Journal of Environmental Economics and Management 68 (2014) 226–242

Contents lists available at ScienceDirect

Journal of
Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

# Risk aversion and adaptive management: Insights from a multi-armed bandit model of invasive species risk ☆

Michael R. Springborn *

*Department of Environmental Science and Policy, University of California Davis, 2104 Wickson Hall, Davis, CA 95616, United States*

A R T I C L E I N F O

*Article history:*
Received 4 March 2013
Available online 10 July 2014

*Keywords:*
Adaptive management
Approximate dynamic programming
Multi-armed bandit
Decision-making under uncertainty
Bayesian learning
Invasive species
Risk aversion

A B S T R A C T

This article explores adaptive management (AM) for decision-making under environmental uncertainty. In the context of targeting invasive species inspections of agricultural imports, I find that risk aversion increases the relative value of AM and can increase the rate of exploratory action. While calls for AM in natural resource management are common, many analyses have identified modest gains from this approach. I analytically and numerically examine the distribution of outcomes from AM under risk neutrality and risk aversion. The inspection decision is framed as a multi-armed bandit problem and solved using the Lagrangian decomposition method. Results show that even when *expected* gains are modest, asymmetry in the distribution of outcomes has important implications. Notably, AM can serve to buffer against large losses, even if the most likely outcome is a small loss.

## Introduction

Most environmental management problems involve taking action under uncertainty. Doremus (2007) argues that uncertainty is the “signature challenge” of resource policy. The role of parametric or model uncertainty in particular is a “dominant” issue in the economics of climate change (Kelly and Kolstad, 1999). It features prominently in the dynamic management of many planning problems such as water resources (Freeze et al., 1990), fisheries (Sainsbury et al., 1997), and invasive species (Eiswerth and van Kooten, 2002). In contrast to stochasticity or noise, model uncertainty can be reduced over time through learning.

Learning in environmental management may occur in several ways. Engaging in controlled R&D, separated from the system under management, is one way to gather information. For example, Kaplan et al. (2003) consider a dynamic learning situation in which pollution abatement effort can be diverted to collecting information about the contribution of different sources to the problem. Information acquisition and the resulting management (abatement) actions are treated as *separate* activities—opportunities to learn while undertaking abatement are not considered. However, hands-on management experience itself is often the primary way in which uncertainty in environmental models is reduced (Walters, 1986). This reasoning grounds the concept of adaptive management (AM) as outlined in foundational discussions by Holling (1978) and Walters and Hilborn (1978), which emphasized the need to “break out of the passive mode and learn to treat the acquisition of functional information” as deliberatively “experimental” (Walters and Hilborn, 1978, p. 183).

---

☆ This research was supported by NSF/IGERT Program Grant DGE-0114437 and the U.S. Department of Agriculture's (USDA) Economic Research Service (ERS). The research benefited from discussion and data assistance from Donna Roberts and Peyton Ferrier of the USDA's ERS.

* Fax: +1 530 752 3350.
*E-mail address:* mspringborn@ucdavis.edu

http://dx.doi.org/10.1016/j.jeem.2014.05.004
0095-0696/

Despite many calls for AM in natural resource management (Doremus, 2010), in the rare cases where such optimal endogenous learning is explicitly modeled the returns have been modest when compared to a "passive" AM approach in which there is no experimentation but policies are adjusted as information is incidentally revealed. For example, in the context of species translocation, Rout et al. (2009) find a relative expected gain from AM of 1–3%. Bond and Loomis (2009) explore the use of AM for water pollution management and estimate a 0.1% improvement. However, consideration of the dynamics of choice and learning under AM raises the question of whether a focus on the expected value of AM might be missing important elements of the story. The immediate opportunity cost of AM is given by the difference in reward between the best non-learning choice and the exploratory AM choice. Such AM opportunity costs are limited in the long run since exploration typically attenuates with learning. The benefit of AM lies in reducing future management error from imperfect understanding of a system, which is potentially large and ongoing. This asymmetry in upside risk versus downside risk suggests that the distribution of possible returns from AM might also be skewed. Furthermore, if beliefs about system mechanics are unbiased, then the most likely outcome might be a modest net loss, e.g. when exploration confirms initial notions of the best option.

In this article, I explore the determinants and distribution of returns to AM, first through a simple two-period analytical model, and second through a fully dynamic numerical case study. The analysis is motivated by research questions relating to the existence and form of asymmetry in the distribution of outcomes under AM. If outcomes are indeed asymmetric, (1) what are the implications for the expected versus most likely outcome, and (2) what is the potential for AM to buffer against large losses? The latter prospect leads to further questions: what role does risk aversion play in determining both the extent of exploration and the relative gains from AM?

A misconception among some economists and others is that an increase in risk aversion will increase the value of information (Eeckhoudt and Godfroid, 2000). However, Hilton (1981) shows using a simple example that increased risk aversion can result in a lower value of information. Hilton and those who have reiterated the result focus on the value of *perfect* information. In contrast, AM processes typically involve incremental learning. Furthermore, optimal AM involves consideration of the opportunity cost of any exploratory action. In the context of fisheries management, Hilborn and Walters (1992) have argued that risk aversion "will make experimental or probing adaptive policies look less worthwhile" (p. 498). This argument follows from the observation that risk aversion enhances the opportunity cost of an exploratory policy. However, if learning reduces the likelihood of large losses, one might expect the potential benefits of AM (the value of information) to also increase. While the AM literature is lacking in explicit analysis of risk aversion, it is acknowledged qualitatively that this element could play an important role (e.g., Groeneveld et al., 2013).

One possible explanation for the modest expected returns from AM summarized above (relative to a passive learning approach) is that AM is not differentially informative. As explained by Walters (1986), it is possible in certain cases for a passively adaptive policy to "produce essentially as informative a sequence" as AM (p. 249). Springborn and Sanchirico (2013) illustrate this in a fisheries context, finding that returns to an exploratory policy are small unless a departure from the passively adaptive policy is required to facilitate learning. I explore conditions under which AM fails or succeeds in producing a differentially informative path and thus net gains from exploration.

To assess the research questions above, the value of AM is explored in the context of adaptive allocation of scarce trade inspection resources for invasive species. Agricultural and environmental systems are vulnerable to the introduction of pests and diseases which arrive mainly via the movements of traded goods. Despite a large scale effort to mitigate risk through border inspections in the U.S., resources for inspection are constrained. Thus the need to gather and use information to target inspections is widely acknowledged. Addressing the decision problem of targeting, for example based on commodity and origin, is complicated by imperfect information over the actual hazard posed by any particular source.

In general terms, the dynamic management problem of interest here involves repeatedly allocating a given pool of management resources across a set of alternatives, each with an uncertain probability of providing a reward. Examples include allocating land parcels to different treatments, water resources to various demands, and compliance inspections across potential targets. Each of these involves apportioning units of constrained resources (parcels of land, units of water, inspections) to a particular management option (land use, water use, inspection target) generating both an immediate payoff and information for future decisions.

Fully adaptive management has an intuitive appeal but can be difficult to implement in reality (Lee, 1999; Doremus, 2007). For example, when there exists only a rough sense that one management option has greater learning value than another, the value of learning may be only subjectively assessed and not immediately comparable with the opportunity cost of exploration (e.g. McDaniels, 1995). Another typical limitation is to largely constrain the potential true states of nature. For example, Bond and Loomis (2009) advance the typical treatment of a shallow lake pollution model to incorporate learning, but limit uncertainty over a nutrient threshold to two possible levels. Other examples of using a limited set of alternatives include applications to waterfowl (Johnson, 2011) and fisheries (Sainsbury et al., 1997). Another common simplifying constraint imposed on AM problems is to limit the information gathering phase, e.g. either to occur exclusively in the first period (e.g. Costello and Karp, 2004) or until a subjective level of confidence is achieved (e.g. Sainsbury et al., 1997). The stochastic dynamic optimization framework illustrated here contributes to the adaptive resource management literature by allowing for learning about multiple parameters that may take any value in a continuous range while explicitly estimating the value of information from exploration.

A useful framework for addressing the exploitation–exploration tradeoff in this context is the multi-armed bandit (MAB) model, in which a decision maker chooses between alternatives with reward processes that operate like slot machines or