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

Robert P. Lieli† and Michael Springborn‡

March 1, 2010

Abstract

In this paper we assess multiple alternatives for designing a screening system to make decisions to allow or exclude novel imported goods which may be accompanied by an undesirable side effect, such as biological invasion. The statistical decision problem is to use available information on previous imports to parameterize a predictive model of the key unknown—a proposed good's latent status as damaging or benign. We develop the first side-by-side comparison of two classical approaches—maximum likelihood and Bayesian—against a third, recently developed "maximum utility" (MU) estimation methodology. We demonstrate the implications for expected payoffs (benefits and potential damages) of a risky import for the risk estimation problem. While Bayesian methods for incorporating actual expected costs of error into statistical estimation are known, for the class of discrete action/outcome problems of interest here, where the probability of a future state is the focal unknown, Bayesian estimates are independent of payoffs and do not take full advantage of the structure of the problem. In contrast, the MU approach utilizes the insight that a global fit of the statistical model is less important than the localized problem of identifying the best switching point from one discrete decision to another, e.g. from rejection to acceptance of a proposed import. We develop an empirical application using Australian data based on the problem of choosing to reject or accept novel plant imports when the primary unknown is whether or not the proposal will become a damaging invasive species. For each methodology, we show how to account for a non-random, endogenously stratified sample, a problem commonly encountered with rare events data. We demonstrate when the MU method is likely to offer significant incremental gains relative to the alternatives and estimate this annual value to be $34-$49 million (AU$).

Keywords: risk assessment, invasive species, maximum utility estimation, Bayesian decision theory

---

*Working paper – please do not cite without authors' permission. We thank seminar participants at the 2009 Western Economic Association International Annual Conference and the 2009 Agricultural & Applied Economics Association Annual Meeting for helpful comments and suggestions.

†Department of Economics, University of Texas, Austin.

‡Department of Environmental Science & Policy, University of California, Davis, 2104 Wickson Hall, Davis, CA 95616, mspringborn@ucdavis.edu

1 Introduction

New forms of economic activity are an important engine for growth. However, novel goods may also bring health or environmental risks that are hard to quantify. For example, new international trade can benefit both importers and exporters but also unintentionally lead to the transfer of invasive pests or pathogens. Managing such economic activity to balance the tradeoff between benefits and risks involves forming expectations about low probability, high-impact events and then deciding whether the benefits justify the risks. This is typically treated as a two-step process of classical model parameter estimation followed by decision making based on expected costs and benefits. While a standard goal in decision making under uncertainty is to maximize payoffs (minimize losses), a segregated two-step estimation and decision process introduces an unnecessary intermediate objective. If the estimation involves standard linear regression, for example, parameter estimates are chosen to minimize the sum of squared residuals. In many cases of environmental risk, the true cost of estimation error is unlikely to be symmetric. For example, it is common for losses from mistakenly classifying a product as safe (false negative) to far outweigh losses from mistakenly identifying it as unsafe (false positive).

In this paper we extend and compare methods for using predictive information to estimate risk and set a course of action. We focus on accounting for the real economic cost of errors in the context of screening international plant trade for invasive species risk. The intensional global movement of non-indigenous species presents a significant policy problem for countries seeking to maximize the net benefits of trade. While non-native plant species may generate crop production (Ewel et al. 1999) or ornamental value (Knowler and Barbier 2005) they can also pose a threat to biodiversity (Reichard and White 2001) and agriculture (Pimentel et al. 2005). A key uncertainty in the decision to allow or exclude a proposed import is the propensity of a species to establish and cause harm. Some ecologists have expressed doubts that attributes predictive of future invasive status can be identified (Williamson 1999; Enserink 1999). However, in a more recent review of the literature, Kolar and Lodge (2001) argue that substantial progress has been made in using quantitative statistics to predict which species are likely invaders, especially for the taxon of plants. Observable attributes encompassing species ecology, history and biogeography are believed to be partially predictive of invasive species risk (Pheloung et al. 1999; Kolar and Lodge 2001). The current leading approach for classifying potential plant imports according to their risk of invasiveness using predictive covariates is embodied in the Australian Weed Risk Assessment (WRA) model (Pheloung et al. 1999). This approach involves making decisions on proposed imports based on inference from a previously assembled training data set of species known to be either invasive or non-invasive in the given host habitat, along with values for the predictive covariates. While the WRA model makes extensive use of expert assessments and is appealing in its transparency and ease of use, it is not based on formal statistical or economic foundations (Caley et al. 2006).

We compare two classical methods, which approach estimation of the invasive species threat independently of the decision to be made, with a third,recently developed technique, which integrates this process into a single step. Under either a maximum likelihood (ML) estimation or Bayesian estimation approach, 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. Incorporating implications of actual losses into the estimation process has long been discussed in a Bayesian framework (e.g. Berger 1985). However, we show in Section 2.3 that for a standard discrete action/outcome problem, where the parameters of interest inform a probability model of outcomes, the Bayesian loss function approach only makes use of the posterior expected probability of an outcome. This statistic is wholly independent of the expect cost of prediction error. In contrast, in the "maximum utility" (MU) estimation approach, expected consequences have a direct influence on parameter estimation itself (Elliott and Lieli 2007; Lieli and White, forthcoming). The method exploits the idea that, for prediction of a discrete variable (e.g. 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 errors. The ML estimation approach has been the method of choice in previous efforts to add statistical foundations to the WRA model (Caley et al. 2006; Hughes and Madden 2003). In this paper we develop the first side by side examination of the MU, ML and Bayesian classification methods.

To assess relative economic performance, we develop an empirical application using data from the Australian WRA program (Pheloung et al. 1999). The data set is novel as it is not subject to a particular endogeneity problem from an action-outcome feedback which has plagued previous empirical examinations of the MU methodology. The application allows for the full exploitation of the MU framework, supporting an analysis of the model under a more realistic payoff structure where utility also varies with the covariates. In Section 4.3 we present results for the case in which a classical method is sufficient and an alternative case in which the MU approach provides significant gains. The findings suggest that the covariate-responsive utility framework is an important driver of the improvements generated by the MU methodology. Results show that adopting a statistically rigorous approach can generate a several hundred thousand dollar increase in expected net benefits per species assessed. Furthermore, the MU approach can offer additional, statistically significant, incremental gains.

The empirical application requires an extension to the MU model to address a sampling problem. Our training data set is a non-random, endogenously stratified sample, a common occurrence with rare events data. In order to ensure that such training data includes sufficient information on the rare event of interest (e.g. species that are invasive), this uncommon event is often over-represented in the sample relative to the population rate. For example, while a large majority (77%) of the observations in the training data set we use are weedy species, one assessment pegged the most likely value for the population probability of plant weediness at 2% (Smith 1999). Methods for addressing such a stratified sample in a frequentist framework are well explored (Manski and Lerman 1977; Cosslett 1993; Imbens and Lancaster 1996; King and Zeng 2001). We show how these methods may be logically extended for the MU and Bayesian approaches.

2 Methodology

2.1 General elements of the classification model

The essential information for our statistical decision problem includes, for each species, a vector of observed covariates X and a binary indicator Y of invasiveness. We set $Y = 1$ for "invasive" and $Y = -1$ for "not invasive." The objective is to determine whether the optimal action a is to "ban" ($a = 1$) or "accept" ($a = -1$) based on the covariates of the proposed import, without direct knowledge of whether it will be invasive (Y).

Utility for the four possibilities in the action-outcome space over an infinite horizon is given by:

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

We will present specific estimates for these utility measures in Section 3. In general we assume that utility from correctly matching action and outcome (e.g. "ban" when "invasive") is greater than from incorrectly matching; formally, $u_{1,1}(X) > u_{-1,1}(X)$ and $u_{-1,-1}(X) > u_{1,-1}(X)$ for any possible value of X. The flexibility of allowing damages or benefits to vary systematically with the covariates could convey a significant advantage to certain methodologies. For example, it may be the case, as we will argue in the empirical application, that some covariates predictive of invasiveness are also correlated with expected damages from accepting an invasive import. This relationship can be captured by a specification of utility which varies with X, as above.

The decision maker's goal is to find the decision rule $a^*(X) \in \{-1, 1\}$ that maps the observed covariates into the action space (accept/ban) in an optimal way. Optimality, as usual, is in the sense of maximizing expected utility:

$$\max_{a(\cdot)} E_{XY}[U(a(X), Y, X)], \tag{2}$$

where the subscript denotes expectation w.r.t. the joint distribution of X and Y and maximization is undertaken over all (measurable) decision rules. The solution to this problem can be constructed "pointwise", i.e. for any possible value x of X one sets $a^*(x)$ equal to the solution of the problem

$$\max_{a\in\{-1,1\}} E[U(a, Y, X \mid X = x)] = \max_{a\in\{-1,1\}} \{p(x)u_{a,1}(x) + [1 - p(x)]u_{a,-1}(x)\}, \tag{3}$$

where $p(x) = P(Y = 1 \mid X = x)$, the conditional probability of invasiveness given $X = x$. Comparing expected utility under the two possible actions, the optimal decision rule is to predict "invasive" and take the action "ban" if and only if:

$$p(x) > \frac{u_{-1,-1}(x) - u_{1,-1}(x)}{[u_{-1,-1}(x) - u_{1,-1}(x)] + [u_{1,1}(x) - u_{-1,1}(x)]} \equiv c(x), \tag{4}$$

or, more succinctly, $a^*(x) = sign[p(x) - c(x)]$, where $sign(z) = 1$ if $z > 0$ and $sign(z) = -1$ if $z \leq 0$.

We refer to $c(x)$ as the "cutoff function" and note that it is optimal to ban the proposed species if the probability of invasiveness is greater than the value of the cutoff function. The numerator, $u_{-1,-1}(x) - u_{1,-1}(x)$, which is also the first bracketed term in the denominator, is the gain from switching from an incorrect to correct decision when $Y = -1$. The second bracketed term in the denominator, $u_{1,1}(x) - u_{-1,1}(x)$, is the gain from moving to the correct action when $Y = 1$. The greater is the relative gain from switching to the correct action when $Y = -1$ (non-invasive), the greater $c(x)$ will be, expanding the covariate range over which $a = -1$ (accept) is optimal. See Figure 1 for illustration.

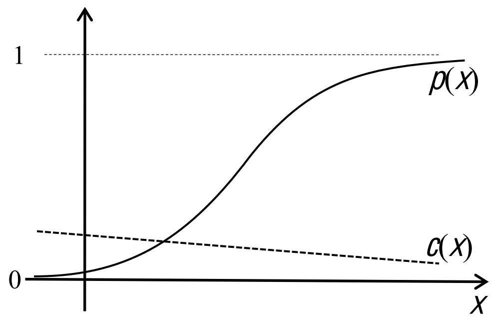

Figure 1: The basic decision/forecasting problem

The framework outlined above is essentially that of Elliott and Lieli (2007). A similar formulation has been used in previous empirical work (Boyes et al. 1989; Granger and Pesaran 2000; Pesaran and Skouras 2001), however, without considering the potential dependence of the utility function on covariates. A related decision-theoretic framework has been used to address the question of when to cease testing to learn about the likelihood of a hazardous outcome and settle on a particular binary response. For example, Olson (1990) considers the problem of when to stop testing to assess the likelihood of carcinogenic harm from a chemical and choose a particular regulatory action. While this search-theoretic literature is grounded in Bayesian learning in a controlled laboratory experiment setting, our problem requires econometric methods for estimating the decision rule $a^*(x)$ based on observed data. We now discuss various approaches.

2.2 Estimating the optimal decision rule: maximum likelihood vs. maximum utility

The "classical" approach to estimating the optimal decision rule starts by specifying a parametric model $p(x;\theta)$ for the conditional probability of invasiveness, $p(x)$, where $p(x;\theta)$ is a known function up to the finite dimensional parameter vector θ.[1] Given a training data set $S_N \equiv \{(X_1, Y_1), \ldots, (X_N, Y_N)\}$, θ can be estimated by maximum likelihood (ML). For example, if the training data is a random sample, then the ML estimator is given by

$$\hat{\theta}^{ML} = \arg\max_{\theta} \prod_{n=1}^{N} f(Y_n \mid X_n; \theta), \tag{5}$$

where $f(y \mid x;\theta)$ represents the Bernoulli density function $p(x;\theta)^{\frac{1+y}{2}}[1-p(x;\theta)]^{\frac{1-y}{2}}$. The optimal cutoff can then be applied to $p(x;\hat{\theta}^{ML})$.

If $p(x;\theta)$ is a correctly specified model of $p(x)$, then the ML estimator will recover $p(x)$ in the limit under general conditions, and, a fortiori, $a^*(x)$ is also consistently estimated. This is not generally the case under model misspecification, in which case the ML estimate of $p(x;\theta)$ provides a *global*

[1] In applications it is common to use parameterizations of the form $F(x'\theta)$, where F is a given cdf. The logit model corresponds to the choice of the logistic cdf, probit corresponds to standard normal, etc. On some of the more "exotic" choices for F see Koenker and Yoon (2007).

asymptotic approximation to $p(x)$.

A key insight in Elliott and Lieli (2007) is that estimation of the entire function $p(x)$ is not necessary for optimal decision making. For a given value of x, it is enough to know whether $p(x)$ is above or below the cutoff function $c(x)$—it does not matter by how much. What needs to be estimated with precision is the intersection of $p(x)$ and $c(x)$, and this is often possible even if the model $p(x;\theta)$ is not fully correctly specified. Elliott and Lieli (2007) accomplish this by using the sample analog form of the decision maker's expected utility maximization problem in estimating θ. The result is the maximum utility (MU) estimator, an extension of Manski's (1975, 1985) maximum score method. The output of the procedure is best interpreted as a decision rule [an estimate of the sign of $p(x) - c(x)$] rather than an estimate of $p(x)$ per se. If misspecification of $p(x;\theta)$ is so severe that even $a^*(x)$ cannot be recovered, then MU still delivers, asymptotically, the best decision rule given the model specification. ML, of course, does not have this property.

The MU estimator is set up as follows. Given a model $p(x;\theta)$, one approximates $a^*(x)$ with a decision rule of the form $sign[p(x;\theta) - c(x)]$. Substituting for $a(X)$ in problem (2) and rewriting yields the equivalent form

$$\max_{\theta} E_{XY} \left\{ b(X)[Y + 1 - 2c(X)] sign[p(X;\theta) - c(X)] \right\}, \tag{6}$$

where $b(x)$ is defined as the denominator of the cutoff function: $b(x) \equiv [u_{-1,-1}(x) - u_{1,-1}(x)] + [u_{1,1}(x) - u_{-1,1}(x)]$. Next, one selects $\hat{\theta}^{MU}$ such that

$$\hat{\theta}^{MU} = \arg\max_{\theta} N^{-1} \sum_{n=1}^{N} b(X_n) \left[Y_n + 1 - 2c(X_n) \right] sign \left[p(X_n;\theta) - c(X_n) \right]. \tag{7}$$

Because the objective function is a step function in θ and will generally feature multiple local maxima, typical optimization routines involving the gradient vector are not suitable. The preferred alternative is the simulated annealing algorithm which has been shown to perform well over multi-modal functions with flat ranges (Corana et al. 1987; Goffe et al. 1994).

2.3 The Bayesian decision-theoretic approach

We now develop a fully Bayesian classification method in which the decision theoretic foundations and the approach to estimating $p(x;\theta)$ are both Bayesian.[2] This approach is different than the ML and MU methodologies in that it does not involve an optimal selection of the coefficient vector θ (e.g. to maximize likelihood or utility). Instead the training sample S_N is used to update prior beliefs about the true value of θ, leading to a posterior distribution. Integrating over the posterior, we find the expected probability that a proposed species will be invasive. However, once this estimate of $p(x)$ is determined, the optimal decision rule takes the same general form as derived in equation (4).

A Bayesian decision-theoretic approach involves choosing a Bayes action, $\tilde{a}$, which minimizes expected loss (maximizes expected utility) given the information contained in the training sample S_N and the covariate X. Any pre-existing information about the vector θ is captured by the prior distribution $\pi(\theta)$. When working with the full data set we assume a noninformative uniform prior over

[2] We thank Graham Elliott for his input into this section.

θ since we are working with a sufficient number of observations "to let the data speak for themselves" (Gelman et al. 2004, p. 61).

Incorporating the information in a random training data sample S_N, the posterior distribution of beliefs over θ is given by:

$$\pi(\theta|S_N) = \frac{\pi(\theta)\prod_{n=1}^{N} f(Y_n \mid X_n;\theta)}{\int \pi(\theta)\prod_{n=1}^{N} f(Y_n \mid X_n;\theta)d\theta}, \tag{8}$$

where $f(y \mid x;\theta)$ again represents the Bernoulli density. Given also the covariates of a proposed import, the posterior expected probability that the species is invasive is given by

$$\tilde{p}(x) = \int p(x;\theta)\pi(\theta|S_N)d\theta. \tag{9}$$

An essential element to the Bayesian theoretic approach is the specification of a "loss function," $L(a,\theta)$, which identifies the level of loss if action a is taken and the true state of nature is θ. Berger (1985, p. 60) argues that while loss functions should ideally be developed from a utility framework, often certain "standard" losses are assumed, such as mean square loss or mean absolute loss, without reference to underlying consequences or utility functions (Parmigiani 2002, p. 84). We construct a utility-based loss function given by

$$\begin{aligned} L(a,\theta,x) &= -E^{\theta}[U(a,Y,x) \mid X = x] \\ &= -(1/2)(1+a)\left\{p(x,\theta)u_{1,1}(x) + [1-p(x,\theta)]u_{1,-1}(x)\right\} - \\ &\quad (1/2)(1-a)\left\{p(x,\theta)u_{-1,1}(x) + [1-p(x,\theta)]u_{-1,-1}(x)\right\}, \end{aligned} \tag{10}$$

where the expectation is taken with respect to $p(x;\theta)$, the parameterized conditional distribution of Y given $X = x$, and the superscript on the expectations operator indicates conditionality on a fixed level of θ. Our goal is to identify the Bayes action $\tilde{a}$, defined as the argument which minimizes the posterior expected loss:

$$\min_a \int L(a,\theta,x)\pi(\theta|S_N)d\theta. \tag{11}$$

Note that integrating the loss function (10) over the posterior (8) returns the loss function with $p(x;\theta)$ replaced by $\tilde{p}(x)$, the posterior expected probability of invasiveness. This estimate is independent of any expected damages or benefits. Finally, minimizing the posterior expected loss, we find an action rule of the form presented in equation (4); in particular, it is optimal to predict "invasive" ($Y = 1$) and take the action "ban" ($a = 1$) if and only if

$$\tilde{p}(x) > c(x), \tag{12}$$

or, more succinctly, $\tilde{a}(X) = sign[\tilde{p}(X) - c(X)]$.

2.4 Model comparison

While all three models involve optimization, it is instructive to consider in turn what the explicit or implicit objectives are. The ML estimator maximizes the (log-)likelihood function and, in doing so, selects parameter values that make what has been observed (e.g. in the training sample) more likely to have occurred than under any other parameter value. A non-frequentist description favored by ML's early users (e.g. Pierre-Simon Laplace and Carl Friedrich Gauss) was simply that the ML estimate was the "most probable value" (Jaynes and Bretthorst 2003, p. 175). Focusing on the mode of the likelihood function, the ML approach implies that "we only care about being exactly right; and, if we are wrong, we don't care how wrong we are" (Jaynes and Bretthorst 2003, p. 414). The likelihood function, maximized in equation (5), plays a central role in determining Bayesian posterior beliefs, equation (8), particularly given a diffuse prior and large sample size. But whereas the estimate of the conditional probability $p(x)$ under ML, $p(x; \hat{\theta}^{ML})$, is the conditional probability evaluated at the most probable level of θ, the Bayesian estimate, $\tilde{p}(x)$, is the value of $p(x; \theta)$ averaged across the range of beliefs over θ. In either case, the estimate of $p(x; \theta)$ is determined in isolation of the consequences or utility, which is maximized in a second stage, taking the output of the first stage as given.

In contrast, the MU approach maximizes the value of classification performance directly, balancing false positives and false negatives in a manner sensitive to the economic consequences. It takes advantage of the insight that all that matters in a utilitarian sense is making the best binary decision, i.e. estimation of where $p(x; \theta)$ intersects $c(x)$. One implication is that the MU approach is likely to be less sensitive to misspecification of $p(x; \theta)$ since getting the shape exactly correct is not generally important for estimating the optimal action. This framework also highlights the potential importance of exploring a cutoff function, $c(x)$, which truly varies over x, as this serves to shift the observations which should be emphasized. By construction, the MU method places extra weight on those parts of the covariate space over which correct classification produces the largest net economic benefit.

3 Adjustments under endogenously stratified sampling

Our training sample data, originally analyzed by Pheloung et al. (1999)[3], includes 286 species classified as weeds and 84 species classified as non-weeds. A particular challenge presented by this data set is that the sampling process by which the data were obtained cannot be regarded as random. The data set is best described as a "response based" sample, a type of endogenous stratification in which the sampling process is conditioned on the outcomes $Y = 1$ and $Y = -1$, and then a random sample is drawn separately from the two strata, rather than the joint distribution of (Y, X). As a result, the sample proportion of weeds is not a consistent estimate of $\tau \equiv P(Y = 1)$, the unconditional population proportion, or base rate, of weeds. The proportion of weeds in the data set (77%) is much greater than expert assessments of what the population proportion might be. Reviewing the appropriate literature, Smith (1999) reports the range of assessments to be 0.01%–17%, with a likely value of 2%. Methods exist for addressing this type of stratified sample, though many require information or assumptions about the base rate τ.

[3]We are grateful to the authors for generously sharing their data.

3.1 Correction to ML and MU

Caley et al. (2006) correct for sample imbalance using a bootstrap approach in which the weed observations are under-sampled to achieve a given target ratio for τ. This approach, while intuitive, involves setting aside some potentially useful information (i.e. excluded weed observations) in each bootstrap sample. An alternative procedure involves reweighting the objective function used in estimation to adjust for a given value of τ. To understand how this method works, consider the population objective function $Q(\theta) = E[q(X, Y; \theta)]$, where the expectation is with respect to the joint distribution of X and Y. If a random sample of size N from this distribution is available, maximizing the sample objective $\hat{Q}(\theta) = N^{-1} \sum_{n=1}^{N} q(X_n, Y_n; \theta)$ typically provides a consistent estimator of the maximizer of $Q(\theta)$. If, instead, random samples from the strata $Y = 1$ and $Y = -1$ are available, then the decomposition

$$Q(\theta) = E[q(X, Y; \theta) \mid Y = 1]\tau + E[q(X, Y; \theta) \mid Y = -1](1 - \tau)$$

suggests that consistent estimation can be based on the modified sample objective

$$\hat{Q}_\tau(\theta) = \frac{\tau}{N\bar{Y}_N} \sum_{n:Y_n=1} q(X_n, Y_n; \theta) + \frac{1-\tau}{N(1-\bar{Y}_N)} \sum_{n:Y_n=-1} q(X_n, Y_n; \theta),$$

where $\bar{Y}_N = N^{-1} \sum_{n=1}^{N} 1_{\{Y_n=1\}}$ indicates the sample proportion of observations with $Y = 1$. Thus, correction for response-based sampling can be achieved simply by multiplying each term in $\hat{Q}(\theta)$ by a weight, w_n, where $w_n = \tau/\bar{Y}_N$ if $Y_n = 1$ and $w_n = (1 - \tau)/(1 - \bar{Y}_N)$ if $Y_n = -1$.

In the context of estimating logistic regressions, this type of correction to the likelihood function was first proposed by Manski and Lerman (1977), resulting in what they called the weighted exogenous sampling (WES) ML estimator. Correction of the MU estimator for a response-based sampling process has not previously been addressed but is straightforward to implement based on the scheme outlined above.

While the reweighting described above restores consistency under response-based sampling, it relies on knowledge of τ and does not result in an asymptotically efficient estimator of θ (at least in the context of ML estimation). Alternative estimators under general stratified sampling schemes that overcome these problems have been proposed by Cosslett (1993), Imbens and Lancaster (1996) and King and Zeng (2001). Given certain conditions on $p(x; \theta)$, these methods allow efficient estimation of θ and τ simultaneously from a response-based sample without needing to parameterize the distribution of X. Unfortunately, if $p(x; \theta)$ is a logit model with a constant term, then simultaneous identification of the constant and τ is impossible. While consistent estimation of τ is, in principle, possible under alternative functional form assumptions, we found via numerical simulations that the likelihood surface in our problem tends to be very flat and identification is tenuous at best. Therefore, we do not pursue these methods to estimate τ. Even with τ given, we opt for the Manski-Lerman type correction, as it is extremely simple to implement, does not depend on the model specification used, and is directly applicable to the MU estimator as well.

In the next section we address correcting for the stratified sample in the Bayesian model which involves several more steps than in the case of MU or ML.

3.2 Endogenous stratified sampling in the Bayesian model

Statistics of interest in the decision framework for the Bayesian model ultimately depend on the posterior distribution for the parameter vector, specified in the random sample case in equation (8). In practice, since the denominator of (8) is a constant, we need only find the numerator, the unnormalized posterior density, a product of the prior $\pi(\theta)$ and the likelihood function of the sample. Typically, when specifying the likelihood of the sample, the joint likelihood of X and Y is eschewed in favor of the conditional likelihood (where X is treated as given) since X contains no information about θ. This is no longer the case under endogenous stratified sampling, necessitating the consideration of the distribution of X.

Let $g(x, y)$ represent the joint *sampling* distribution of Y and X and $g(x; y)$ the sampling distribution of X given Y. The marginal sampling probability for the stratum $Y = y$ is denoted H_y. The natural estimator of H_1 is of course the sample proportion $\bar{Y}_N$ of weeds (in fact, it is the ML estimator). *Population* distributions will be denoted by $f(\cdot)$. In particular, let $f(x; \lambda)$ represent the population distribution of X given a parameter vector λ. Following Cosslett (1993), we specify the likelihood function under stratified sampling where τ is treated as given:

$$
\begin{aligned}
L(\theta, \lambda) &= \prod_{i=1}^{N} g(Y_i, X_i) \\
&= \prod_{i=1}^{N} \left(\frac{H_1}{\tau}\right)^{\frac{Y_i+1}{2}} \left(\frac{1-H_1}{1-\tau}\right)^{\frac{1-Y_i}{2}} f(Y_i; X_i, \theta) f(X_i; \lambda),
\end{aligned}
\tag{13}
$$

where the second line follows from the fact that $g(x; y) = f(x; y)$ and Bayes rule.[4] Both τ and the density of the covariate sample $f(X_i; \lambda)$—ignored when the sample is random—are incorporated in the likelihood function above. For a given value of τ, the values of θ and λ are related to each other through the following constraint:

$$
\tau = \int_{-\infty}^{\infty} p(x; \theta) f(x; \lambda) dx. \tag{14}
$$

The framework above necessitates specifying the functional form of $f(x; \lambda)$. We assume that the covariate follows a normal distribution, conditional on Y: $X_i|Y_i \sim N(\mu_{Y_i}, \sigma^2_{Y_i})$. Supporting this assumption, we observe that the empirical conditional distributions of $X|Y = 0$ and $X|Y = 1$ are both symmetric and conform closely to a straight line on a normal probability plot. The unconditional population density of X is then given by a mixture model:

$$
f(x; \lambda, \tau) = f(x; \mu, \sigma, \tau) = (1 - \tau) f(x; \mu_0, \sigma_0) + \tau f(x; \mu_1, \sigma_1), \tag{15}
$$

where $\mu = [\mu_0, \mu_1]$, $\sigma = [\sigma_0, \sigma_1]$ and $\lambda = [\mu, sigma]$.

The final component of the Bayesian involves encoding any available pre-existing information into the priors over parameters for the Bernoulli probability model, $\pi(\theta)$, and for the covariates, $\pi(\lambda)$. The

[4]Specifically, $g(y, x) = H_y g(x; y) = H_y f(x; y) = [H_y/f(y)] f(y; x, \theta) f(x; \lambda)$.

unnormalized posterior distribution is given by:

$$\pi(\theta, \lambda; S_N) \propto \pi(\theta)\pi(\lambda)L(\theta, \lambda), \tag{16}$$

subject to the population proportion constraint in equation (14). To clarify this constraint, equation (14) is re-expressed here, solving explicitly for τ:

$$\tau = \frac{\int_{-\infty}^{\infty} p(x;\theta) f(x;\mu_0,\sigma_0)dx}{1 - \int_{-\infty}^{\infty} p(x;\theta)[f(x;\mu_1,\sigma_1) - f(x;\mu_0,\sigma_0)]dx}. \tag{17}$$

To implement the Bayesian estimation for the results presented in the next section we set $\pi(\theta)$ to be a constant since the data set is large enough for the observations to dominate the posterior, obviating the need for an informative prior. We explored several models of increasing complexity for the prior on the vector λ, starting from the simplest case in which the mean and variance parameters are set to their maximum likelihood estimates with the variance parameters assumed equal ($\sigma_0 = \sigma_1$) through a highly flexible structure where all four parameters are modeled as random. Since the simplest structure for the normal mixture model consistently performed best in terms expected welfare, the results we present in the empirical section are for this case of a degenerate prior over λ. We estimate the posterior distribution for the parameters of interest in the vector θ using Markov chain Monte Carlo (MCMC) techniques where the vector θ is sampled using the Metropolis-Hastings algorithm for the posterior in (16). We run 11 chains with starting points dispersed throughout the target distribution and then assess convergence using the standard metric proposed by Gelman and Rubin (1992). Methods for implementing the flexible random model for the normal mixture model using a Gibbs/Metropolis-Hastings sampling approach are described in the appendix.

4 Empirical application

4.1 Economic parameters

For an assessment of the costs and benefits of accepting exotic plant imports we follow Keller et al. (2007) who assemble estimates of the value and potential damages of ornamental imports. All monetary figures are in 2002 Australian dollars (AU$). Keller et al. (2007) argue that the best available estimate of economic losses from invasive plants in Australia is from Sinden et al. (2004) who assess the costs of control and output loss in agriculture and of control in natural environments. The estimate is incomplete but the direction of the bias is uncertain. Non-market impacts are omitted, biasing the measure down. However, the increase in control costs from invasives is likely overstated given the likely need for domestic weed control in agriculture in the absence of exotic weeds. Following Keller et al. (2007) we take Sinden et al.'s mean estimate of total economic loss ($4,039K) and, using estimates from Virtue et al. (2004), weight the estimate by the percentage of invasive plants thought to be attributable to ornamental plant trade (70%). Dividing the result by the number of invasive plants from this sector (1,366) returns an annual expected weed damage estimate D of $2,068K.

To estimate benefits of imports we would ideally use a measure of combined consumer and producer surplus. However, Keller et al. (2007) state that the information necessary to formulate an "accurate...value per species do(es) not exist" (p. 206). A coarse estimate, again consistent with Keller et al. (2007), is constructed as follows. The 2003-2004 fiscal year total value of the ornamental plant sector ($5.55 billion, Nursery and Garden Industry Australia 2004) is deflated by the percentage of plant sales attributed to exotic plants (64.6%, Nursery Industry Association of Australia 1999) and divided by the total number of introduced species (25,360, Virtue et al. 2004). The annual expected benefit of an imported plant B is therefore set to $141K. Upward bias in this measure stems from the use of plant sector revenue instead of surplus. Downward bias is introduced by omitting consumer surplus and assuming that all 25,360 species that have been introduced were still being sold in 2003-2004.

4.2 Predictive covariates

The weed risk assessment (WRA) score is an aggregate numerical measure based on responses to 49 questions regarding attributes of a plant that are correlated with weediness. The methodology and rationale are described in detail by Pheloung (1995). The questions are grouped into three main categories. Biogeographic attributes include the observed distribution, climate preferences and existing global weediness history of a plant. The second section covers undesirable traits such as whether the plant is noxious or parasitic. The final category of biology/ecology encodes the perceived potential of the species to "reproduce, spread and persist" (Pheloung 1995, p. 11). The training sample data includes 370 non-native plant species present in Australia drawn from all sectors, including the undeveloped environment, agriculture, horticulture and garden and service areas. Multiple plant scientists evaluated each species in the set, including whether the exotic plant is considered a weed within Australia.

4.3 The evaluation exercise and numerical results

To illustrate how the relative performance of different classification rules can vary depending on the modeling of costs and the correction for the stratified sample, we present four cases. In the first two cases costs are modeled as independent of the covariates of the species under consideration, and the assumed base rate is chosen to be 5% and 2%, respectively. The former value is the approximate mean of the Caley et al. (2006) prior, the latter is the mode. In the second pair of cases we work under the more realistic assumption that the decision maker's payoff matrix (and hence the optimal cutoff) depends on the charactersitics of the given species, and we condition on the same two values for the base rate.

We assume that the existing policy (status quo) is a "closed door", that is, without an assessment novel plant species are not accepted for importation. This has no effect on the optimal decision under any of the methodologies; it serves only to establish a point of reference for welfare calculations. The decision maker is assumed to be risk neutral—the utility of each possible outcome is given by the resulting monetary payoff.

When the benefits from importing a useful species, as well as the cost of importing a harmful one, are assumed to be constant, we use the cost-benefit estimates described in Section 4.1. The implied payoff matrix is presented in Table 1. Using the optimal decision rule in equation (4), the decision maker bans a species for import if the probability of a species being invasive, conditional on the information at hand, is greater than the constant $c = 0.0682$. As indicated above, the two subcases

Table 1: Infinite horizon costs and benefits of weed classification, assumed constant

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

Note: Source: Benefits and damages as reported by Keller et al. (2007), discussed in Section 3. Assumed discount rate: $r = 0.03$. Figures are in AU$.

Table 2: Infinite horizon costs and benefits of weed classification, given dependence on X

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

Note: Assumed discount rate: $r = 0.03$. Figures are in AU$.

within the "constant cutoff" case correspond to assuming $\tau = 0.05$ and $\tau = 0.02$, respectively.

Conditional on a plant import being weedy, it is however quite reasonable to expect that the damage it generates is correlated with its WRA score (or at least some of its components). For example, the sub-score for undesirable traits captures several species characteristics with obvious implications for damage, including whether a species is believed to be parasitic, toxic to humans or animals, or creates a fire hazard (Pheloung et al. 1999). For purposes of illustration we assume that the dependence of utility on covariates is as given in Table 2. Here it is assumed that if a species has an undesirable traits score (`Sc_Undes`) of -1 (the minimum possible), then the cost of it becoming a weed, minus benefits, is zero. Moreover, the average damage of the (don't ban, weed) option over all weeds in the sample is calibrated to be precisely $2,068K, chosen to recover the constant damage level specified in Table 1. Using equation (4), under this cost-benefit specification the theoretically optimal decision rule is to ban imports if the probability of a species being invasive, conditional on the information at hand, is greater than the optimal cutoff function

$$c(\texttt{Sc\_Undes}) = 0.14148/(0.9283 + 0.7868 \cdot \texttt{Sc\_Undes}).$$

The two subcases within the "variable cutoff" case correspond to conditioning on the two possible values of τ given above.

In each of the four cases we compare six classification rules:

1-2. (ML) Reject the species if and only if

$$F(\hat{\theta}_0^{ML} + \hat{\theta}_1^{ML}\texttt{WRA\_SCORE}) > \text{optimal cutoff}, \tag{18}$$

where the link function F is a c.d.f. chosen by the researcher and $\hat{\theta}^{ML}$ is the maximum likelihood estimate of θ under the distributional assumption made. The log-likelihood function of the observations is reweighted to correct for the stratified sample. While it is common to employ the logistic c.d.f. in modeling conditional probabilities, this choice is often dictated by convention and is rarely subjected to testing. In addition to the logistic c.d.f. (giving rise to the logit model), we also use c.d.f. of the Cauchy distribution as a link function (the "cauchit" model). While the latter choice might appear somewhat exotic, there is no a priori reason to rule it out, and it will be shown that the classification results are actually sensitive to the specification of F. Clearly, this method takes the decision maker's preferences into account in choosing the optimal cutoff, but not in estimation.

3-4. (Bayes) Reject the species if and only if

$$E_{\pi(\theta|S_N)}F(\theta_0 + \theta_1 \texttt{WRA\_SCORE}) > \text{optimal cutoff}, \tag{19}$$

where the expectation on the LHS is w.r.t. the posterior density for θ. Again, the two choices of F examined are logit and cauchit. The estimated posterior for θ is corrected for the stratified sample. Again, the decision maker's preferences are taken into account in choosing the optimal cutoff, but not in estimation.

5. (MU) Reject the species if and only if

$$F(\hat{\theta}_0^{MU} + \hat{\theta}_1^{MU} \texttt{WRA\_SCORE}) > \text{optimal cutoff}, \tag{20}$$

where $\hat{\theta}^{MU}$ is the maximum utility estimate of θ. In estimating the model the decision maker's sample objective is reweighted to correct for the stratified sample. As seen in equation (7), the MU objective function ignores entirely the choice of F as long as it is strictly increasing. Therefore, in obtaining the numerical results, F is simply set equal to the identity function. Here the decision maker's preferences are taken into account in estimation as well as in choosing the cutoff.

6. (WRA) As a benchmark, we also evaluate the decisions returned by the WRA system. This means reject if $\texttt{WRA\_SCORE} > 5$; accept if $\texttt{WRA\_SCORE} < 1$ and evaluate further if $1 \leq \texttt{WRA\_SCORE} \leq 5$ (Pheloung et al. 1999). Here, for purposes of illustration, we force a binary decision by rejecting species with WRA scores of 3 or larger. The decision maker's preferences, as parameterized in this paper, are not explicitly incorporated in this decision rule.

We perform the following exercise. Using the available sample (370 observations), we estimate the six decision rules described above in each of the four cases defined by the different assumptions about $c(x)$ and τ. Each species in the sample is classified as weed (ban) or non-weed (do not ban) by each decision rule. Predictions from a given decision rule are then compared with the actual outcome, resulting in economic net benefits given by the corresponding entry in Table 1 (constant cuttoff) or Table 2 (variable cutoff). By averaging over the predictions, while correcting for the assumed base rate of weeds, we obtain an (in-sample) estimate of the *per species* expected economic net benefit associated with each decision rule. In addition, we calculate sensitivity (the "true positive rate", or proportion of actual weeds classified as such) and specificity (the "true negative rate", or proportion of non-weeds classified as such).

In comparing decision rules 1-2 (ML) with decision rule 5 (MU) based on this exercise, one must keep in mind that, by construction, MU will always outperform ML in terms of in-sample utility for a given model specification. To compensate for potential in-sample overfitting of utility by the MU

Table 3: Sensitivity, specificity and expected benefits of six classification rules. The constant cutoff case ($c = 0.0682$).

Method	Sensitivity		Specificity		Expected net benefit (AU$, millions)		Relative net benefit (P.F.=100)	
	$\tau = 0.05$	$\tau = 0.02$	$\tau = 0.05$	$\tau = 0.02$	$\tau = 0.05$	$\tau = 0.02$	$\tau = 0.05$	$\tau = 0.02$
IN-SAMPLE EVALUATION								
Perfect Foresight	1.00	1.00	1.00	1.00	4.47	4.61	100.0	100.0
ML (logit)	0.74	0.57	0.86	0.96	3.01	3.89	67.5	84.4
ML (cauchit)	0.56	0.45	0.97	0.98	2.92	3.78	65.3	82.1
Bayes (logit)	0.75	0.55	0.86	0.98	3.03	3.93	67.8	85.3
Bayes (cauchit)	0.55	0.36	0.98	0.99	2.93	3.74	65.6	81.1
MU (lin. index)	0.71	0.58	0.91	0.97	3.11	3.94	69.6	85.6
WRA (weed if ≥ 3)	0.85	0.85	0.74	0.74	2.80	3.21	62.7	69.7
BOOTSTRAP EVALUATION								
Perfect Foresight	1.00	1.00	1.00	1.00	4.47	4.61	100.0	100.0
ML (logit)	0.74	0.57	0.86	0.96	3.00	3.88	67.2	84.1
ML (cauchit)	0.56	0.45	0.96	0.97	2.91	3.77	65.1	81.8
MU (lin. index)	0.70	0.58	0.88	0.97	2.99	3.90	66.8	84.7
WRA (weed if ≥ 3)	0.85	0.85	0.74	0.74	2.80	3.20	62.6	69.5

method, we extend the evaluation exercise to artificially generated bootstrap samples. In particular, we generate two bootstrap samples from the original sample and use one for estimation step and then the other for evaluation step.[5] We then repeat the exercise 2500 times and report the Monte Carlo average of each statistic computed over the evaluation samples. The Bayesian method (decision rules 3 and 4) is excluded from this exercise as computationally prohibitive since a single run of the MCMC algorithm involves a large number of iterations (20,000) replicated for 11 independent chains.

Estimation results are presented in Table 3 (the constant cutoff case) and Table 4 (the variable cutoff case). As an additional benchmark the tables also contain per species welfare figures computed under the hypothetical scenario that the decision maker has perfect foresight in predicting weed outcomes. This perfect information measure identifies the upper bound on the per species expected net benefit that is possible in each case considered. The relative net benefit statistic in Table 3 conveys the expected percentage of this maximum possible net benefit achieved under each decision rule.

The results presented in Table 3 show that, the MU decision rule can outperform the other decision rules from an expected net benefits perspective, although the advantage is modest for this constant cutoff case and attenuates in the bootstrap evaluation where ML and MU perform similarly. The WRA decision rule is conservative from a welfare standpoint given the particular cost-benefit tradeoff specified in Table 1. The WRA system does detect weeds with high probability—it has by far the highest sensitivity. However, the tradeoff is loss of specificity—it ends up excluding a greater proportion of non-weedy species. A distinct increase in relative net benefits for the alternative approaches indicate that there are welfare gains to be had by taking the decision maker's preferences into account in the classification procedure. The magnitude of these potential gains will generally

[5] Technically, we resample from the weed and non-weed subsamples separately so that each bootstrap sample has the same composition of weeds versus non-weeds as the original. This is, however, not essential.

Table 4: Sensitivity, specificity and expected benefits of six classification rules. The variable cutoff case.

Method	Sensitivity		Specificity		Expected net benefit (AU$, millions)		Relative net benefit (P.F.=100)	
	$\tau = 0.05$	$\tau = 0.02$	$\tau = 0.05$	$\tau = 0.02$	$\tau = 0.05$	$\tau = 0.02$	$\tau = 0.05$	$\tau = 0.02$
IN-SAMPLE EVALUATION								
Perfect Foresight	1.00	1.00	1.00	1.00	4.48	4.62	100.0	100.0
ML (logit)	0.62	0.50	0.94	0.98	3.66	4.16	81.7	89.9
ML (cauchit)	0.51	0.43	0.98	0.99	3.52	4.10	78.5	88.7
Bayes (logit)	0.62	0.48	0.95	0.98	3.71	4.13	82.8	89.4
Bayes (cauchit)	0.49	0.34	0.98	1.00	3.49	4.05	77.9	87.7
MU (lin. index)	0.59	0.55	0.98	0.99	3.90	4.35	87.0	94.1
WRA (weed if ≥ 3)	0.85	0.85	0.74	0.74	3.01	3.29	67.1	71.2
BOOTSTRAP EVALUATION								
Perfect Foresight	1.00	1.00	1.00	1.00	4.48	4.62	100.0	100.0
ML (logit)	0.63	0.50	0.94	0.98	3.65	4.15	81.4	89.7
ML (cauchit)	0.51	0.43	0.97	0.98	3.51	4.09	78.4	88.5
MU (lin. index)	0.58	0.54	0.96	0.98	3.80	4.28	84.8	92.5
WRA (weed if ≥ 3)	0.85	0.85	0.74	0.74	3.01	3.30	67.1	71.3

depend on the exact preference specification and the assumed base rate of weeds. It should be noted that the WRA model has been simplified to exclude the "evaluate further" classification and was not designed to maximize the objective specified here. The true expected costs and benefits of potentially invasive plants are also not known with certainty. Welfare measures presented here are conditional on the assumed cost and benefit figures in Table 1. Another perspective to take would be to ask what parameter values would lead to a higher expected net benefit per species evaluated under the WRA system given the sensitivity and specificity reported in Table 3. Under the payoff assumptions in Table 1 our assumed baseline damage to benefit ratio is $D/B = 14.6$. This ratio would need to be significantly greater for expected welfare under the WRA system to surpass welfare under the other approaches. For example, in the case of $\tau = 0.02$, the damage to benefit ratio would need to be around four times greater ($D/B = 58.8$) to motivate the WRA decision rule.

Table 3 also shows that ML, Bayes and MU estimation of the logit model generate similar levels of sensitivity and specificity and thus similar welfare results, at least in this simple setup with a constant cutoff, c. What the closeness of the estimators indicates is that the logit functional form is likely well-specified for $P(Y = 1 \mid \texttt{WRA\_SCORE})$, at least in the range where this probability is close to the cutoff $c = 0.0678$. Thus, even if the MU methodology does not significantly improve on traditional likelihood-based procedures in the constant cutoff case, it can still be used to check their soundness. Nevertheless, unless the logit model (or some other parametric specification) is exactly correctly specified for $P(Y = 1 \mid \texttt{WRA\_SCORE})$ over the entire range of observed WRA scores, MU can still potentially outperform ML and Bayes under alternative cost-benefit specifications as demonstrated next in the variable cutoff case.

In Table 4 we present results for the case in which utility given an invasive outcome is not constant but rather allowed to vary as a function of the covariates. Here the improvement provided by the

MU method in the in-sample evaluation is more substantial and holds up through the bootstrap evaluation. For both levels of τ, the in-sample difference between MU and the best classical method is about 190K$ per species assessed. For the bootstrap evaluation this difference is 150K$ for $\tau = 0.05$ and 130K$ for $\tau = 0.02$. While this is a modest relative net benefit on a per species basis, it is statistically significant (it holds up over 1,000 Monte Carlo cycles) and suggests a substantial annual figure when the rate of proposals is considered. For Australia the average number of proposals for 1997-2002 was 260 per year (Thorp 2002) and therefore an estimate of $34-$49 million in additional annual expected net benefits. The broader point suggested by the result is that if one constructs a reasonably realistic and flexible cost-benefit model, then MU can outperform classical approaches in an economically meaningful way. Previous research shows that it is precisely in these situations where the MU estimator is preferred to ML (c.f. Lieli and White, forthcoming).

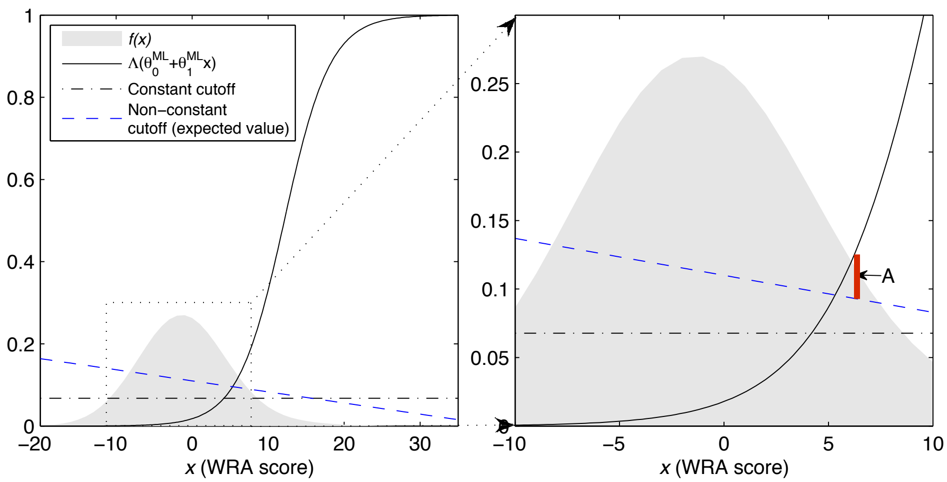

Figure 2: Graphical representation of the decision problem with $\tau = 0.06$: estimated covariate density $f(x)$, logit probability function given ML estimates, constant and non-constant cutoff functions. A magnified region of the left panel appears in the right panel. The length of line segment A measures the degree to which the probability of weediness exceeds the non-constant cutoff for a given level of x.

The relative net benefit figures in the bootstrap evaluation from Table 4 characterize the degree to which uncertainty constrains our ability to capture the entirety of expected net benefits under perfect information. Given a baseline weediness rate of 2% (5%), the percentage estimate of expected net benefits captured under the various decision rules ranges from 71.3% (67.1%) under the WRA decision rule, up to 92.5% (84.7%) under the MU decision rule.

Intuition for what drives our results emerges from examining overlapping plots of the logit probability function, the density of the covariate and the constant and non-constant cutoff functions as presented in Figure 2. While the true function $p(x;\theta)$ is unknown, the S-shaped logit function with ML estimates serves as a proxy for discussion. First, recall that the decision rule is to reject the proposed import if $p(x;\theta) > c(x)$, which occurs near 4.3 for the constant cutoff function and 5.3 for the non-constant cutoff function. The estimated covariate population density $f(x)$ appears in the background as a shaded, bell-shaped plot, and conveys a sense of how important the region of intersection is. For example, if the intersection were to fall near the mode of X, around -1, then 'drawing' species near the intersection would be relatively common and estimating this point well would be of greater importance. Alternatively, if the intersection were to fall near -20 or 20, it would be much less common to draw a species that could easily be misclassified. The impact of a misclassification is also important. The distance between $p(x;\theta)$ and $c(x)$ is a measure of the expected welfare gain from correcting a misclassification, $E\Delta U_{\bar{a}^* \to a^*}$, normalized by damages. This follows from defining $E\Delta U_{\bar{a}^* \to a^*} \equiv |p(x;\theta)D(x) - B(x)|$ and observing from equation (4) that $c(x) = B(x)/D(x)$, which implies $|p(x;\theta) - c(x)| = E\Delta U_{\bar{a}^* \to a^*}/D(x)$. For example, line segment A in Figure 2 characterizes the degree to which the probability of weediness exceeds the non-constant cutoff for a given level of x and thus a sense of the degree to which expected damages outweigh benefits (as a proportion of damage from invasion). Note that when the cutoff function is sloped in the opposite direction of $p(x;\theta)$, as in Figure 2, the expected normalized impact of misclassification grows faster over distance from the intersection, relative to the case of a constant cutoff.

5 Precaution motivated by risk aversion and measurement error

Up to this point we have ignored the implications of a risk averse decision maker and the possibility of measurement error (ME) in the WRA score of the new proposal. Next we demonstrate how reasonable models of either of these elements motivate a more conservative decision. First, suppose that X, the covariate for a proposal, is assessed with ME from a symmetric, mean-zero distribution. We consider the case consistent with our empirical approach and depicted in Figure 1: $c(x)$ is not strictly convex and $p(x,\theta)$ is a monotonically increasing convex-concave function of x, for example given a probit or logit link function with an index increasing in x. Since the invasive outcome is considered a rare event—for example Caley et al. (2006) set the 99th percentile for τ at 17%—the intersection of $c(x)$ and $p(x,\theta)$ will occur in the convex range of $p(x,\theta)$. Modifying the objective function in equation (3) to account for uncertainty over X, the only change in decision rule of equation (4) is to replace $p(x;\theta)$ with its expected value: $E^{X^0}[p(x;\theta)] = \int_{-\infty}^{\infty} p(x;\theta)g(x)dx$, where g is a symmetric, mean-zero pdf. Since $g(X^0)$ is symmetric and centered in the convex range of $p(x,\theta)$ at X, Jensen's inequality implies that $E^{X^0}[p(x;\theta)] > p(X;\theta)$. Since the relevant estimate of the expected value of the probability of invasiveness is increased under ME, this uncertainty induces a more conservative decision, that is it expands the covariate range over which "ban" is the optimal action.

Next, we relax the assumption of risk neutrality over monetary payoffs. Since we are interested in attitudes towards outcomes leading to absolute gains or losses from current wealth (as opposed to percentage changes) we employ a utility function with absolute risk aversion. Letting m represent monetary payoffs (e.g. as characterized in Table 1) we consider the exponential utility function $u = -\exp(-\omega m)$, where $\omega > 0$ is the risk aversion coefficient. This function is unique in manifesting constant absolute risk aversion. Using equation (4), the optimal cutoff is given by $c = [1 - \exp(\omega B/r)]/[1 - \exp(\omega D/r)]$. For various levels of the base rate τ, we identify the level of constant absolute risk aversion at which the MU approach generates the same decision rule as the WRA rule. These results are presented in Table 5. To interpret the reported level of risk aversion, we

Table 5: Levels of absolute risk aversion for which the WRA rule is optimal for different levels of τ.

Base rate (τ)	Cutoff (c)	Abs. risk aversion (ω)	Indiff. at $4.7M	Indiff. at $68.9M
1%	0.008	0.052	62%, 1.6:1	99.9%, 1,260:1
2%	0.017	0.036	58%, 1.4:1	99.3%, 137:1
6%	0.051	0.009	52%, 1.1:1	77%, 3.3:1
8%	0.068	risk neutral	50%, 1:1	50%, 1:1

Note: Benefits and damage are measured in millions, i.e. $B = 0.141$ and $D = 2.068$. [*Note to reader: results in this table will be expanded to include the case of* $\tau = 0.05$ *to conform with previous sections.*]

report the win percentage for a Bernoulli gamble necessary for such a decision maker to be indifferent over winning or losing a given amount. The two amounts considered are the present values of the expected stream of trade benefits ($4.7M) and invasive damages ($68.9M). For example, if the true base rate (τ) is 2% (the modal level from Caley et al. (2006)), then the risk aversion parameter that equates the WRA and optimal (MU) decision rules is $\omega = 0.036$. This implies that in order for such a decision maker to be indifferent to a Bernoulli gamble of $4.7M, the odds of winning would have to be almost 3-to-2 (58%). If the true base rate is 8%, the optimal decision rule is equal to the WRA rule under risk neutrality. Overall, for the modal and mean levels of the base rate (2% and 6%) the WRA rule is optimal for seemingly reasonable levels of risk aversion, given gambles in the neighborhood of $4.7M. For a larger bet at $68.9M and $\tau = 0.02$, the level of risk aversion appears high.

6 Discussion

While the ML, Bayes and MU approaches all offer welfare gains over an existing subjective classification method, the MU approach shows particular promise when covariate-dependent payoffs are considered. We illustrate this effect, showing how covariate dependent payoffs can reflect a situation where losses from misclassification grow more quickly as a proposal covariate deviates from the intersection or indifference point. In the case of a base rate of weeds set near the expected value, estimated additional benefits of the MU approach per species range from $130K (out-of-sample) to $190K (in-sample). Given the average rate of proposals in the area of our case study this translates to a annual estimated benefit of $34-$49 million.

Conditional on our damage and benefit estimates, the assumption of risk neutrality and the dichotomization of the WRA decision rule, the existing WRA system appears conservative. The damage to benefit ratio would have to be roughly four times greater for expected welfare under the WRA system to approach that of the alternatives presented here. However, given limited resources for risk assessment it is reasonable to suppose that a given predictive covariate will be measured with some degree of ME. For standard link functions and models of damages and benefits, we find that symmetrically distributed ME in the covariate of a proposal leads to a more conservative decision rule, i.e. the range over which proposals are rejected increases. The relationship between assessment effort and the degree of ME is an empirical question for future study. An additional component of the decision problem that would also lead to a more conservative decision rule is risk aversion in the utility function. While our baseline assumptions include risk neutrality, this is not to say that a particular entity might not reasonably choose to reflect some degree of risk aversion in their decision-making.

Future steps involve implementing a multi-variate conditional probability model and taking advantage of the flexibility of the Bayesian approach to incorporate additional expert knowledge, of particular value at low training sample sizes. We intend to explore the relative performance of the Bayesian estimation across varying sizes of the training sample and varying quality and quantity of expert prior information.

Appendix: Markov chain Monte Carlo simulation for the Bayesian model

To implement the flexible random structure for the parameters of the normal mixture model we again set a vague prior and assumed prior independence of location and scale and uniformity over μ_{Y_i} and $\log \sigma_{Y_i}$ (see Gelman et al. 2004, p. 74):

$$\pi(\mu_{Y_i}, \sigma_{Y_i}) \propto (\sigma^2_{Y_i})^{-1}. \tag{21}$$

The unnormalized posterior distribution is then given by:

$$\pi(\theta, \mu, \sigma | S_N) \propto \pi(\mu_0, \sigma_0)\pi(\mu_1, \sigma_1)L(\theta, \mu, \sigma). \tag{22}$$

Conveniently the posterior for (μ, σ) depends only on X_N and is independent of θ. Let $S_{X|Y}$ represent the set of N_Y observations on X in S_N for which $Y = y$. The posterior density for σ^2_Y is a scaled inverse chi-square distribution(Gelman et al. 2004, p. 75):

$$\sigma^2_Y | S_{X|Y} \sim \text{Scale-inv-}\chi^2 \left(N_Y - 1, \frac{1}{N_Y - 1} \sum_{x_i \in S_{X|Y}} (X_i - \bar{X}_Y)^2 \right), \tag{23}$$

where $\bar{X}_Y$ is the mean of X in $S_{X|Y}$. The posterior for μ_Y is given by:

$$\mu_Y | \sigma^2_Y, S_{X|Y} \sim \text{Normal}(\bar{X}_Y, \sigma^2_Y / N_Y). \tag{24}$$

In the MCMC simulation to estimate the posterior, each iteration begins with Gibbs sampling of (μ, σ) using (23) and (24). Then the vector θ is sampled using the Metropolis-Hastings algorithm for the posterior in (16).

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag.

Boyes, W., D. Hoffman, and S. Low (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics 40*, 3–14.

Caley, P., W. M. Lonsdale, and P. C. Pheloung (2006). Quantifying uncertainty in predictions of invasiveness. *Biological Invasions 8*, 277–286.

Corana, A., M. Marchesi, C. Martini, and S. Ridella (1987). Minimizing multimodal functions of continuous variables with the"simulated annealing" algorithm. *Transactions on Mathematical Software 13*, 262–280.

Cosslett, S. (1993). Estimation from endogenously stratified samples. *Handbook of statistics 11*, 1–43.

Elliott, G. and R. P. Lieli (2007). Predicting binary outcomes. Working papers. Department of Economics, University of Texas, Austin. Currently revise and resubmit at *Econometrica.*

Enserink, M. (1999). Predicting invasions: biological invaders sweep in. *Science 285*(5435), 1834.

Ewel, J., D. O'Dowd, J. Bergelson, C. Daehler, C. D'Antonio, L. Gómez, D. Gordon, R. Hobbs, A. Holt, K. Hopper, et al. (1999). Deliberate introductions of species: research needs. *BioScience 49*(8), 619–630.

Gelman, A., J. Carlin, H. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (2 ed.). Washington, District of Columbia: Chapman and Hall/CRC.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*(4), 457–472.

Goffe, W., G. Ferrier, and J. Rogers (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics 60*, 65–99.

Granger, C. and M. Pesaran (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting 19*, 537–560.

Hughes, G. and L. V. Madden (2003). Evaluating predictive models with application in regulatory policy for invasive weeds. *Agricultural Systems 76*, 755–774.

Imbens, G. and T. Lancaster (1996). Efficient estimation and stratified sampling. *Journal of Econometrics 74*(2), 289–318.

Jaynes, E. and G. L. Bretthorst (2003). *Probability theory: The logic of science.* Cambridge, England: Cambridge University Press.

Keller, R., D. M. Lodge, and D. C. Finnoff (2007). Risk assessment for invasive species produces net bioeconomic benefits. *Proceedings of the National Academy of Sciences 104*, 203–207.

King, G. and L. Zeng (2001). Logistic regression in rare events data. *Political Analysis 9*, 137–163.

Knowler, D. and E. Barbier (2005). Importing exotic plants and the risk of invasion: are market-based instruments adequate? *Ecological Economics 52*(3), 341–354.

Kolar, C. and D. Lodge (2001). Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution 16*(4), 199–204.

Lieli, R. and H. White (forthcoming). Construction of empirical credit scoring models based on maximization principles. *Journal of Econometrics*.

Manski, C. (1975). The maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–228.

Manski, C. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*, 313–333.

Manski, C. F. and S. R. Lerman (1977). The estimation of choice probabilities from choice based samples. *Econometrica 45*(8), 1977–1988.

Nursery and Garden Industry Australia (2004). New report shows latest trends in the Australian garden market.

Nursery Industry Association of Australia (1999). How important are Australian natives in the trade?

Olson, L. (1990). The search for a safe environment: The economics of screening and regulating environmental hazards. *Journal of Environmental Economics and Management 19*(1), 1–18.

Parmigiani, G. (2002). *Modeling in Medical Decision Making: A Bayesian Approach.* Chichester, United Kingdom: John Wiley and Sons.

Pesaran, M. and S. Skouras (2001). Decision-based methods for forecast evaluation. In M. Clemens and D. Hendry (Eds.), *Companion to Economic Forecasting.* Oxford: Blackwell Publishers.

Pheloung, P. (1995). Determining the weed potential of new plant introductions to australia. Technical report, Agriculture Protection Board, Western Australia.

Pheloung, P. C., P. A. Williams, and S. R. Halloy (1999). A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management 57*, 239–251.

Pimentel, D., R. Zuniga, and D. Morrison (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics 52*(3), 273–288.

Reichard, S. and P. White (2001). Horticulture as a pathway of invasive plant introductions in the United States. *BioScience 51*(2), 103–113.

Sinden, J., R. Jones, S. Hester, D. Odom, C. Kalisch, R. James, and O. Cacho. (2004). The economic impact of weeds in Australia. Technical report, Report to the CRC for Australian Weed Management.

Smith, C. S. (1999). Studies on weed risk assessment. Masters of applied science thesis, University of Adelaide, South Australia.

Thorp, J. (2002). Report 19972002. Technical report, National Weeds Strategy Executive Committee, Launceston, Australia.

Virtue, J., S. J. Bennett, and R. P. Randall (2004). Plant introductions in Australia: How can we resolve "weedy" conflicts of interest? In B. Sindel and S. Johnson (Eds.), *Proceedings of the 14th Australian Weeds Conference*, Sydney, pp. 42–48. Weed Society of New South Wales Inc.

Williamson, M. (1999). Invasions. *Ecography 22*(1), 5–12.