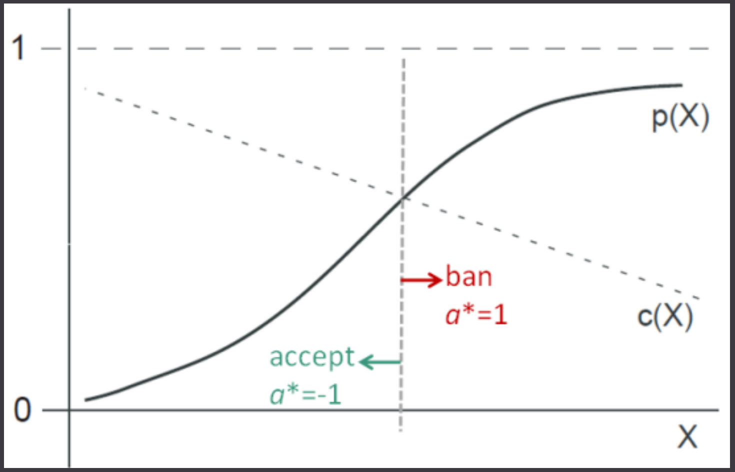
$$p(x;\theta) > \frac{u_{-1,-1}(x) - u_{1,-1}(x)}{[u_{-1,-1}(x) - u_{1,-1}(x)] + [u_{1,1}(x) - u_{-1,1}(x)]} \equiv c(x)$$



		ML	Bayes	MU
Steps	**1**	θ: maximize the likelihood function •most probable value (mode) from the likelihood function over θ. •“we only care about being exactly right; and, if we are wrong, we don't care how wrong we are" (Jaynes and Bretthorst 2003)	θ: Find $E[p(x;\theta)\mid\pi(\theta\mid S_N)]$, •$p(x;\theta)$ averaged across the entire posterior distribution •posterior can be similar to or exactly the likelihood function used in ML	θ: maximize sample average of utility function •*given* that a particular action is triggered by the sign of $p(x;\theta) - c(x)$ •θ: No assumed distribution for θ
	2	a: max EU given: $p(x;\theta^{ML})$	a: min PEL (max PEU) given: $E[p(x;\theta)\mid\pi(\theta\mid S_N)]$	max EU given: $a = \text{sign}[p(x;\theta) - c(x)]$
Emphasis on particular obs. in range of X		Observations given equal weight. Global fit.	Observations given equal weight. Global fit	Emphasis on observations that are important for identifying intersection [where $p(x;\theta) = c(x)$]

Maximum utility (MU) estimation

- Idea: estimation of the entire function *p*(*x*;θ) is a not necessary for optimal binary decision (Elliott and Lieli, 2007)
  - It is enough to know whether *p*(*x*;θ) is above or below the cutoff function, *c*(*x*)
  - Task: the *intersection* of *p*(*x*;θ) and *c*(*x*)



- Estimate: parametric models of *p*(*x*;θ) [e.g. logit link function]
  - Sample analog of objective function

- Result: the MU estimator
  - an extension of Manski's (1975, 1985) maximum score method
  - output best interpreted as a **decision rule** (technically, an estimate of the sign of [*p*(*x*;θ) - *c*(*x*)] rather than an estimate of *p*(*x*;θ) per se.

Maximum utility estimation

Social planner's objective function:

$$\max_a \{p(x;\theta)u_{a,1}(x) + [1 - p(x;\theta)]u_{a,-1}(x)\}$$

- **Recall: action *a*=1 (ban) is optimal iff: $p(x;\theta) - c(x) > 0$**
- **Replace *a* with: $sign[p(x;\theta) - c(x)]$ =**
  1, if $p(x;\theta) - c(x) > 0$
  -1, if $p(x;\theta) - c(x) < 0$

  Embed optimal decision *a** as a function of θ
- **Rearrange and simplify objective:**

$$\max_\theta E \quad \{b(X)[Y + 1 - 2c(X)]sign[p(X;\theta) - c(X)]\}$$

$$c(X) \equiv \frac{u_{-1,-1}(X) - u_{1,-1}(X)}{[u_{-1,-1}(X) - u_{1,-1}(X)] + [u_{1,1}(X) - u_{-1,1}(X)]} \equiv \frac{u_{-1,-1}(X) - u_{1,-1}(X)}{b(X)}$$

- Find estimates using the sample analog:

$$\hat{\theta}^{MU} = \arg\max_\theta N^{-1} \sum_{n=1}^{N} b(X_n)\,[Y_n + 1 - 2c(X_n)]\,sign\,[p(X_n;\theta) - c(X_n)]$$

Bayesian decision theoretic approach

- Difference from ML & MU:
  - does not involve an estimate of the coefficient vector θ
- Use the training sample to update beliefs about the true value of θ:

$$\underbrace{\pi(\theta|S_N)}_{\text{posterior (unnormalized)}} \propto \underbrace{\pi(\theta)}_{\text{prior}} \underbrace{\prod_{n=1}^{N} f(y_n|\theta, X_n)}_{\text{likelihood function}}$$

Bayesian decision theoretic approach

Specify a utility-based "loss function" (Berger, 1985):

$$\begin{aligned} L(a,\theta,x) &= -E_\theta[U(a,Y,x)] \\ &= -(1/2)(1+a)\left\{p(x,\theta)u_{1,1}(x) + [1-p(x,\theta)]u_{1,-1}(x)\right\} - \\ &\quad (1/2)(1-a)\left\{p(x,\theta)u_{-1,1}(x) + [1-p(x,\theta)]u_{-1,-1}(x)\right\}, \end{aligned}$$

Goal: identify the "Bayes action" *a** that minimizes the posterior expected loss:

$$PEL(a^*) = \min_a \int L(a,\theta,x)\pi(\theta|S_N)d\theta$$

$$\pi(\theta|S_N) \propto \pi(\theta)\prod_{n=1}^{N} f(y_n|\theta,X_n)$$

Posterior distribution for θ (unnormalized)

Yields a similar decision rule as before, *a**=1 iff:

$$\int p(x,\theta)\pi(\theta|S_N)d\theta - c(x) > 0$$

		ML	Bayes	MU
Steps	**1**	θ: maximize the sample average of the likelihood function •most probable value (mode) from the likelihood function over θ. •"we only care about being exactly right; and, if we are wrong, we don't care how wrong we are" (Jaynes and Bretthorst 2003)	θ: Find $E[p(x;\theta)\|\pi(\theta\|S_N)]$, •$p(x;\theta)$ averaged across the entire posterior distribution •posterior can be similar to or exactly the likelihood function used in ML	θ: maximize sample average of utility function •*given* that a particular action is triggered by the sign of $p(x;\theta) - c(x)$ •θ: No assumed distribution for θ
	2	a: max EU given: $p(x;\theta^{ML})$	a: min PEL (max PEU) given: $E[p(x;\theta)\|\pi(\theta\|S_N)]$	max EU given: $a = \text{sign}[p(x;\theta) - c(x)]$
Emphasis on particular obs. in range of X		Observations given equal weight. Global fit.	Observations given equal weight. Global fit	Emphasis on observations that are important for identifying intersection [where $p(x;\theta) = c(x)$]

Current practice

- Australian Weed Risk Assessment (WRA) model (Pheloung et al. 1999).
  - Make decisions on proposed imports based on inference from a previously assembled training data set
  - Makes extensive use of expert assessments
  - Ease of use
  - Transparent process
    - though not necessarily in value judgments of where to draw the cutoff
  - Not based on formal statistical or economic foundations (Caley et al. 2006)

Data

- $S_N=\{(X_1,Y_1),\ldots,(X_N,Y_N)\}$: training sample data
  - 370 plant species present in Australia (Pheloung et al. 1999)
    - 286 weedy (77%), 84 non-weedy ($Y_1,\ldots,Y_{370}$)
    - Drawn from all sectors (agriculture, environment, horticulture, garden and service areas)
  - Experts scored multiple covariates in three categories: biogeography, undesirable traits and biology/ecology ($X_1,\ldots,X_{370}$)




Weed Risk Assessment question sheet:

Weed Risk Assessment system question sheet: Answer yes (y) or no (n), or don't know (leave blank), unless otherwise indicated

Botanical name:	Outcome:
Common name:	Score:
Family name	Your name:

		History/Biogeography			
A	**1**	*Domestication/ cultivation*	1.01	Is the species highly domesticated? If answer is 'no' got to question 2.01	
C			1.02	Has the species become naturalised where grown?	
C			1.03	Does the species have weedy races?	
	2	*Climate and Distribution*	2.01	Species suited to Australian climates (0–low; 1–intermediate; 2–high)	2
			2.02	Quality of climate match data (0–low; 1 intermediate; 2–high)	2
C			2.03	Broad climate suitability (environmental versatility)	
C			2.04	Native or naturalised in regions with extended dry periods	
			2.05	Does the species have a history of repeated introductions outside its natural range?	
C	**3**	*Weed elsewhere*	3.01	Naturalised beyond native range	
E			3.02	Garden/amenity/disturbance weed	
A			3.03	Weed of agriculture/horticulture/forestry	
E			3.04	Environmental weed	
			3.05	Congeneric weed	
		Biology/Ecology			
A	**4**	*Undesirable traits*	4.01	Produces spines, thorns or burrs	
C			4.02	Allelopathic	
C			4.03	Parasitic	

Data: endogenously stratified sample

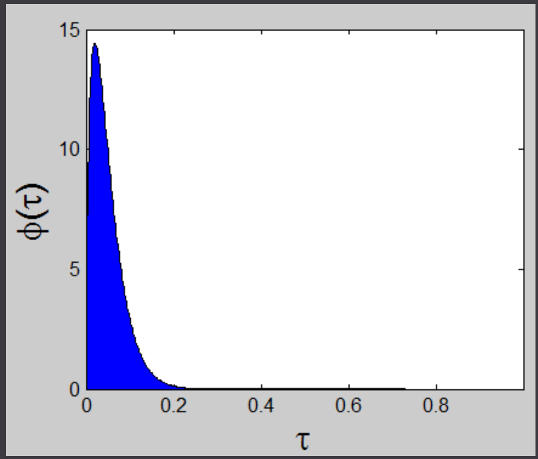
- Common with rare events data

- Rare event (Y=1, "invasive") is overrepresented in the sample relative to the population rate, τ:
  - τ=2%: most likely value of the population probability of plant weediness (Smith 1999)
  - Training data set: mean(Y)=77% invasive ("weedy") species

- Stratified sampling econometrics literature (frequentist framework):
  - **Manski and Lerman 1977 (WESML)**
  - **Cosslett 1993**; Imbens and Lancaster 1996;
  - King and Zeng 2001.

Economic and biological parameters

- τ: population proportion invasive
  - Review by Smith (1999) reports the range of assessments to be 0.01%-17%, with a **likely value of 2%**.
  - Caley et al. (2006}: uncertainty over the population proportion $\varphi(\tau)$ is characterized by a Beta distribution
    - Parameters selected such that:
      - mode equal to the most likely value (0.02)
      - 99% quantile = upper bound of assessed range (0.17)



- Cost and benefit parameters (Keller et al. 2007)
  - *D*, annual expected damage from a weed: $2,068K
  - *B*, annual expected benefit of an imported plant: $141K
  - Based on Australian data (Sinden et al. 2004; Virtue et al. 2004; Nursery and Garden Industry Australia 2004; Nursery Industry Association of Australia 1999).
    - Units: 2002 Australian dollars (AUD)

Welfare calculations

- What is the per species expected net benefit (*ENB*) of implementing one of the four screening methodologies? (ML, Bayes, MU, WRA)
  - Depends on the status quo:
    - open versus closed door

- Optimal decision for a given species doesn't depend on status quo but the *ENB* does.

Action-outcome utility matrix:

- Constant case, $c(x) = c$
- Starting point is **close door** policy

	$Y = 1$ (weed)	$Y = -1$ (non-weed)
Ban	0	0
Don't ban	$(B - D)/r = (141K - 2,068K)/0.03$	$B/r = 141K/0.03$

Results

	$Y = 1$ (weed)	$Y = -1$ (non-weed)
Ban	0	0
Don't ban	$(B - D)/r = (141K - 2,068K)/0.03$	$B/r = 141K/0.03$

Constant cutoff: $c = 0.0678$. Assumed base rate of weeds: $\tau = 0.02$.

Table 2: Sensitivity, specificity and expected benefits of three classification rules. Estimation sample size = 250; Evaluation sample size=120; Reps=1,000

Method	Pr(*a*=1|*Y*=1) Sensitivity	Pr(*a*=-1|*Y*=-1) Specificity	Expected net benefit (AUD, millions)	Relative net benefit (ML=100)
IN-SAMPLE				
ML (logit)	0.566	0.970	3.886	100.0
Bayes (logit)	0.549	0.976	3.932	101.2
MU (logit)	0.567	0.973	3.901	100.4
WRA (weed if $>$ 3)	0.862	0.728	3.157	81.2
OUT-OF-SAMPLE				
ML (logit)	0.567	0.970	3.885	100.0
MU (logit)	0.569	0.968	3.880	99.9
WRA (weed if $\geq$ 3)	0.856	0.720	3.111	80.1

D/B = 14.6. Would need to increase to 62.8 for *ENB* of WRA to surpass others.

Utility: Dependence on covariate.

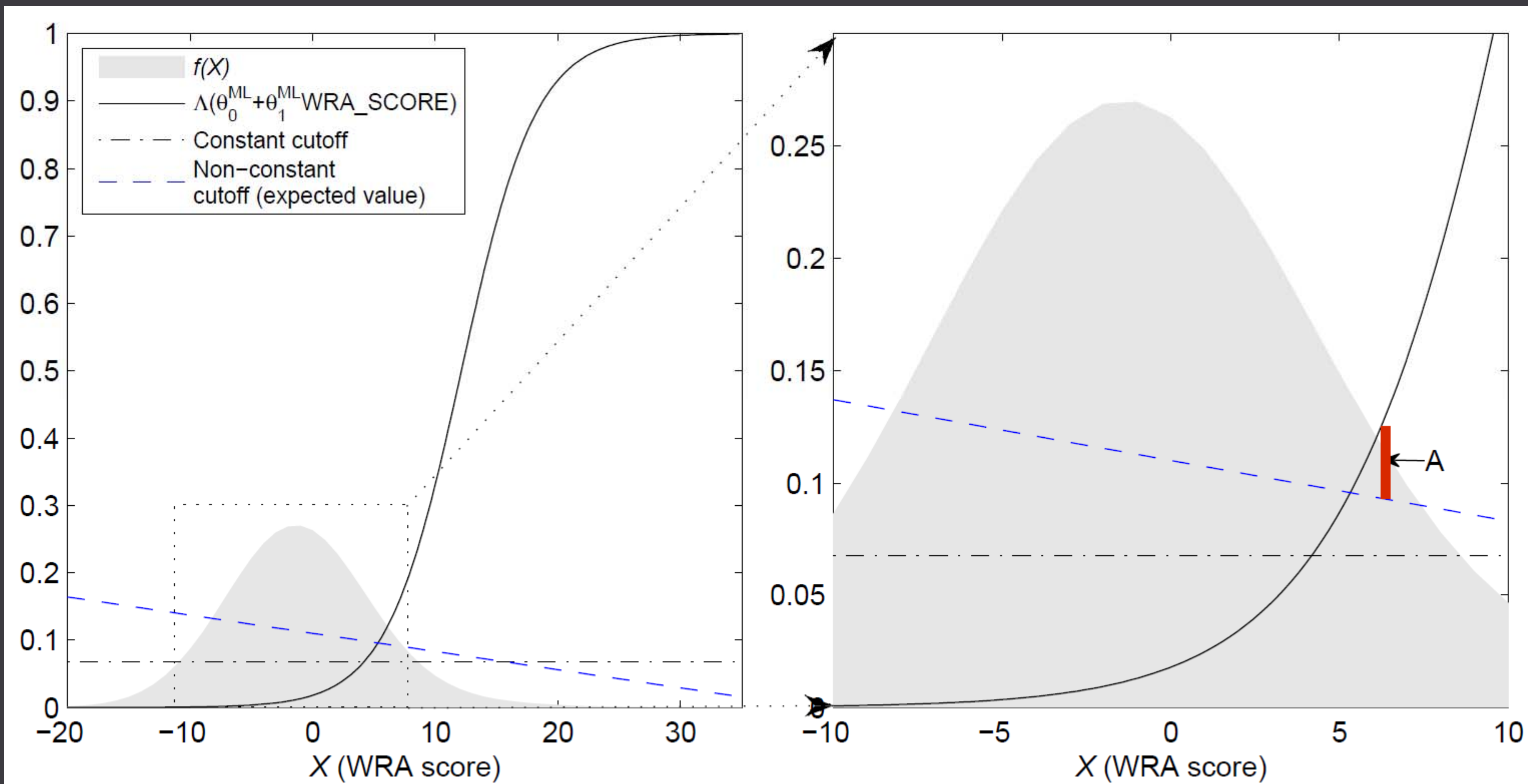
	$Y = 1$ (weed)	$Y = -1$ (non-weed)
Ban	0	0
Don't ban	$-(928K + 787K \cdot \texttt{Sc\_Undes})/r + 141K/r$	$141K/r$

Assumed base rate of weeds: $\tau = 0.06$.

Table 4: Estimation sample size= 250; Evaluation sample size=120; Reps=1,000

Method	Pr(a=1|Y=1) Sensitivity	Pr(a=-1|Y=-1) Specificity	Average expected net benefit (AUD, millions)		Relative net benefit (ML=100)
IN-SAMPLE					
ML (logit)	0.649	0.909	3.465		100.0
Bayes (logit)	0.640	0.917	3.505	$253K /species	101.2
MU (logit)	0.607	0.971	3.758		108.5
WRA (weed if ≥ 3)	0.860	0.726	2.898		83.6
OUT-OF-SAMPLE					
ML (logit)	0.651	0.903	3.434	$125K/species	100.0
MU (logit)	0.599	0.933	3.558		103.6
WRA (weed if ≥ 3)	0.860	0.727	2.903		84.5

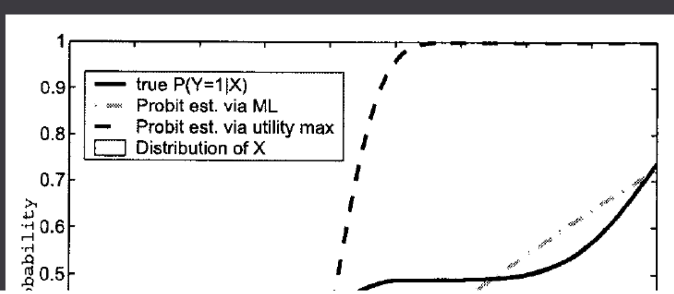
Graphical depiction of MU advantage



Results

- covariate-dependent payoffs can be an important driver of the improvements generated by the MU methodology
  - More likely the case when
    - *c*(*x*) is steeply and oppositely sloped to *p*(*x*;θ) [Although at some point this will reverse as *c*(*x*) approaches a vertical line at which point the policy is degenerate.]
    - Intersection is at a point in the range of X where such draws are common.

- MU can outperform other methods, though gain appears to decay if training sample is not representative


true P(Y=1|X)
Probit est. via ML
Probit est. via utility max
Distribution of X
1
0.9
0.8
0.7
0.6
0.5

Context

- intensional global movement of non-indigenous species
- screening international plant trade for invasive species risk
- regulators who must make decisions over what to allow and exclude based on incomplete information

Scolymus hispanicus
(Common Golden Thistle)

Elodea canadensis
(American Waterweed)

Mimosa pigra
(Giant Sensitive Tree)

Onopordum acanthium
(Cotton Thistle)

Sagittaria montevidensis
(California Arrowhead)

Sorghum halepense
(Johnson grass)

Model comparison

- ML -- Estimation and decision in 2 steps:
  1. θ:Maximize the sample average of the likelihood function
     - most probable value (mode) from the likelihood function over θ.
       – "we only care about being exactly right; and, if we are wrong, we don't care how wrong we are" (Jaynes and Bretthorst 2003)
     - X: emphasizes fit of model given observed values of X.
  2. Maximize utility conditional on $p(x;\theta^{ML})$

- Bayes -- Minimize posterior expected loss. Effectivley 2 steps:
  1. θ: Find $E[p(x;\theta)|\pi(\theta|S_N)]$, or $p(x;\theta)$ averaged across the entire posterior distribution (posterior can be similar to or exactly the likelihood function used in ML).
     - X: emphasizes posterior given observed values of X.
  2. Minimize loss (maximise utility) conditional on expected values given the posterior: $E[p(x;\theta)|\pi(\theta|S_N)]$.

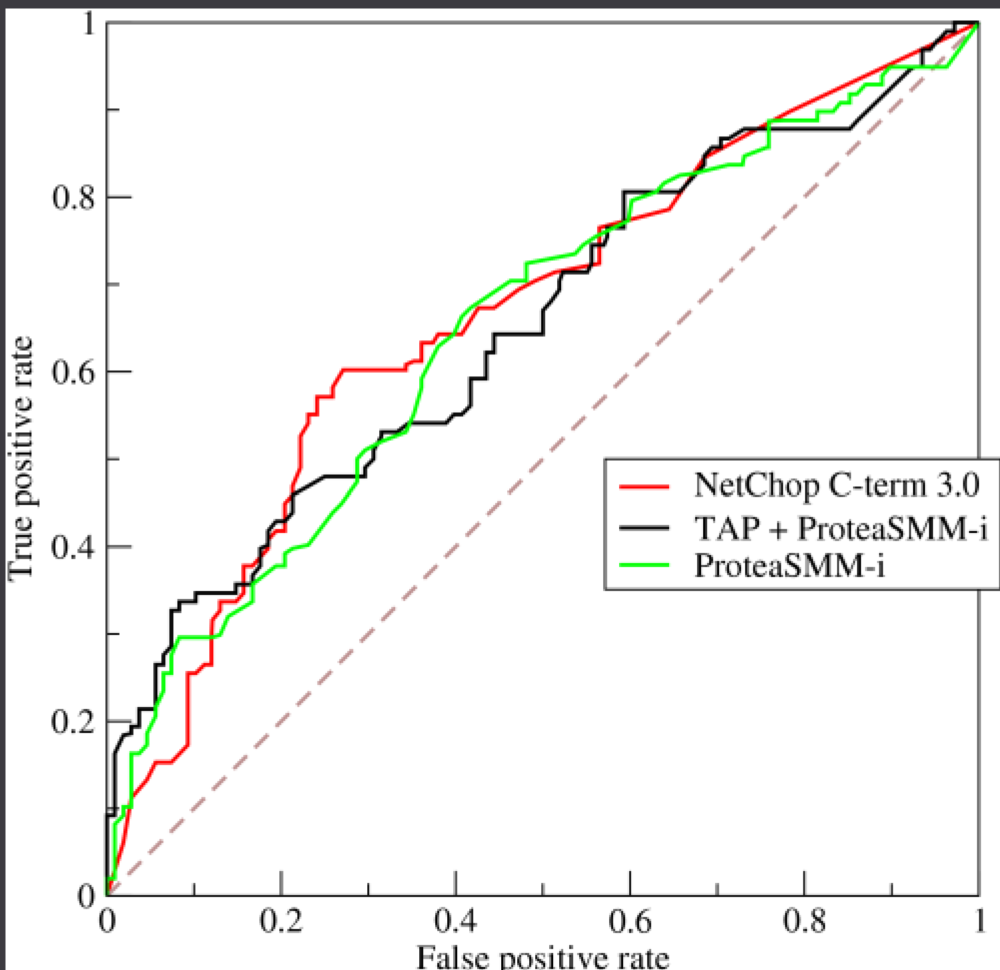
- MU – Estimation and decision in one step:
  – Select θ to to maximize utility given that a particular decision is triggered by the sign of $[p(x;\theta) - c(x)]$
  – θ: No assumed distribution for θ
  – X: Emphasis is placed not on the global fit of the model but rather on getting the intersection right, i.e. where $p(x;\theta) = c(x)$.

Comparison of alternatives

- Standard 2-step approaches: conditional probabilities of invasiveness are estimated in isolation before consequences of outcomes are considered in making the decision of whether to ban or allow a novel plant import.
  - maximum likelihood (ML) estimation
  - Bayesian estimation

- “Maximum utility" (MU) estimation
  - expected consequences have a direct inuence on parameter estimation itself (Elliott and Lieli 2007; Lieli and White, forthcoming).
  - Key idea: for prediction of a binary variable (invasive/non-invasive), a global fit of the model is less important than a localized fit which partitions the information (covariate) space in a way that minimizes the economic cost of classification (ban/accept) errors.
    - Errors: (1) ban a non-invasive species, (2) accept an invasive species

ROC curve



true positive (TP) eqv. with hit
true negative (TN) eqv. with correct rej
false positive (FP) eqv. with false alarm,
false negative (FN) eqv. with miss, Type
true positive rate (TPR) eqv. with hit rat
false positive rate (FPR) eqv. with false
accuracy (ACC) *ACC* = (*TP* + *TN*) / (*P* + *N*)
specificity (SPC) or True Negative Rate *S*
positive predictive value (PPV) eqv. wit
negative predictive value (NPV) *NPV* = *T*
false discovery rate (FDR) *FDR* = *FP* / (*F*