Simulating the Effects of a Green Payment Program on the Diffusion Rate of a Conservation Technology

Kenneth A. Baerenklau

The decision to adopt a potentially profitable but unfamiliar conservation technology is cast in a multi-period Bayesian framework. Specifically, dairy farmers who are both risk-averse and susceptible to peer group influence progressively learn about the true impact of adopting reduced phosphorus dairy diets on their income distributions as they repeatedly experiment with this new technology. Empirically calibrated simulations are used to examine the effects of a voluntary green payment program on the rate of technological diffusion. Results suggest that (*a*) green payments can accelerate learning and produce significant, permanent changes in behavior relatively quickly and for a reasonable cost; (*b*) shorter contracts offering larger incentives may be more cost-effective when learning plays an important role in behavioral change; and (*c*) unknown prior beliefs can reduce the efficacy of a green payment program, implying efforts to verify these priors or to ensure against them by increasing the payment level may be worthwhile.

Key Words: conservation technology adoption, cost sharing, endogenous learning, green payments, neighborhood effects, risk aversion, voluntary programs

The Empirical Problem

Madison, Wisconsin—a city of 208,000 people—has a 76,000-seat football stadium. The Wisconsin livestock industry, primarily dairy farmers, produces enough manure to completely fill the stadium *each day* (*The Economist*, 2001). Among other problems, this manure contains a high concentration of phosphorus. Phosphorus is good for dairy cows because it is necessary for milk production and reproduction; but it is bad for the environment because it eventually finds its way into lakes and streams where it causes noxious blue-green algal blooms, decreased dissolved oxygen levels, fish kills, and changes in aquatic vegetation and fish populations.

While livestock manure has been regulated based on nitrogen content for many years, the phosphorus problem has grown to the point where both state and federal agencies are beginning to change their focus (Connors, 2000; Ritchie, 2001). Recently two important findings have been receiving increased attention. First, animal scientists have known for many years that phosphorus exhibits a threshold effect in dairy cows: when cows are fed above approximately 3.3–3.8 grams of phosphorus per kilogram of dry matter (g/kg DM), it appears to have no marginal effect on either pregnancy rates or milk production; but below this concentration, both appear to decline rather quickly (Satter, 2000; Wu, Satter, and Sojo, 2000; Satter and Dhiman, 1996).

Second, agricultural extension agents have observed recently that most dairy farmers appear to be feeding their milking herds phosphorus concentrations well above this threshold level, i.e., 4.8 g/kg

Kenneth A. Baerenklau is assistant professor, Department of Environmental Sciences, University of California–Riverside. This research was funded by a Science to Achieve Results Fellowship from the U.S. Environmental Protection Agency, by a Hatch grant from the U.S. Department of Agriculture, and by the University of Wisconsin–Madison while the author was a graduate student at the University of Wisconsin–Madison. The author thanks Bill Provencher, Brad Barham, Rich Bishop, Dan Bromley, and Fred Buttel for comments and suggestions on previous drafts. Any remaining errors are the responsibility of the author.

This paper was presented at a workshop sponsored by: USDA-NRI, USDA-ERS, and NERCRD.

DM on average (McGraw, 1999). Were the state's entire dairy industry (1.3 million cows) to adopt the threshold level, phosphorus loading into the environment would be reduced by approximately 30% (a reduction of 27,000 kg daily). At the same time, each farmer would save on average $13 annually per cow (Satter and Dhiman, 1996), for a total cost savings to the industry of nearly $17 million each year. Considering the low profit margins in the dairy sector, this is not a trivial sum of money.

Despite these findings, the vast majority of farmers are not adopting reduced-phosphorus diets. There are several candidate explanations for this behavior. First, producers may perceive using a higher level of phosphorus as an inexpensive form of insurance against uncertain (but potentially large) losses from milk yield and reproductive deficiencies. This uncertainty is exacerbated by the second candidate explanation: conflicting information. Feed suppliers, veterinarians, extension agents, crop consultants, and others all influence the decisions made by farmers, but typically the goals of these individuals are not aligned with those of farmers. Feed suppliers clearly have an incentive to recommend higher levels of phosphorus and other additives to the extent they can charge higher prices for feed. Veterinarians may recommend higher phosphorus levels either to hedge against liability or as a convenient way to rule out phosphorus deficiency as the cause of a recurring problem. And even extension agents may be reluctant to recommend the latest scientific results if they are not convinced of their validity and fear a loss of rapport with farmers if they suggest counterproductive changes in farming practices.

Finally, there is the possibility that many farmers do not consider the amount of phosphorus in their feed to be a decision worth making themselves, and thus leave this decision to others. A recent survey conducted by the University of Wisconsin asked 98 dairy farmers, "Do you usually feed your milk cows supplemental phosphorus?" Eighty percent of respondents said "yes," 8% said "no," and 12% said "not sure." Next the survey asked, "What percent of phosphorus is usually included in a typical dairy cow ration?" The average for those who reported a value was 5.2 g/kg DM (high even by current standards), but 71% of respondents replied they were not sure (Jackson-Smith, 2000).

Wisconsin regulators appear to have a Pareto dominant solution to the state's phosphorus problem: if dairy farmers were to adopt the recommended level of phosphorus, they would reap higher net incomes and the public would reap better environmental quality. [One estimate (Stumborg, Baerenklau, and Bishop, 2001) suggests the benefits would be large—at least $43 million for reducing the phosphorus load in a single watershed near the city of Madison by 50%.] However, there are no indications that such changes will take place anytime soon. Although larger milk producers in various parts of the United States have started to use reduced-phosphorus diets (Satter, 2000), the vast majority of smaller farmers in Wisconsin clearly have not yet given the issue much attention.

A Possible Solution

Wisconsin agricultural producers, and in particular dairy farmers, have significant lobbying power within the state. Wisconsin long ago proclaimed itself "America's Dairyland" and most residents, whether urban or rural, continue to feel compassion for farmers. The history of agricultural regulation in Wisconsin therefore is dominated by wealth transfers from the public to farmers (often called "green payments" when they aim to promote adoption of conservation technologies) rather than the opposite, and politicians understand that instruments like taxes and standards involve significant implementation costs. A case in point: The current nonpoint source pollution control program is based primarily on voluntary cost-sharing agreements for "best management practices," and proposed modifications of this program involve even more subsidies.

Considering the scarcity of information associated with the phosphorus issue as well as the reality of regulating the dairy sector in Wisconsin, this problem would seem well suited for a green payment program to promote adoption of reduced-phosphorus diets. The subsidy would encourage farmers to scrutinize their phosphorus input levels in more detail and to consider reducing their input levels at least temporarily. If scientists are correct about the threshold level, farmers should learn this from experience, and ultimately they should become willing to keep phosphorus input levels lower even after the incentive contract has expired. Because little is known about how such an incentive program might best be designed and implemented, the goal of this study is to examine this policy mechanism in detail with empirically calibrated simulations.

Literature Review[1]

Despite a significant body of work on technology adoption theory and an ever-increasing number of empirical applications, the existing literature has rather little to say from a quantitative perspective about how economic incentives can be used to control the rate of adoption of a new technology (Kemp, 1997). One exception is Jaffe and Stavins (1995), who found evidence that technology subsidies may have a greater potential to influence conservation technology adoption than do input taxes. However, their model does not allow for any uncertainty (and therefore learning) regarding the new technology. Another exception is Shampine (1998), who addresses subjective uncertainty and considers how a social planner might hasten the rate of adoption with subsidies. Although his simulations are not empirically calibrated, Shampine concludes that net gains attainable through subsidies are small; but this result is largely due to the initially fast rate of adoption without subsidies which, in turn, is a result of the assumptions that agents are risk-neutral and observe information perfectly from a very large sample.

In contrast to Shampine's purely theoretical approach, other studies have estimated empirical models of technology adoption and conservation program participation, but these efforts generally tend to oversimplify the adoption decision. Examples here include a static logit analysis of participation in a Tennessee forest stewardship program by Bell et al. (1994), a similar probit analysis of participation in an Indiana forestry program by Nagubadi et al. (1996), a bivariate probit analysis of conservation technology adoption by Cooper and Keim (1996), and a static multinomial logit analysis of irrigation technology adoption by Green et al. (1996). While these studies draw similar conclusions regarding the importance of education, economic incentives, attitudes, awareness, and agent-specific characteristics in the adoption (or participation) decision, none provides reliable quantitative insights into the cost-effectiveness of incentive programs for promoting environmental quality. Two recent papers which do examine this issue directly are Khanna, Isik, and Zilberman (2002), and Isik and Khanna (2003), but each of these uses a single-period framework and does not incorporate endogenous learning or any type of agent interactions.

The topics these empirical papers have failed to consider have been examined in more theoretical models of technology adoption. With regard to endogenous learning, the literature has recognized two types: learning by doing and learning from others. Learning by doing is the simpler of the two processes to model and was first addressed by Lindner, Fisher, and Pardy (1979), and Stoneman (1981). Each of these models is based on the idea that when a new technology is introduced, agents initially are unfamiliar with it, and therefore attach subjective beliefs to the output distribution. By experimenting with the new technology, individual agents learn about the true output distribution from their own experiences, and then update their beliefs. An agent's optimal adoption decision therefore changes with the evolution of his or her subjective beliefs, and an intra-firm diffusion curve is generated. Jensen (1982) extended this thinking to a population of agents who differ with respect to their subjective prior beliefs regarding the new technology. In this model, the entire population continues to receive new information over time from an exogenous source. As private subjective beliefs are updated, an aggregate inter-firm diffusion curve emerges.

Learning from others is similar but relatively more complicated and has been addressed most recently. When potential adopters are uncertain about a new technology and the outcomes of others' experiments with the technology are observable, it may become rational for a forward-looking agent to postpone adoption (at least to some extent) until new information regarding the expected benefit of adoption becomes available. This means we should expect to observe agents tending to "wait and see" what happens to their neighbors (i.e., tending to free-ride on their neighbors' experiences) before they assume the private cost of experimenting with a new idea themselves. The result is an information externality that produces an inefficiently slow rate of adoption.

Because the effects of this information externality are realized only in future periods (i.e., through the posterior distribution of beliefs), agents must be modeled as forward-looking to examine the impact of these effects on their behavior. And because the optimal action by any agent is thought to depend on the actions of all other agents, agents should be modeled as strategic, as well. Unfortunately, solving for the equilibria of these strategic-dynamic games is difficult unless simple analytical forms are chosen for all relevant functions. But because such

[1] More details regarding the discussion here and in the "behavioral model" section that follows are available in Baerenklau (2003).

simplifications have undesirable consequences for empirical work, there has been extremely limited use of such models with actual data. Manski (1993a), and Bolton and Harris (1999) both examine dynamic models, but only in theoretical aspects. Munshi (2000), and Conley and Udry (2000) both use empirical data, but restrict their agents to myopic behavior.

Two papers which do estimate empirical strategic-dynamic models are Besley and Case (1997), and Foster and Rosenzweig (1995). Both take advantage of painstakingly collected data that track changing agricultural practices in developing countries in great detail over several years. But despite having high-quality data and employing structural models, both studies found only very weak evidence of forward-looking behavior. One candidate explanation for these negative results is the possibility that although this dynamic information externality may be relevant for modeling large, capital-intensive, irreversible decisions, it may be of secondary importance for smaller, less costly, reversible decisions (such as which type of seed to plant, as is the case for both of these earlier papers; or how much feed additive to use, as is the case for this study). For adoption decisions of the latter type, an alternative and nondynamic type of behavioral interaction—commonly referred to as a neighborhood effect—may have greater relevance, but has yet to receive much attention in the resource economics literature.

Neighborhood Effects

Besley and Case (1997) chose a model with forward-looking noncooperative behavior in an attempt to measure the effect on behavior of waiting to see how well a new technology works; Foster and Rosenzweig (1995) did the same to measure the effect of waiting to see how best to use a new technology. But these are not the only ways to model social interactions under uncertainty. There is also the possibility that, although agents may be very concerned with how well something works and how best to use it, they may be more directly influenced by the contemporaneous decisions made by their peers. As Banerjee (1992) states:

> There are innumerable social and economic situations in which we are influenced in our decision making by what others around us are doing. Perhaps the commonest examples are from everyday life: we often decide on what stores and restaurants to patronize or what schools to attend on the basis of how popular they seem to be. But it has been suggested by Keynes [1936], for example, that this is also how investors in asset markets behave (the famous 'beauty contest' example). In the literature on fertility choices it has frequently been suggested that various fertility decisions (how many children to have, whether or not to use contraception, etc.) are heavily influenced by what other people in the same area are doing. *It has also been suggested that the same kind of factor also influences the decision to adopt new technologies.* Voters are known to be influenced by opinion polls to vote in the direction that the poll predicts will win; this is another instance of going with the flow. The same kind of influence is also at work when, for example, academic researchers choose to work on a topic that is currently 'hot' (pp. 797–798, emphasis added).

Most generally, a neighborhood effect captures the idea that an agent's behavior can be influenced by exposure to the behavior of other agents. Such effects have been discussed and examined by social scientists for decades and have been thought to be important in a variety of situations. One earlier example is Coleman et al. (1966), who argued that academic performance by disadvantaged students could be improved far more easily through exploiting peer group dependencies than through increased school expenditures. Another example is Schelling (1971), who derived one of the first quantitative models of interactions to help explain racial segregation. Labov (1972a, b) examined the choice of spoken dialect in an interactions framework, and Loury (1977) derived a theory of persistent racial income differences that was in part based on spillovers. Consumer demand, charitable contributions, the behavior of political parties and scientific communities, smoking habits, teenage pregnancy, crime, school dropout rates, and other emergent social phenomena all have been examined in some part of the social interactions literature.[2]

Relevance to This Study

As noted above, previous researchers have speculated that neighborhood effects would be expected to influence technology adoption decisions. For example, the notion of a "network externality" would apply to situations in which the benefit derived from using a technology is a function of the total number of users of that technology.[3] Alternatively,

[2] Refer to Brock and Durlauf (2001), and Manski (1993b) for citations and additional discussion.

[3] Manifestations of this effect include the predominance of the "QWERTY" keyboard (David, 1985), VHS tape players, and the Microsoft Windows operating system.

the idea of an "informational cascade" would apply to situations in which agents who receive noisy signals regarding expected payoffs of different technologies choose to emulate the choices of "first-movers" who act relatively quickly because they have received more precise signals. And pure "conformity preference" would seem appropriate for modeling choices that produce ancillary benefits from social acceptance, being part of the "in-crowd," or the existence of a support group.

However, due in part to identification problems,[4] there remains a relative dearth of empirical research which incorporates social interactions like these into modern decision-theoretic behavioral models in order to measure the impacts of these interactions on technology adoption. But both intuition and anecdotal evidence suggest such an approach is highly relevant for modeling adoption behavior, especially among Wisconsin dairy farmers. Intuitively, for any technology (whether conservation or otherwise), making the "mainstream" adoption decision produces private benefits to the extent that it allows a producer to participate more integrally in his or her farming community. While this may be of lesser importance in areas dominated by large "factory" farms, it would seem to have greater relevance in the state of Wisconsin with its smaller farms and their attendant farming cooperatives and local community orientations.

Furthermore, for conservation technologies in particular, pressure to "do what's right" for the environment and to "be a team player" should enhance these interaction effects to the extent that farmers dislike being perceived by their communities as selfish holdouts who are unwilling to cooperate. And last, there also appears to be a mentality among farmers that while it is unwise to operate a farm as a slave to fads, no one wants to be last to adopt a new innovation shown to be truly worthwhile. According to one Wisconsin Cooperative Extension agent, "Two drivers determine whether a farmer will adopt a new technique: if he thinks it's profitable and if his peers accept it" (Andersen, 2001).

Derivation of the Behavioral Model

The adoption model used here incorporates three appealing elements: risk preferences, endogenous learning, and neighborhood effects. Risk preferences are included to account for the propensity of economic agents to care not only about the expected level of income, but also about higher moments of the income distribution as well. Endogenous learning is included because rational farmers clearly have a greater propensity to use a technology that has produced good results for them in the past, as opposed to one that has not (learning by doing), and because of the abundance of empirical evidence suggesting the importance of farmer-to-farmer communications in technology adoption decisions (learning from others). And neighborhood effects are included to capture the possible influence of contemporaneous social interactions on the rate of diffusion.

The adoption model is best presented in layers. First, a net farm income function is posited that exhibits locally constant returns to scale with respect to the size of the milking herd (h_{it}) and captures the effect of various farmer and production characteristics on farm profits:[5]

$$(1) \qquad \frac{\pi_{it}}{h_{it}} = \kappa_i + \mathbf{x}_{it}\boldsymbol{\beta} + y_{it}\beta_y + \left[\exp(\mathbf{z}_{it}\boldsymbol{\gamma} + y_{it}\gamma_y)\right]^{1/2} \cdot u_{it} + v_t ,$$

where the left-hand side, π_{it}/h_{it}, is annual profit per milk cow. Because panel data are used to estimate the model, the first term on the right-hand side, κ_i, is a farm-specific fixed effect used to account for persistent unobservable characteristics of individual farms that would tend to bias the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The next term includes a vector of farm and operator characteristics, $\mathbf{x}_{it}$, which are thought to influence the mean income level and an associated parameter vector, $\boldsymbol{\beta}$. The third term breaks out the technology choice of concern, y_{it}, and its associated coefficient, β_y, which expresses the estimated "true" impact of the technology on mean profit. If panel data on phosphorus input levels were available, these data would enter equation (3) as y_{it}. However, as mentioned previously, these data currently do not exist. Therefore, in order to obtain estimates of the parameters needed to conduct the simulations that follow, an alternative adoption scenario is utilized. Details regarding this adoption scenario are provided after the model has been fully introduced.

The remaining two terms on the right-hand side of (1) are the error components. The first shock, u_{it},

[4] Refer to Brock and Durlauf (2001), and Manski (1993b) for discussions of identification issues associated with neighborhood effects.

[5] Note that risk preferences, learning, and social interactions do not play a role in the estimation of the profit function. These factors enter into the choice model shown in equation (2).

is a time- and farm-specific shock, assumed to be drawn independently from a standard normal distribution, representing the combined idiosyncratic effect of unobservable characteristics on profits. The variance of this shock is assumed to exhibit the commonly used multiplicative form of heteroskedasticity and therefore is given by $\exp(\boldsymbol{\gamma}\mathbf{z}_{it} + y_{it}\gamma_y)$. Here, as before, $\mathbf{z}_{it}$ is a vector of farm and operator characteristics which are thought to influence the variance of income, and $\boldsymbol{\gamma}$ is an associated parameter vector. The second term in the exponential breaks out the technology choice of concern, y_{it}, and its associated coefficient, γ_y, that expresses the estimated "true" impact of the technology on the variance of income. The second shock, v_t, is a time-specific shock which is common to all farms and accounts for the expected correlation in farm profits that may be caused by weather events or factor price fluctuations. This shock is assumed to be normal with mean zero and variance σ_v^2.

The next layer of the model concerns adoption behavior. To derive a reduced-form model for empirical estimation, assume each farmer exhibits preferences over the first two moments of his or her income distribution and over a measure of peer group influence. Denote each farmer's anticipated wealth level by $\Pi_{it} \equiv W_{it} + E[\pi_{it}]$, and the farmer's anticipated standard deviation of this wealth level by Σ_{it}, where W_{it} is nonrandom wealth (here, nonfarm net worth plus nonfarm income) and $E[\pi_{it}]$ is (subjective) expected farm profit. Each farmer's optimization problem is then expressed as:

$$(2) \qquad y_{it}^* \equiv \arg\max_{y_{it}} \Bigg[\left((\Pi_{it})^{\alpha_1} - (\Sigma_{it})^{\alpha_2}\right) + \rho \cdot \left(\max\left\{\left(\frac{1}{n_{it}-1}\sum_{j\neq i} y_{jt}^* - y_{it}\right), 0\right\}\right)^2 \Bigg].$$

In each period t, each agent i selects y_{it} to maximize this reduced form.

Equation (2) has two components. The first is a function of agent wealth that captures the combined impact of risk preferences and income uncertainty on technology choice. This functional form $(\Pi^{\alpha_1} - \Sigma^{\alpha_2})$, known as "nonlinear mean standard deviation utility," was proposed by Saha (1997) and has been used recently by Isik and Khanna (2003). It is attractive because of its inherent flexibility: it can exhibit risk aversion, neutrality, and affinity corresponding to $\alpha_2 > 0$, $\alpha_2 = 0$, and $\alpha_2 < 0$; it can exhibit decreasing, constant, and increasing absolute risk aversion corresponding to $\alpha_1 > 1$, $\alpha_1 = 1$, and $\alpha_1 < 1$; and it can exhibit decreasing, constant, and increasing relative risk aversion corresponding to $\alpha_1 > \alpha_2$, $\alpha_1 = \alpha_2$, and $\alpha_1 < \alpha_2$. Therefore, instead of imposing a type of risk preferences on the model, the data are used to derive $\hat{\alpha} \equiv [\hat{\alpha}_1, \hat{\alpha}_2]$ with which hypothesis tests may be conducted. Furthermore, this functional form easily accommodates heterogeneity in risk preferences. The form of heterogeneity used here is kept relatively simple due to model tractability concerns: one set of coefficients $[\hat{\alpha}_1^H, \hat{\alpha}_2^H]$ is estimated for farmers who have completed some kind of post-high school degree program (i.e., higher education, including trade school, college, and graduate school), and one set of coefficients $[\hat{\alpha}_1^L, \hat{\alpha}_2^L]$ for those who have not (i.e., lower education).

The expectation of net farm income, $E[\pi_{it}]$, is taken over two random vectors. The first is the vector of shocks to income that enter through the profit function: $[u_i, v]$. The second is the vector of subjective beliefs regarding the profitability of the new technology: $[\beta_y, \gamma_y]$. These beliefs are assumed to be commonly held by all agents[6] and evolve through time as new information regarding the technology is revealed through agents' adoption decisions in each period. Each belief is represented by a normal random variable with equations of motion for the mean and variance derived from Bayes' rule (Box and Tiao, 1973):

$$(3.1) \qquad \beta_{yt+1} = \frac{\tau \cdot \tilde{\sigma}^2_{\beta_{yt}} \cdot \beta_{yt} + \sigma^2_{\beta_{yt}} \cdot \tilde{\beta}_{yt}}{\tau \cdot \tilde{\sigma}^2_{\beta_{yt}} + \sigma^2_{\beta_{yt}}},$$

$$(3.2) \qquad \sigma^2_{\beta_{yt+1}} = \frac{\tau \cdot \tilde{\sigma}^2_{\beta_{yt}} \cdot \sigma^2_{\beta_{yt}}}{\tau \cdot \tilde{\sigma}^2_{\beta_{yt}} + \sigma^2_{\beta_{yt}}},$$

$$(3.3) \qquad \gamma_{yt+1} = \frac{\tau \cdot \tilde{\sigma}^2_{\gamma_{yt}} \cdot \gamma_{yt} + \sigma^2_{\gamma_{yt}} \cdot \tilde{\gamma}_{yt}}{\tau \cdot \tilde{\sigma}^2_{\gamma_{yt}} + \sigma^2_{\gamma_{yt}}},$$

and

$$(3.4) \qquad \sigma^2_{\gamma_{yt+1}} = \frac{\tau \cdot \tilde{\sigma}^2_{\gamma_{yt}} \cdot \sigma^2_{\gamma_{yt}}}{\tau \cdot \tilde{\sigma}^2_{\gamma_{yt}} + \sigma^2_{\gamma_{yt}}},$$

where any quantity denoted by a tilde (~) represents either the new information (the signal) revealed about that quantity in time period t, or the variance (the noise) of that new information. The parameter

[6] This assumption is made primarily for model tractability, but seems plausible based on the extent of information sharing exhibited by the study group. Besley and Case (1997) proceed similarly.

τ is an additional scale factor: the larger is τ, the more noisy is the signal and the more weight is given to the prior belief rather than the signal. This parameter is needed because empirical estimation of the noise component is problematic. Specifically, although the analyst may be able to use the sample data to generate an unbiased estimate of the signal received by the agents in each period (here, $\tilde{\beta}_{yt}$ and $\tilde{\gamma}_{yt}$), estimating the noise associated with that signal would require knowledge of each agent's effective sample size. Therefore, τ specifies a data-driven relationship between the noise perceived by the agents and the noise perceived by the analyst in each period (here, $\tilde{\sigma}^2_{\beta_{yt}}$ and $\tilde{\sigma}^2_{\gamma_{yt}}$).

The second component of equation (2) represents the impact of social interactions on individual choices. Following Brock and Durlauf (2001), this component is adapted from mean-field theory, where the relevant neighborhood effect is assumed to be a function of the deviation from the mean behavior exhibited by an agent's peer group. Here, ρ is a parameter to be estimated; y_{it} is the choice made by agent i in period t;[7] n_{it} is the total number of members of agent i's peer group at time t; and y^*_{jt} is the expected choice (from agent i's perspective) made by member j of agent i's peer group at time t. When ρ is negative, this framework implies an agent suffers a utility loss when he or she is perceived by his or her peer group as a "laggard" with regard to adoption (specifically, when the agent's adoption level is less than the peer group average).[8] This specification is consistent with the anecdotal evidence presented earlier, and furthermore does not penalize "innovators" whose relatively early adoption decisions generate benefits for other farmers through information sharing. Moreover, a *positive* value for ρ implies an agent incurs a utility gain by lagging behind his or her peers, which could be interpreted as evidence of the type of strategic behavior investigated by Besley and Case (1997), and Foster and Rosenzweig (1995).

Parameter Estimation

Because adoption of low phosphorus diets remains a hypothetical scenario in Wisconsin, the behavioral model presented above must be calibrated with an alternative adoption scenario for which data are available. Provided certain key characteristics of this alternative scenario mimic those of the phosphorus scenario, this approach should produce parameter estimates for risk preferences and neighborhood effects which are applicable to the phosphorus scenario and can be used in the policy simulations that follow. This approach essentially treats the phosphorus scenario as an out-of-sample prediction problem; but instead of predicting how a different group of agents would behave when faced with the same adoption problem, it predicts how the same group of agents would behave when faced with a different (but similar) adoption problem.

The only other approach for providing regulators with ex ante information about expected adoption behavior would be to ask farmers to state how they would behave if offered various incentives. The debate regarding stated versus revealed preference methods continues in the field of applied economics, and further research is needed to determine the merits of each approach for technology adoption policy guidance. But because so few farmers appear to know anything about their phosphorus levels (Jackson-Smith, 2000), a stated preference method is not used here. Instead, an alternative scenario involving the selection of improved (non-native) forage varieties by Wisconsin dairy farmers is used to obtain parameter estimates. This adoption scenario exhibits several of the same key characteristics as the phosphorus scenario: it focuses on animal nutrition, it requires low capital costs, and it is easily reversible. Furthermore, this scenario shows significant variability in adoption rates during the study window—average adoption levels grew from 11.6% in 1996 to 29% in 2000—which greatly facilitates parameter identification in a model with both risk preferences and endogenous learning. And finally, the adopters of these improved varieties have organized themselves into distinct professional networks that serve as convenient indicators of peer group membership.

The behavioral model is estimated in three sequential stages. First, the error structure specified in equation (1) requires the use of a maximum-likelihood method (as developed by Griffiths and Anderson, 1982) to obtain estimates for the profit function parameters. Then the profitability signals in equation (3) are estimated by replacing β_y and γ_y in equation (1) with time-specific versions of these parameters in order to permit the technology to have different observable impacts on profits in each year. Given these signals, a total of 10 coefficients must

[7] For notational simplicity, y_{it} in the second component of equation (2) is understood to represent the extent of adoption of the new technology (0 = no adoption, 1 = full adoption).

[8] Squaring the deviation from the group mean implies the marginal utility loss is increasing in the magnitude of the deviation, and maintains continuous differentiability of the utility function.

then be estimated in equations (2)–(3): β_{y1} and γ_{y1}, the initial values for the subjective beliefs regarding the new technology; $\sigma^2_{\beta_{y1}}$ and $\sigma^2_{\gamma_{y1}}$, the initial values for the subjective uncertainty regarding these beliefs; four values for the risk preference vector, $\boldsymbol{\alpha}$; τ, the scale factor for new information; and ρ, the social interaction parameter. Recent work has estimated technology and risk preference parameters jointly (e.g., Isik and Khanna, 2003), but has excluded the roles of endogenous learning and social interactions. Introducing either of these factors complicates the estimation problem significantly and motivates the use of a sequential approach.[9]

Estimation of (2)–(3) is accomplished using maximum entropy (Golan, Judge, and Miller, 1996). This approach requires reparameterizing the estimable coefficients and the error terms whereby each is defined by a discrete support space and a probability distribution over that support space. Then, given a prior distribution specified by the analyst, the maximum entropy principle determines a posterior set of probability distributions that are as "close" as possible to these priors but that also are consistent with the observed data. Integrating these posterior distributions over the support spaces gives the coefficient point estimates.

The main advantage of the entropy framework over traditional estimation approaches is that, because the entropy metric is globally concave and the parameter space is compact, it guarantees a solution regardless of the complexity of the model and whether or not the regression system is identified in the traditional sense. This enables the analyst to say at least something instead of nothing about the information content of a complex system and/or a sparse data set. Furthermore, it facilitates the incorporation of ex ante information into the model by allowing the analyst to specify both the supports and the prior distributions for the parameter spaces.[10] Empirical applications of maximum entropy in the economics literature include Kaplan, Howitt, and Farzin (2003); Fernandez (1997); and Golan, Judge, and Karp (1996). Additional details regarding this application are provided in Baerenklau (2003).

Summary statistics for the regressors used to estimate the profit function are given in table A1 of the appendix; estimation results for the profit function are given in table A2, for the profitability signals in table A3, and for the risk preference and neighborhood effects parameters in table A4. The point estimates for the profit function generally have the correct signs and magnitudes, and those for the adoption model suggest that agents exhibit increasing absolute and relative risk aversion as well as a negative neighborhood effect. However, confidence intervals are large enough such that it is not possible to reject other forms of risk aversion or a positive neighborhood effect at the typical significance levels. But because the model does an adequate job of predicting average annual adoption levels (table A5), these point estimates are used to calibrate the simulations that follow.

Model Calibration

To simulate the effects of changes in phosphorus input levels and the incentive mechanism on net farm income, the behavioral model must be modified slightly. First, some new terms must be introduced into equation (1):

$$\text{(4)}\qquad \pi_{it} = \Big[\kappa_i + x_{1it}(\psi_{it})\cdot\beta_1 + \mathbf{x}_{it}\boldsymbol{\beta} + C(\psi_{it}) + \big[\exp(\mathbf{z}_{it}\boldsymbol{\gamma})\big]^{1/2}\cdot u_{it} + v_t\Big]\cdot h_{it}.$$

For notational simplicity, $y_{it}\beta_y$ and $y_{it}\gamma_y$ are now subsumed by $\mathbf{x}_{it}\boldsymbol{\beta}$ and $\mathbf{z}_{it}\boldsymbol{\gamma}$, respectively. The first structural change involves the revenue term (regressor x_1 in table A1). In equation (1), this term is calculated as $x_{1it} = p_{it}w_{it}$, where p_{it} is the reported price per pound of milk received by agent i in time period t, and w_{it} is the reported average daily milk production per cow (in pounds). In equation (4), this term and its associated coefficient has been extracted from $\mathbf{x}_{it}\boldsymbol{\beta}$ to emphasize that milk revenue now depends on the chosen phosphorus concentration, ψ_{it}.

Specifically, the effect of phosphorus on revenue per cow is given by:

$$\text{(5)}\qquad x_{1it}(\psi_{it}) = p_{it}\cdot w_{it}\cdot\lambda(\psi_{it}, \theta, \hat{\psi}),$$

where $\lambda(\psi_{it}, \theta, \hat{\psi}) \in [0, 1]$ is a normalized loss function that depends on the chosen phosphorus concentration (ψ_{it}), the threshold level (θ), and a lower-bound concentration ($\hat{\psi} < \theta$) below which it is assumed a farmer would never decrease phosphorus concentration due to the risk of serious health problems affecting the milking herd. This loss function takes a value of one when $\psi_{it} \geq \theta$, declines at an increasing rate as ψ_{it} is reduced below θ, and reaches zero when $\psi_{it} = \hat{\psi}$.

[9] Note that inclusion of the social interactions term requires solving for a Nash equilibrium in each period. In a similarly complex model, Besley and Case (1997) use an analogous sequential estimation procedure which is motivated by concerns regarding tractability and identification.

[10] This is a particularly useful property here because initial attempts to use maximum likelihood produced nonsensical coefficient estimates.

Equation (4) also includes a term to account for the effect of phosphorus concentration on production costs. The annual per cow cost savings a farmer can earn by reducing the phosphorus concentration fed to his or her herd is given by:

$$(6)\qquad C(\psi_{it}) = (c_P + c_I)\cdot\max\left\{\left(\psi_{it}^0 - \psi_{it}\right), 0\right\} - c_R\cdot\left(1 - \lambda(\psi_{it}, \theta, \hat{\psi})\right),$$

where c_P is the annual savings per cow per g/kg DM of phosphorus reduction, c_I is the additional per unit incentive offered by the regulator, ψ_{it}^0 is the farmer's initial baseline phosphorus concentration, and c_R is the cost to replace a milk cow after a failed pregnancy in order to maintain a constant herd size. The first component of equation (6) captures the cost savings from phosphorus reductions and the second component employs the same loss function from equation (5) to capture the additional cost of increased pregnancy failures. Equations (5) and (6) incorporate the observations presented earlier, namely that phosphorus appears to have no marginal effect on either milk production or pregnancy rates when the feed concentration exceeds the threshold level; but below this level, both begin to decline.

Note, as specified in equations (4)–(6), phosphorus affects the mean but not the variance of income.[11] When data become available from animal scientists regarding the impact of phosphorus on the variance of milk production, including this effect would be a useful extension of the model. Also note that the loss function $\lambda(\psi_{it}, \theta, \hat{\psi})$ is assumed to be quadratic. The actual shapes of these phosphorus-induced losses in milk production and pregnancies are not yet well known, but the quadratic form has several appealing properties. First, unlike a linear approximation, it generates a continuously differentiable function for both milk production and pregnancy rates. This is both beneficial for the optimization routine and perhaps more realistic than "kinked" functions. Second, because each function is anchored at two points—one point corresponding to the threshold input level (θ) and the baseline milk production/pregnancy rate, and the other point corresponding to the minimal feed concentration ($\hat{\psi}$) and no milk production/reproduction—using a higher order exponent tends to make each function flatter near θ and steeper near $\hat{\psi}$, thereby shifting the effective threshold level down. Because animal scientists believe the threshold is near θ, this is an undesirable effect. Smaller exponents (e.g., 3/2) and sigmoid-shaped functional forms had little effect on the simulation results, so the quadratic specification is adopted.

Given these changes to the profit function, the optimization problem in equation (2) becomes:

$$(7)\qquad \psi_{it}^* \equiv \arg\max_{\psi_{it}} \left[\left(\left(\Pi_{it}(\psi_{it})\right)^{\alpha_1} - \left(\Sigma_{it}(\psi_{it})\right)^{\alpha_2}\right) + \rho\cdot\left(\max\left\{\left(\frac{1}{n_{it}-1}\sum_{j\neq i}\psi_{jt}^* - \psi_{it}\right), 0\right\}\right)^2\right],$$

where ψ_{it}^* is the optimal phosphorus concentration chosen by agent i at time t, again expressed as the extent (or percentage) of adoption of the new technology; and Π_{it} and Σ_{it} are as defined previously. Compared with equation (2), this expression differs primarily in terms of its sources of uncertainty. The first random vector over which expectations are taken $[u_i, v]$ is unchanged. But the second now represents the subjective and uncertain belief regarding the true phosphorus threshold, θ. As before, this belief is commonly held by all agents and evolves through time as new information regarding the true threshold is revealed through agents' phosphorus decisions in each period.

The threshold belief is represented by a normal distribution with equation of motion given directly by Bayes' rule:

$$(8)\qquad f_{t+1}(\theta|\xi_t) \equiv \frac{\ell(\xi_t|\theta)\cdot f_t(\theta)}{\int \ell(\xi_t|\theta)\cdot f_t(\theta)\cdot d\theta}.$$

Here, ξ_t represents the set of new information revealed at time t and includes farm characteristics ($\mathbf{x}_t$ and $\mathbf{z}_t$), phosphorus input levels ($\boldsymbol{\psi}_t$), and realized net income ($\boldsymbol{\pi}_t$); $f_t(\theta)$ is the belief at time t that θ is the true threshold; and $\ell(\xi_t|\theta)$ is the conditional likelihood function for ξ_t.

The equations of motion derived from Bayes' rule and shown in equation (3) are not applicable here because agents now are learning about a threshold level. The implication of this is the following. If, in any period t, all new observations occur above the true threshold level, the signals used in equation (3) to update beliefs will not be identified. Rather, an infinity of signals will describe the new information equally well. However, direct application of Bayes' rule remains possible because it is not necessary to determine these signals in order to calculate the posterior distribution in equation (8). In addition, this methodological change does not

[11] However, partial adoption remains possible (and likely) due to the combined effects of risk aversion and uncertainty.

affect the fundamental learning structure assumed in the previous estimation—in both cases, learning is Bayesian.

Policy Simulations

To incorporate additional key features of the Wisconsin dairy industry and of the phosphorus decision faced by dairy farmers, the following assumptions are made:

- The farms used to calibrate the behavioral model presented above constitute a single peer group and are contained in a single watershed. The relevant characteristics of these farms (including baseline milk yields and herd sizes) are set equal to reported values for the year 2000.
- The true threshold level for phosphorus in feed is assumed to be 3.5 g/kg DM, but farmers are uncertain about this true level and initially believe it is significantly higher.
- The lower-bound concentration, $\hat{\psi}$ in equations (5) and (6), is assumed to be 1.5 g/kg DM (Valk and Šebek, 1999; Call et al., 1987).
- The annual savings per cow from reducing phosphorus by 1 g/kg DM is set equal to $10 (Satter, 2000; Satter and Dhiman, 1996).
- The cost to replace a milk cow after a failed pregnancy in order to maintain a constant herd size is set equal to $335, the difference between the cost of a replacement heifer and the price received for a slaughter cow (Wisconsin Agricultural Statistics Service, 2001).

The first set of simulations presented here characterizes a baseline scenario from which to measure the effects of a green payment program on the adoption process. All simulations are conducted over 15-year time horizons, and a discount factor of 0.96 is used to calculate the present value of program costs. Other simulation parameters characterizing this baseline scenario are summarized in table 1. As shown in table 1, prior beliefs are assumed to be normal with a mean of 4.5 g/kg DM and a variance of 0.09 in the baseline scenario. No payment is offered in this scenario, and therefore the contract length is not applicable. The scale parameter adds a small random disturbance to each agent's optimal choice to simulate the unexplained portions of these decisions. Both risk preferences and neighborhood effects are permitted to influence adoption decisions in the baseline scenario, and therefore the estimates reported in table A4 of the appendix are used in this simulation.

Table 1. Simulation Parameters for the Baseline Scenario

Parameter	Value
Initial mean belief about threshold level	4.5 g/kg DM
Measure of initial uncertainty about threshold level (variance of initial belief)	0.09
Incentive for reducing phosphorus	none
Number of years incentive is offered	NA
Scale parameter for choice disturbances	0.1

Before presenting the results for this scenario, it is important to note that although there is no incentive payment here, it should not be considered analogous to the case of no government intervention. This is because the current scenario regarding phosphorus use in Wisconsin—the true "no intervention" case—is characterized by the fact that most farmers do not yet consider the amount of phosphorus in their feed to be a decision worth making themselves, and consequently leave this decision to others. The first scenario presented here instead should be thought of as a baseline *conditional on farmers examining their phosphorus input decisions in more detail.* Achieving this higher level of scrutiny itself may require some form of government intervention.

The baseline scenario in table 1 produces initial phosphorus choices in the first period of the simulations which are well within the realistic range: these initial feed concentrations vary between 4.28 and 5.20, with an average of 4.74 g/kg DM. Figure 1 shows the subsequent evolution of the mean threshold belief and the mean choice level for each period in the baseline scenario, as well as the results for the same model with no social interactions ("no group effect"). In both models, agents clearly are learning that the true threshold level is less than 4.5 g/kg DM, but their rate of learning is rather slow and appears to diminish over time. The neighborhood effect has only a small impact on behavior, but it has relatively more importance later in the simulations after uncertainty about the threshold level has been reduced. While this implies risk aversion and subjective beliefs may be the primary determinants of this adoption decision, social interactions are nonetheless retained as part of the baseline scenario.

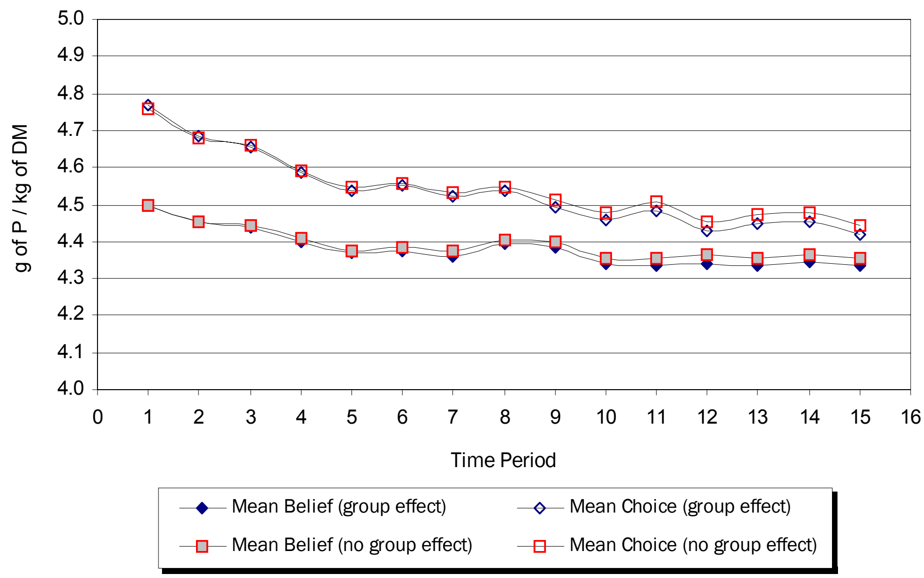

Figure 1. Evolution of mean threshold belief and mean choice level in baseline scenario

Green Payments: Uniform Input Reduction Subsidy

To hasten the rates of learning and adoption, a uniform input reduction subsidy is considered. This mechanism most closely resembles the cost-sharing arrangements currently used in Wisconsin and other states. Here, the regulator chooses a uniform per unit payment for phosphorus reductions and a common contract length in order to achieve a desired pollutant load reduction. Table 2 shows the parameters characterizing these simulations, and figure 2 presents the results graphically along with the original baseline from figure 1.

Increasing the payment level has two related effects. First, it promotes faster learning. Figure 2 shows that as the payment increases from \$0 to \$200, beliefs converge faster toward the true value of the threshold. But after the contracts expire in the tenth period, learning rates slow and are (not surprisingly) similar in nature to the path for the baseline scenario with no subsidy payment. Second, increasing the payment level promotes lower phosphorus input levels both during and after the contract window. For sufficiently high payment levels (e.g., \$200), farmers temporarily choose input levels that are *below* the true threshold level, thereby incurring temporary production losses for which they are compensated via the payment mechanism. But regardless of the payment level, choice levels

Table 2. Simulation Parameters for the Subsidy Mechanism

Parameter	Value
Initial mean belief about threshold level	4.5 g/kg DM
Measure of initial uncertainty about threshold level (variance of initial belief)	0.09
Incentive for reducing phosphorus: annual payment per cow for reducing phosphorus concentration in feed by 1 g/kg DM [a]	\$50, \$100, \$200
Number of years incentive is offered [b]	10
Scale parameter for choice disturbances	0.1

[a] For example, if a farmer with 100 cows reduced phosphorus concentration from 4.8 to 3.8 g/kg DM, this farmer would receive a total annual subsidy payment of \$5,000, \$10,000, and \$20,000 in each respective simulation.
[b] Ten years is a typical contract length for Wisconsin's nonpoint source water pollution control program.

"rebound" when the contracts expire and the payments are discontinued.[12]

The expected post-contract choice levels are of particular importance to a regulator who is tasked with meeting an exogenously imposed load reduction standard within a fixed amount of time. Also of interest are the expected total (present value) costs to achieve each of these post-contract levels

[12] Note that to attain the socially optimal pollution level indefinitely, payments also must continue indefinitely. But questions of social efficiency are not the focus here.

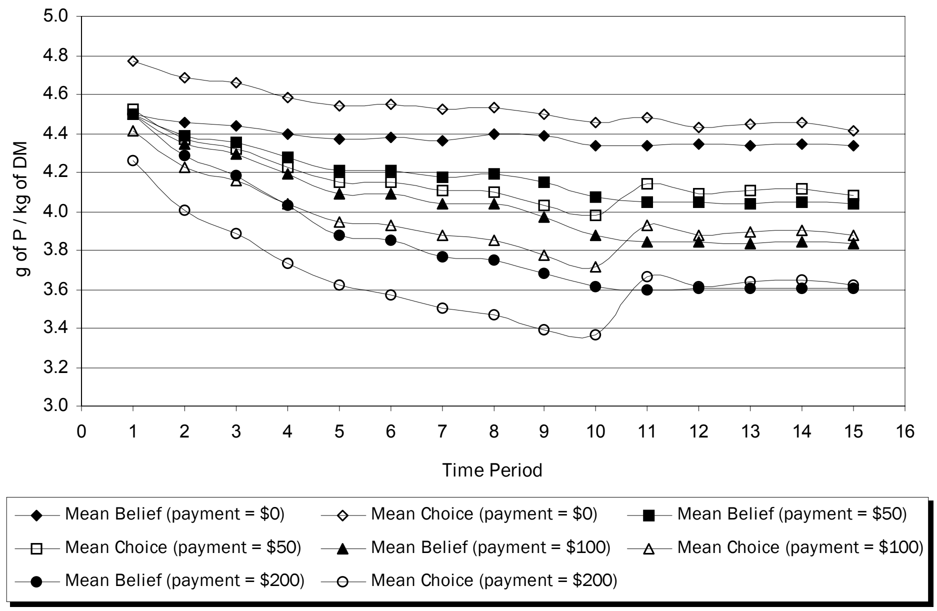

Figure 2. Evolution of mean threshold belief and mean choice level for varying subsidy payments

(summarized in table 3). By the end of the fifteenth period, the $50 contract achieves an expected permanent load reduction of 9,049 kg of phosphorus per year for an expected total cost of $535,179 to the regulator. The $100 contract achieves an expected annual load reduction of 11,751 kg of phosphorus for an expected cost of $1,425,875. And the $200 contract achieves an expected annual load reduction of 15,070 kg of phosphorus for an expected cost of $3,948,617. Each of these may be compared with an expected annual load reduction of 4,716 kg of phosphorus with no incentive payment in the baseline scenario.[13]

As a "reality check," the results of these simulated policies also may be compared with figures from an actual watershed currently enrolled in Wisconsin's nonpoint program.[14] The Lake Mendota Watershed is a 590 square-kilometer drainage basin located in south-central Wisconsin. Land use in this region is dominated by agricultural activity which accounts for 60% of the total area and is the main cause of several different water quality problems in the basin. To address these problems, the Wisconsin

Table 3. Summary of Load Reductions and Total Present Value Costs for Contracts Presented in Figure 2

Payment ($/g/kg DM)	Total Load Reduction (kg P)	Total Cost ($)
0	4,716	0
50	9,049	535,179
100	11,751	1,425,875
200	15,070	3,948,617

Department of Natural Resources (WDNR) added this watershed to its priority list in 1993, and has since developed a pollution control plan for the watershed expected to cost $17.8 million over 10 years.

Although this figure includes the costs to achieve several different water quality objectives, the watershed's namesake, Lake Mendota, is the predominant receiving body in the drainage basin and is a main focus of the pollution control plan. Lake Mendota's primary pollution problem is excessive phosphorus loading. Based on WDNR estimates, the lake receives approximately 34,000 kg of the nutrient each year. The control plan specifies that this load must be reduced by 17,000 kg within 10 years. Considering the total cost of the control plan, even a conservative estimate of the WDNR's cost to achieve

[13] Each of these load reductions is calculated as in Wu, Satter, and Sojo (2000) by converting phosphorus in feed to manure phosphorus. Also note that very little load reduction occurs after the tenth year in any scenario.

[14] This discussion is taken from Stumborg, Baerenklau, and Bishop (2001).

this single objective likely would exceed the $3.95 million needed to achieve the 15,070 kg load reduction after 10 years in the $200 contract simulations, implying a voluntary program may be a viable cost-effective alternative for the WDNR.[15]

In addition to altering the payment level as in figure 2, the simulations also permit examining the impact of varying the contract length on beliefs and behavior. For example, it may be more cost-effective to achieve the same load reduction by offering a larger incentive over a shorter period of time than a smaller incentive over a longer period of time. Figure 3 provides two scenarios for comparison: $200 over ten years and $325 over five years.[16] The long-run effects of the two scenarios are similar, but the shorter contract is slightly more cost-effective. The $200, ten-year contract achieves a permanent load reduction of 15,070 kg for a total cost of $3,948,617, and the $325, five-year contract achieves a permanent load reduction of 15,177 kg for a total cost of $3,833,387. These results suggest the WDNR should consider offering shorter contracts with larger incentives when learning plays an important role in adoption.[17]

The simulations also permit examination of the effect of alternative prior beliefs on the outcome of the incentive mechanism. This would be of particular interest to a regulator who is unsure of the priors held by the agents who will be offered the incentive (a likely scenario) to the extent that an incorrect assumption regarding these priors can have a detrimental impact on the efficacy of the incentive scheme. To explore this possibility, the $100-for-10-years contract with prior mean and variance of 4.5 g/kg DM and 0.09 (as in figure 2) is used as a baseline. Two alternative sets of priors which produce similar distributions of initial choices are then considered: mean = 4.4 g/kg DM and variance = 0.16, and mean = 4.6 g/kg DM and variance = 0.04.

Figure 4 shows the results for all three scenarios. It is apparent that the final outcome of the incentive program is fairly sensitive to the initial beliefs (note how the paths for each scenario *diverge* through time), implying this type of miscalculation by the regulator could be costly, and suggesting that efforts to discover the actual initial beliefs rather than relying on the initial distribution of choices may be worthwhile. An alternative approach when a regulator is uncertain of the prior beliefs would be to increase the incentive payment to ensure against priors that would produce relatively slow rates of learning (such as those given by the third scenario in figure 4). Though not shown here, additional simulations demonstrate that the $200-for-10-years contract appears sufficient for each set of priors considered above: the mean post-contract choice levels for all scenarios are between 3.59 and 3.68 g/kg DM, while the expected program costs remain between $3.9 and $4.1 million.

Summary and Conclusion

The simulation results presented here are based on a novel microeconomic model of rational choice under uncertainty which incorporates three key behavioral elements: risk preferences, endogenous learning, and peer group influence. Because data are unavailable for the hypothetical situation addressed by the simulations, parameter calibration is accomplished using an alternative but structurally similar adoption decision faced by the sample population. The calibration results generally are good and suggest all three elements affect adoption behavior, but peer group influence is of lesser importance.

Based on the simulation results, a green payment program can accelerate learning and produce significant, permanent changes in behavior relatively quickly and for a reasonable cost. In addition, the simulations suggest that, compared with typical cost-sharing arrangements, shorter contracts offering larger incentives may be able to achieve load reduction targets more cost-effectively when learning plays an important role in behavioral change. But they also demonstrate the potential impact of unknown prior beliefs on the efficacy of a green payment program, implying that efforts to verify these priors or to ensure against them by increasing the payment level may be well justified.

Although a green payment program appears to be a viable approach to Wisconsin's current phosphorus problem, additional research is needed to determine the best course of action. Possible topics

[15] This comparison assumes reductions in excess manure phosphorus (i.e., the load reductions calculated for these simulations) produce equivalent reductions in ambient phosphorus (i.e., the load reductions referenced for Lake Mendota), and therefore represents a best-case scenario. If only a fraction of excess manure phosphorus eventually migrates to the receiving water body, the cost of a voluntary program that meets the ambient loading objective would increase. However, Ebeling et al. (2002) recently have shown that converting to a low phosphorus diet can reduce runoff concentrations by up to 90%, implying actual program costs may not be much higher than those calculated here.

[16] The first scenario is identical to that in figure 2.

[17] Additional present value cost savings of approximately $700,000 can be obtained by postponing implementation of the shorter contract (i.e., offering the five-year contract during years 6–10 instead of during years 1–5). The drawback of this approach is higher interim pollution damages, but it nonetheless attains a post-contract pollution level similar to the 10-year contract.

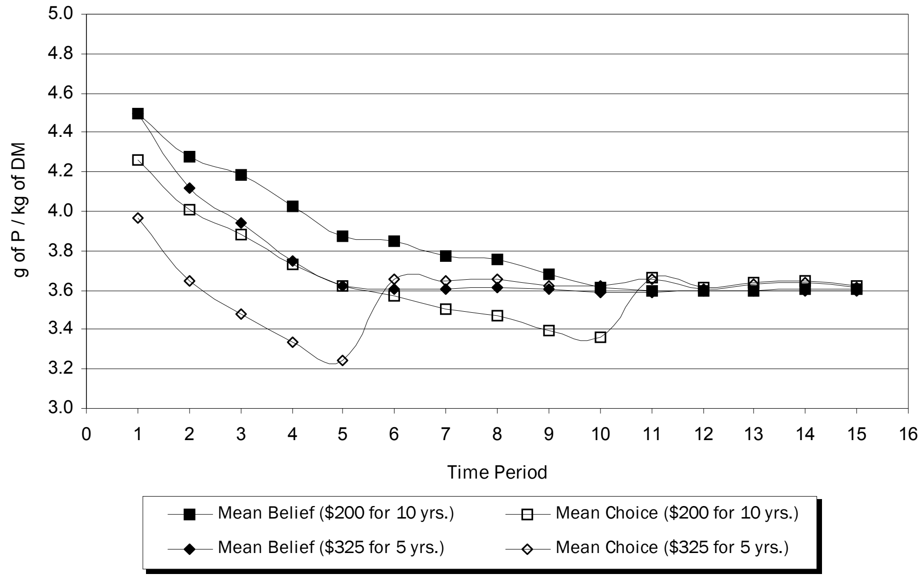

Figure 3. Evolution of mean threshold belief and mean choice level for varying contract lengths

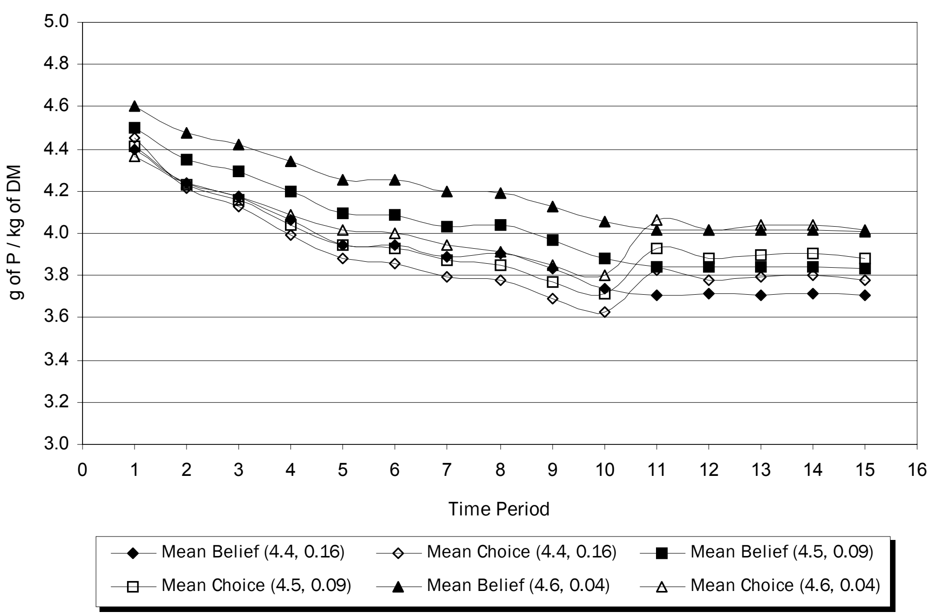

Figure 4. Evolution of mean threshold belief and mean choice level for differing prior beliefs

include: (*a*) obtaining better information on current phosphorus input levels and willingness to adopt reduced phosphorus diets, (*b*) targeting incentives at a subset of the population to reduce program costs, (*c*) examining alternative specifications of the adoption model and/or the learning mechanism, and (*d*) investigating how "green insurance" might be used to achieve the same pollution reduction goals.

References

Andersen, N. (2001, January). Dairy/livestock agent, University of Wisconsin Cooperative Extension, Dane County. Personal communication.

Baerenklau, K. A. (2003). "Testing for Risk Preferences, Endogenous Learning, and Neighborhood Effects in the Diffusion of an Innovation with Uncertain Profitability." Working paper, Department of Environmental Sciences, University of California, Riverside.

Banerjee, A. (1992). "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* CVII(3), 797–817.

Bell, C. D., R. K. Roberts, B. C. English, and W. M. Park. (1994). "A Logit Analysis of Participation in Tennessee's Forest Stewardship Program." *Journal of Agricultural and Applied Economics* 26, 463–472.

Besley, T., and A. Case. (1997). "Diffusion as a Learning Process: Evidence from HYV Cotton." Staff paper, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ.

Bolton, P., and C. Harris. (1999). "Strategic Experimentation." *Econometrica* 67(2), 349–374.

Box, G., and G. Tiao. (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

Brock, W., and S. Durlauf. (2001). "Interactions-Based Models." In J. J. Heckman and E. E. Leamer (eds.), *Handbook of Econometrics*, Vol. 5 (pp. 3297–3380). New York: Elsevier Science.

Call, J., J. Butcher, J. Shupe, R. Lamb, R. Boman, and A. Olson. (1987). "Clinical Effects of Low Dietary Phosphorus Concentrations in Feed Given to Lactating Dairy Cows." *American Journal of Veterinary Research* 48(1), 133–136.

Coleman, J., E. Campbell, J. Hobson, J. McPartland, A. Mood, F. Weinfeld, and R. York. (1966). *Equality of Education Opportunity.* Washington, DC: U.S. Government Printing Office.

Conley, T., and C. Udry. (2000). "Learning About a New Technology: Pineapple in Ghana." Discussion Paper No. 817, Economic Growth Center, Yale University, New Haven, CT.

Connors, K. (2000, August). County conservationist, Dane County Land Conservation Department, Madison, WI. Personal communication.

Cooper, J. C., and R. W. Keim. (1996). "Incentive Payments to Encourage Farmer Adoption of Water Quality Protection Practices." *American Journal of Agricultural Economics* 78, 54–64.

David, P. (1985). "Clio and the Economics of QWERTY." *American Economic Review* 75, 332–337.

Ebeling, A. M., L. G. Bundy, J. M. Powell, and T. W. Andraski. (2002). "Dairy Diet Phosphorus Effects on Phosphorus Losses in Runoff from Land-Applied Manure." *Soil Science Society of America Journal* 66, 284–291.

The Economist [staff]. (2001, October 20). "Manure Happens." *The Economist*, p. 35.

Fernandez, L. (1997). "Estimation of Wastewater Treatment Objectives Through Maximum Entropy." *Journal of Environmental Economics and Management* 32, 293–308.

Foster, A., and M. Rosenzweig. (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy* 103(6), 1176–1209.

Golan, A., G. Judge, and L. Karp. (1996). "A Maximum Entropy Approach to Estimation and Inference in Dynamic Models, or Counting Fish in the Sea Using Maximum Entropy." *Journal of Economic Dynamics and Control* 20, 559–582.

Golan, A., G. Judge, and D. Miller. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* West Sussex, England: John Wiley & Sons, Ltd.

Green, G., D. Sunding, D. Zilberman, and D. Parker. (1996). "Explaining Irrigation Technology Choices: A Microparameter Approach." *American Journal of Agricultural Economics* 78, 1064–1072.

Griffiths, W., and J. Anderson. (1982). "Using Time-Series and Cross-Section Data to Estimate a Production Function with Positive and Negative Marginal Risks." *Journal of the American Statistical Association* 77, 529–536.

Isik, M., and M. Khanna. (2003). "Stochastic Technology, Risk Preferences, and Adoption of Site-Specific Technologies." *American Journal of Agricultural Economics* 85(2), 305–317.

Jackson-Smith, D. (2000). "Report to 1999 Wisconsin Dairy Herd Feeding Study Participants, Part I: A Profile of How Wisconsin Dairy Farm Operators Feed Their Milking Herds." Program on Agricultural Technology Studies, University of Wisconsin, Madison.

Jaffe, A., and R. Stavins. (1995). "Dynamic Incentives of Environmental Regulations: The Effects of Alternative Policy Instruments on Technology Diffusion." *Journal of Environmental Economics and Management* 29, S43–S63.

Jensen, R. (1982). "Adoption and Diffusion of an Innovation of Uncertain Profitability." *Journal of Economic Theory* 27(1), 182–193.

Kaplan, J., R. Howitt, and Y. Farzin. (2003). "An Information-Theoretical Analysis of Budget-Constrained Nonpoint Source Pollution Control." *Journal of Environmental Economics and Management* 46, 106–130.

Kemp, R. (1997). *Environmental Policy and Technical Change.* Cheltenham, UK: Edward Elgar.

Khanna, M., M. Isik, and D. Zilberman. (2002). "Cost-Effectiveness of Alternative Green Payment Policies for Conservation Technology Adoption with Heterogeneous Land Quality." *Agricultural Economics* 27, 157–174.

Labov, W. (1972a). *Language in the Inner City.* Philadelphia: University of Pennsylvania Press.

———. (1972b). *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Lindner, R., A. Fischer, and P. Pardey. (1979). "The Time to Adopt." *Economic Letters* 2,187–190.

Loury, G. (1977). "A Dynamic Theory of Racial Income Differences." In P. A. Wallace and A. M. LaMond (eds.), *Women, Minorities, and Employment Discrimination* (pp. 153–186). Lexington, MA: Lexington Books.

Manski, C. (1993a). "Dynamic Choice in Social Settings." *Journal of Econometrics* 58, 121–136.

———. (1993b). "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60, 531–542.

McGraw, L. (1999, July 16). "Cutting Phosphorus in Dairy Cattle Feed Will Save Money, Help Environment." Press release, U.S. Department of Agriculture/Agricultural Research Service. Online at USDA/ARS website: http://www.ars.usda.gov/is/pr/1999/990716.htm.

Munshi, K. (2000). "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." Working paper, Department of Economics, University of Pennsylvania, Philadelphia.

Nagubadi, V., K. T. McNamara, W. L. Hoover, and W. L. Mills, Jr. (1996). "Program Participation Behavior of Nonindustrial Forest Landowners: A Probit Analysis." *Journal of Agricultural and Applied Economics* 28, 323–336.

Ritchie, J. (2001, June/July). "Guess What Hit the Fan ..." *Today's Farmer*, pp. 4–6.

Saha, A. (1997). "Risk Preference Estimation in the Nonlinear Mean Standard Deviation Approach." *Economic Inquiry* 35, 770–782.

Satter, L. (2000, September). Professor, Department of Animal Science, and Researcher, U.S. Dairy Forage Research Center, University of Wisconsin, Madison. Personal communication.

Satter, L., and T. Dhiman. (1996). "Formulating Dairy Diet—Enhancing Profitability and Reducing Environmental Impact." Paper presented at the 1996 Informational Conference with Dairy and Forage Industries, U.S. Dairy Forage Research Center, University of Wisconsin, Madison. Online. Available at http://www.dfrc.ars.usda.gov/Research_Summaries/ind_meet/dfrc13.pdf.

Schelling, T. (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1, 143–186.

Shampine, A. (1998). "Compensating for Information Externalities in Technology Diffusion Models." *American Journal of Agricultural Economics* 80, 337–346.

Stoneman, P. (1981). "Intra-Firm Diffusion, Bayesian Learning, and Profitability." *Economic Journal* 91, 375–388.

Stumborg, B., K. A. Baerenklau, and R. C. Bishop. (2001). "Nonpoint Source Pollution and Present Values: A Contingent Valuation Study of Lake Mendota." *Review of Agricultural Economics* 23(1), 120–132.

Valk, H., and L. Šebek. (1999). "Influence of Long-Term Feeding of Limited Amounts of Phosphorus on Dry Matter Intake, Mild Production, and Body Weight of Dairy Cows." *Journal of Dairy Science* 82, 2157–2163.

Wisconsin Agricultural Statistics Service. (2001, October 2). Monthly Prices Received by Farmers. Online. Available at http://www.nass.usda.gov/wi/misc/agprices.pdf.

Wu, Z., L. Satter, and R. Sojo. (2000). "Milk Production, Reproductive Performance, and Fecal Excretion of Phosphorus by Dairy Cows Fed Three Amounts of Phosphorus." *Journal of Dairy Science* 83, 1028–1041.

Appendix

Table A1. Summary of Regression Variables in Net Farm Income Function (n = 34, t = 5)

Variable	Definition	Sample Mean	Std. Deviation
π / h	Reported net farm income per cow	449.268	474.413
Mean Function:			
κ_i	Farm-specific fixed effect	0.029	0.169
x_1	Estimated daily farm revenue per cow from milk sales (price per pound × pounds produced per cow per day)	6.491	1.408
x_2	Acres of farmable land per cow	3.648	1.709
x_3	Acres of pasture per cow	1.545	0.725
x_4	Percent Holsteins × acres of pasture per cow	1.204	0.758
x_5	Dummy for use of a computerized record keeping system	0.471	0.499
x_6	Dummy for use of freestall housing	0.271	0.444
x_7	Percent of farm assets owned by operator	0.725	0.280
x_8	Acres of pasture planted with improved grasses per cow	0.301	0.371
Variance Function:			
z_1	Dummy for farms located in the southwest region	0.471	0.499
z_2	Dummy for farms located in the north-central region	0.352	0.478
z_3	Dummy for farms located in the east region	0.176	0.381
z_4	Years of experience as a dairy farmer	17.647	10.742
z_5	Pounds of milk produced each day per cow	47.046	9.469
z_6	Acres of farmable land per cow	3.648	1.709
z_7	Acres of pasture per cow	1.545	0.725
z_8	Percent Holsteins × acres of pasture per cow	1.204	0.758
z_9	Dummy for use of a computerized record keeping system	0.471	0.499
z_{10}	Dummy for use of freestall housing	0.271	0.444
z_{11}	Percent of farm assets owned by operator	0.725	0.280
z_{12}	Acres of pasture planted with improved grasses per cow	0.301	0.371

Table A2. Maximum Likelihood Estimates for Net Farm Income Function

Coefficient	Point Estimate	Asymptotic Standard Error	Asymptotic *Z*-Statistic	Asymptotic *p*-Value
Mean Function:				
β_1	117.775	12.776	9.218	< 0.01
β_2	!31.403	24.696	!1.272	0.20
β_3	22.191	48.139	0.461	0.64
β_4	!154.815	45.771	!3.382	< 0.01
β_5	14.993	50.775	0.295	0.77
β_6	!141.637	53.858	!2.630	< 0.01
β_7	!30.045	113.319	!0.265	0.79
β_8	171.889	45.319	3.793	< 0.01
Variance Function:				
γ_1	9.986	0.807	12.378	< 0.01
γ_2	9.732	0.816	11.925	< 0.01
γ_3	8.112	0.934	8.686	< 0.01
γ_4	!0.008	0.014	!0.575	0.57
γ_5	!0.011	0.015	!0.709	0.48
γ_6	0.109	0.074	1.460	0.14
γ_7	0.105	0.256	0.410	0.68
γ_8	!0.506	0.233	!2.168	0.03
γ_9	!0.202	0.257	!0.786	0.43
γ_{10}	1.328	0.307	4.322	< 0.01
γ_{11}	0.443	0.437	1.013	0.31
γ_{12}	0.422	0.339	1.244	0.21
Common Variance Component and Model Fit:				
σ_v^2 = 242.531				
Adjusted R^2 = 0.638				

Table A3. Estimated Profitability Signals, 1996–2000

Year	$\tilde{\beta}_{yt}$	$\tilde{\sigma}^2_{\beta_{yt}}$	$\tilde{\gamma}_{yt}$	$\tilde{\sigma}^2_{\gamma_{yt}}$
1996	228.085	3,395.159	!0.238	0.644
1997	89.701	3,267.265	!0.347	0.568
1998	75.482	6,287.221	1.342	0.385
1999	235.496	2,431.771	0.193	0.307
2000	163.989	1,721.005	0.273	0.172

Notes: $\tilde{\beta}_{yt}$ and $\tilde{\gamma}_{yt}$ are the signals received in each period; $\tilde{\sigma}^2_{\beta_{yt}}$ and $\tilde{\sigma}^2_{\gamma_{yt}}$ are the associated variances.

Table A4. Maximum Entropy Estimates for Adoption Model with Neighborhood Effects

Coefficient	Description	Point Estimate	Asymptotic Std. Error
α_1^L	Risk coefficient on mean income for higher education level	0.631	0.313
α_1^H	Risk coefficient on mean income for lower education level	0.681	0.246
α_2^L	Risk coefficient on standard deviation of income for higher education level	0.897	0.460
α_2^H	Risk coefficient on standard deviation of income for lower education level	0.967	0.303
ρ	Neighborhood effects coefficient	!3,583.5	9,748.6

Table A5. Actual and Predicted Annual Mean Adoption Levels in Calibration Scenario, 1996–2000 (%)

Year	Actual	Predicted
1996	11.6	12.2
1997	14.9	19.9
1998	21.6	21.4
1999	23.5	19.9
2000	29.0	28.0