

Automated Strategy Searches in an Electronic Goods Market: Learning and Complex Price Schedules

Christopher H. Brooks * Scott Fay † Rajarshi Das ‡
Jeffrey K. MacKie-Mason § Jeffrey O. Kephart §
Edmund H. Durfee †

1 June 1999

Abstract

In an automated market for electronic goods new problems arise that have not been well studied previously. For example, information goods are very flexible. Marginal costs are negligible and nearly limitless bundling and unbundling of these items are possible, in contrast to physical goods. Consequently, producers can offer complex pricing schemes. However, the profit-maximizing design of a complex pricing schedule depends on a producer's knowledge of the distribution of consumer preferences for the available information goods. Preferences are private and can only be gradually uncovered through market experience. In this paper we compare dynamic performance across price schedules of varying complexity. We provide the producer with two machine learning methods (function approximation and hill-climbing) which implement a strategy that balances exploitation to maximize current profits against exploration of the profit landscape to improve future profits. We find that the tradeoff between exploitation and exploration is different depending on the learning algorithms employed, and in particular depending on the complexity of the price schedule that is offered. In general, simpler price schedules are more robust and give up less profit during the learning periods even though in our stationary environment learning eventually is complete and the more complex schedules have higher long-run profits. These results hold for both learning methods, even

*Artificial Intelligence Laboratory, University of Michigan, Ann Arbor, MI 48109
{chbrooks,durfee}@umich.edu

†Department of Economics, University of Michigan, Ann Arbor, MI 48109, scottfay@umich.edu

‡Institute For Advanced Commerce, IBM Research, PO Box 704, Yorktown Heights, NY 10598,
{rajarshi,kephart}@watson.ibm.com

§School of Information and Department of Economics, University of Michigan, Ann Arbor, MI 48109, jmm@umich.edu

⁰This work was supported in part by an IBM University Partnership Grant and DARPA grant F30602-97-1-0228, part of the Information Survivability program

though the relative performance of the methods is quite sensitive to choice of initial conditions and differences in the smoothness of the profit landscape for different price schedules. Our results have implications for automated learning and strategic pricing in non-stationary environments, which arise when the consumer population changes, individuals change their preferences, or competing firms change their strategies.

1 Introduction

Electronic goods are very flexible. In contrast to physical goods, marginal costs are negligible and nearly limitless bundling and unbundling of these items are possible. Consequently producers can offer complex pricing and bundling schemes that would be infeasible for traditional commerce in physical goods. Suppose goods can be specified with a listing of A attributes, each of which can take on $V \geq 2$ values. Then, if bundles of any number of such goods can be packaged together, there will be 2^{V^A} different possible bundles. To add further complexity, there are many different pricing structures. Considering only pricing structures that are based on the number of items in a bundle, and not on the identity of the items, there are families of such pricing functions with one free parameter, two parameters, and so forth. In the limit, the most general pricing function for this problem has N parameters, where N is the total number of different information goods under consideration. Within each family are many possible functional forms (e.g., piecewise linear, polynomial, etc.), and different profits can be obtained with different functions drawn from the same family.

Therefore, producers of electronic information goods have a daunting challenge: How to explore the space of all possible bundles and price schemes to find the optimal combination? Nonetheless, search over and experimentation with product bundling and price structures are feasible in an agent-mediated economy, due to lower transaction costs. Our goals in this paper are thus two-fold: to learn something about designing economically-intelligent agents, and to learn about the consequences of interactions between today's not-so-economically-intelligent agents as they search for the best bundle/price niche.

It is generally true that a producer which has more free parameters to control in pricing will be more profitable. What then is there to learn? Why not always use unrestricted nonlinear pricing (a different, unconstrained price for every bundle size)? It turns out that optimal pricing under more complex schemes requires more knowledge about consumer preferences than is required for simple pricing schemes. Learning about consumer preferences takes time; meanwhile, the firm is earning less than the optimal profit. Furthermore, a complex price schedule may be more difficult to explain to consumers and more difficult for consumers to evaluate so as to determine their best response to these prices. If these costs are relatively high compared to the additional profit the complex scheme could potentially provide, then the monopolist is likely to settle on a simpler pricing strategy.

Uncertainty about consumer preferences affects many producer decisions. Early papers in the economics literature studied how agents optimally choose between com-

peting opportunities of unknown reward, often referred to as multi-armed bandit problems [Ber72, Wei79]. Agents weigh the tradeoff of gaining information by experimenting versus the cost of experimentation (such as foregone short-run profits). The tradeoff between exploitation and exploration and the problem of how to determine the optimal sequence of actions over a period of time is a primary focus of the reinforcement learning literature. [SB98] provides an excellent overview of this area. Related to our paper, some authors have studied how a firm chooses a one-parameter linear price when it faces uncertain consumer preferences [Rot74]. One author shows that with incomplete learning, the optimal linear price may never be reached [GLS84]. We have looked at an information economy environment in which there is some simple learning by both producers and consumers [JOKMMng].

There is also an extensive economic literature on how a firm can use multi-parameter pricing schedules to extract greater surplus when the distribution of consumers is known, but individual identities are not (or the firm is not allowed to tailor individual-dependent prices). See [Wil93] for a thorough overview. [MR84] present a method for deriving the most profitable unconstrained nonlinear pricing scheme when consumers are differentiated by a single taste parameter. Most papers on related topics assume that the distribution of consumer valuations are known by the firm. One exception studied the tradeoffs between maximizing current profits (exploitation) and charging lower prices in early periods to learn more about the consumer population [BO94].

Multiagent learning has become a popular research topic; examples of recent research include the works collected in [Sen96] and [WS96]. [VD98a] examines the problem of modeling other agents and discusses conditions under which this sort of modeling is useful.

In this paper we use analytic methods to derive optimal prices under pricing schemes of varying complexity for a model with complete information. We measure the increase in profits as more parameters are controlled by the monopolist. We show that the majority of the gains take place as we move from 1 to 2 parameters. Simulations are used to explore a dynamic model in which the monopolist is uncertain about consumer valuations and thus learns the optimal prices gradually and perhaps imperfectly. The analytical solutions provide a benchmark for the maximum profits that could be attained by the firm in steady-state. These simulations provide a means of measuring the costs of a more complex scheme. As the complexity of a pricing schedule increases, it takes longer to learn, but in some cases, particularly that of two-part tariff, the transitional profits are still similar to those of the simpler pricing schemes, due to the shape of the profit landscape. We also see that mixed bundling tends to perform more poorly than two-part tariff, despite their identical steady-state profits. We find that the choice of learning methods, and the choice of initial conditions, strongly affect the speed of convergence, but more importantly for our purpose, affect the magnitude of foregone profits during the learning period.

2 Agent Behavior With Complete Information

In this section we present our model and analyze consumer and producer behavior when there is no value from learning. This allows us to present an hierarchy of complex price schedules and to separate their differing abilities to produce profits from the costs and benefits of learning. In Section 3 we introduce an opportunity to learn, and then study the interaction between learning and pricing choices.

In our model a monopolist produces N items (articles) in each period. Consumers have identical *a priori* subjective distributions of beliefs over article values. The marginal cost of duplicating and delivering any item is zero. The offered price schedule is denoted by a function $T(Q)$ specifying a payment T for a set Q of delivered articles. Consumers may not read all articles that are delivered (e.g., when they receive a subscription or *bundle*).

2.1 Consumers

For reasons discussed in Section 2.2, we assume all consumers face an identical price schedule. Each consumer chooses to receive a set Q_i that maximizes her value net of the payment $T(Q_i)$.

To make consumer agent behavior more concrete, we follow a convenient model of consumer preferences for information goods proposed in [CS99]. When N items are offered, a consumer has a strictly positive value for only a proportion $k_i < 1$. Suppose these positively valued items have values distributed uniformly from 0 to w_i . Consumer i ranks these items from $n = 1$ to $k_i N$ with $n = 1$ being her most highly valued good. Then, (in expectation) the value of the n th best article is given by

$$\mu(n, k_i, w_i) = w_i \left(1 - \frac{n}{k_i N}\right) \quad (1)$$

For convenience, we assume hereafter that consumers value articles exactly at their expected values, given by (1). *Consumer i 's surplus* from reading the n^* most preferred articles is the aggregate value less any payments made to the producer, $\sum_{j=1}^{n^*} \mu(j, k_i, w_i) - T(Q_i^*)$, where $Q_i^* = \{j_i | j_i \leq n^*\}$ for the article indices j_i defined over the list reordered by descending value for consumer i .

For this paper we further limit consumer heterogeneity by assuming that the value of the most preferred article for each consumer (which will generally be different articles) is the same, so $w_i = w \quad \forall i$. Consumers differ through their k_i , which are distributed uniformly between 0 and \bar{k} . The probability density of article values is thus $f(k) = 1/\bar{k}$. Someone with a higher k_i values a greater portion of the available items and also values each equally ranked item at a higher level than someone with a lower k_i . To simplify the analysis in Section 2.2 we assume quantity is a continuous choice variable for the consumer.¹ Our simulations in Section 3 respect the integer constraint.

¹We have calculated the exact solution for two of our price schedules and found that the continuous approximation was not very good for a market with ten articles, but was quite close for $N = 100$ articles.

For these consumer preferences, the socially efficient outcome would be for each person to consume $k_i N$ items. This would yield a surplus of $\frac{wk_i N}{2}$ to each person for a total value of $\frac{w\bar{k}N}{4}$ for the entire population. A firm that could observe each consumer’s k_i could perfectly price discriminate by making a take-it-or-leave-it offer tailored to each individual and extract this entire surplus. This serves as a baseline for the maximum profit that a monopolist could earn.

We elsewhere studied the dynamics of an agent market in which consumers are incompletely informed and thus try to learn the parameters of the article value distribution [JOKMMng]. In future work we will introduce multiple producers, so that consumers dealing with one producer may be able to learn, at a cost, whether other producers have better offerings.

2.2 Producers

We assume that producers cannot track individual consumers across transactions, and further that the producer is either unable to observe the number of articles read by a consumer, or to gain any advantage by using such information. We choose these somewhat idiosyncratic assumptions so that an informed producer can do no better than offer all consumers the same price function, rather than delve into the complexities of price discrimination. In addition, in this paper we wanted to limit producers to “model-free” learning; that is, trying to learn profitable strategies without an explicit model of consumer preferences. There are several motivations for this. A producer will generally not know the true model generating consumer preferences, and might find it too expensive or error-prone to try to estimate the model. Or the environment might be changing sufficiently quickly (through consumer exit and entry, preference changes, and shifts in the pricing strategies of *other* producers) that the producer is never able to learn enough about the form of preferences to be useful. Our assumptions limit producers to trying to learn the shape, or even just the peak, of the profit landscape that derives from consumer preferences, without discovering the underlying preferences.²

For now we assume that producers know the distribution of consumers’ k_i ’s.³ There is nothing to learn, and both preferences and production costs are independent over time, so the problem simplifies to a once and for all determination of the profit maximizing price schedule. The same schedule will be offered in all subsequent periods.

Since consumers are anonymous and they believe the value of each article is drawn from an identical distribution, the producer’s optimal behavior is to base prices solely on the number of articles purchased, q , for a price schedule $T(q)$. The most general form of this schedule would be to set an independent payment T for every possible

²We plan to explore model-based learning in subsequent work. First we will need to make the environment more complex. Our stationary environment with linear preferences permits a smart producer to learn everything in just one transaction period if the number of articles consumed is observable. Of course, such powerful learning possibilities are not realistic.

³In Section 3 we use simulation methods to analyze a model with incomplete information, in which the producer does not know \bar{k} , or any individual’s k_i .

quantity, $q = 1, \dots, N$. However, N may be large, and for various reasons producers may find it unprofitable to set such a large number of price schedule parameters. For example, consumers may object to the complexity of the price schedule and may purchase less, possibly even switching to a different producer. Further, when we introduce uncertainty about consumer preferences so that the producer must engage in learning, the dimensionality of the profit landscape will be the same as the dimensionality of the number of pricing parameters; search over an N -dimensional landscape may be too costly in time and resources to justify such a complex price schedule. Therefore, we explore producer pricing behavior when it limits itself to functions expressible with small numbers of parameters. For each price schedule we derive the optimal parameter choices for the producer, and evaluate the resulting profits, consumers' surplus and social welfare (sum of the prior two measures).

2.2.1 One Parameter Pricing Models

The two most familiar pricing models in practice have only one parameter: linear pricing (a constant price per item), and subscription or bundling (one price for the entire bundle).

a) Pure Bundling

Bundling has recently received substantial attention for information goods [BB99, CS99]. The price schedule is:

$$T(q) = \begin{cases} 0 & : q = 0 \\ P_B & : q > 0 \end{cases} \quad (2)$$

Profit is

$$\Pi = P_B \int_{k_s}^{\bar{k}} f(k) dk \quad (3)$$

where $k_s = \frac{2P_B}{wN}$, the lowest k_i that will subscribe. For lower k_i , the value of the bundle of N items ($\int_0^{k_i N} \mu(n, k_i) dn$) is not enough to cover the price of the bundle. k_s is defined by $\int_0^{k_s N} \mu(n, k_s) dn = P_B$.

Taking the derivative of profit with respect to P_B and setting it equal to zero yields the profit-maximizing bundle price of $\frac{w\bar{k}N}{4}$, $k_s^* = \frac{\bar{k}}{2}$. Profit equals $\frac{w\bar{k}N}{8}$. Consumer surplus and social welfare are $\frac{w\bar{k}N}{16}$ and $\frac{3w\bar{k}N}{16}$ respectively.

b) Linear Pricing

An alternative one-parameter model is linear pricing. Consumers pay P_A for each item they choose to receive. Thus, the price schedule is:

$$T(q) = P_A q \quad : \quad q \geq 0 \quad (4)$$

Facing this marginal price of P_A , each person will choose their optimal number of articles (n_i^*) so that $\mu(n_i^*, k_i) = P_A$. Then, $n_i^*(k_i) = k_i N (1 - \frac{P_A}{w})$. Profit to the monopolist is:

$$\Pi = P_A \int_0^{\bar{k}} k N (1 - \frac{P_A}{w}) f(k) dk \quad (5)$$

Maximizing profit yields $P_A^* = \frac{w}{2}$. Consumer surplus and social welfare under linear pricing are $\frac{w\bar{k}N}{16}$ and $\frac{3w\bar{k}N}{16}$ respectively.

Profit equals $\frac{w\bar{k}N}{8}$, which is the same as for pure bundling. However not all pricing schemes of the same complexity yield the same profit. For example, if k_i were uniformly distributed between \underline{k} and \bar{k} (where $\underline{k} > 0$) then pure bundling would attain higher profit than linear pricing.

2.2.2 Two Parameter Pricing Models

a) Two-Part Tariffs

Under this pricing scheme, consumers pay a subscription fee (F), and a per article price (P_A) [Oi71]. Thus, the price schedule they face is:

$$T(q) = \begin{cases} 0 & : q = 0 \\ F + P_A q & : q > 0 \end{cases} \quad (6)$$

Profit will be:

$$\Pi = F \int_{k_s}^{\bar{k}} f(k) dk + P_A \int_{k_s}^{\bar{k}} k N \left(1 - \frac{P_A}{w}\right) f(k) dk \quad (7)$$

where $k_s = \frac{2Fw}{N(w-P_A)^2}$. The profit-maximizing subscription fee is $F = \frac{2w\bar{k}N}{27}$ and the per-article charge is $P_A = \frac{w}{3}$. $k_s^* = \frac{\bar{k}}{3}$, and profit is $\frac{4w\bar{k}N}{27}$. Note that this profit is 18.5 % higher than the profits from the one-parameter pricing schedules we analyzed above.

b) Mixed Bundling

With mixed bundling, consumers can choose to buy individual items at P_A each, or N items for P_B . Separate purchase is preferred by consumers who want fewer than P_B/P_A articles, otherwise bundling is preferred. The price schedule is:

$$T(q) = \begin{cases} 0 & : q = 0 \\ \min[P_A q, P_B] & : q > 0 \end{cases} \quad (8)$$

Profit will be:

$$\Pi = P_A \int_0^{k_c} N \left(1 - \frac{P_A}{w}\right) k f(k) dk + P_B \int_{k_c}^{\bar{k}} f(k) d(k) \quad (9)$$

where $k_c = \frac{2P_B}{N[2P_A - \frac{P_A^2}{w}]}$. k_c is defined as the k_i for which the consumer surplus attained from buying the bundle is exactly equal to the consumer surplus from buying articles separately instead. Individuals with $k_i > k_c$ will buy the bundle. Everyone else will purchase (fewer) articles on an individual basis.

Profits are maximized if $p_A^* = \frac{2w}{3}$, $P_B^* = \frac{8w\bar{k}n}{27}$, and $k_c^* = \frac{2\bar{k}}{3}$. Profit will be $\frac{4w\bar{k}N}{27}$. Notice that this is the same as the profit attained under two-part tariffs.

2.2.3 N Parameter Pricing Model: General Non-Linear Pricing

The most general pricing strategy is for the producer to choose a price for each possible quantity, without restriction (except that $T(0) = 0$, since the consumer always has the option of exiting the market. Maskin and Riley provide a method for calculating the optimal continuous non-linear pricing strategy [MR84]. The intuition is this: a profit-maximizing producer wants each different customer type (k_i) to purchase a “constrained” quantity $q_i(p_i)$ at price $T(q_i)$. Let some customer type k_i choose to buy q_1 and pay $T(q_1)$. Call the value this consumer would get from an incremental unit Δ_1 . Now suppose there is another customer with $k_2 > k_1$, who also chooses to buy q_1 . By definition, this customer gets value from the next unit $\Delta_2 > \Delta_1$. Therefore, the producer would make a greater profit by offering $q_1 + 1$ at $T(q_1) + \Delta$ where $\Delta_1 < \Delta < \Delta_2$: the second customer would prefer to purchase $q_1 + 1$ but the first customer would not. Thus the problem can be stated as one in which the producer sets prices to induce each customer to purchase its profit-maximizing quantity, subject to the self-selection constraints that ensure a consumer of type k_1 dose not want to purchase the intended quality for any other $k_i \neq k_1$. The function is found from solving a system of differential equations. To conserve space, only the results are provided here.⁴

We obtain the following optimal unrestricted, nonlinear price schedule:

$$T^*(q) = \frac{w \left[\gamma \left(4\bar{k}N + (\bar{k} - 1)q - \lambda(2\bar{k}^3 N + \gamma)\sqrt{4\bar{k}^3 N + \gamma} \right) - 4\bar{k}^4 N^3 \ln \frac{q + \bar{k}(2\bar{k}N + (\bar{k} - 2)q) + \lambda\sqrt{4\bar{k}N + \gamma}}{\frac{N}{2}} \right]}{4(\bar{k} - 1)^3 \bar{k}N} \quad (10)$$

where

$$\gamma = q(\bar{k} - 1)^3 \quad (11)$$

$$\lambda = \sqrt{q}(\bar{k} - 1). \quad (12)$$

This yields profit, welfare and consumer surplus of

$$\Pi_{NL} = \frac{\bar{k}^2 N w (2 - 2\bar{k} - (1 + \bar{k}) \ln \frac{1}{\bar{k}})}{(\bar{k} - 1)^3} \quad (13)$$

$$SW_{NL} = \frac{\bar{k}^2 N w (11\bar{k} - 8 - 4\bar{k}^2 + \bar{k}^3 + 2(2 + \bar{k}) \ln \frac{1}{\bar{k}})}{4(\bar{k} - 1)^4} \quad (14)$$

$$CS_{NL} = SW_{NL} - \Pi_{NL} \quad (15)$$

For example, when $\bar{k} = .7$ profit with non-linear pricing is 32% higher than either of the one-parameter models.

⁴The details of the derivations are available from the authors on request.

Pricing Scheme	Number of Parameters	Profit	CS	Welfare
Pure Bundling	1	.875	.4375	1.3125
Linear Pricing	1	.875	.4375	1.3125
Two Part Tariffs	2	1.037	.449	1.486
Mixed Bundling	2	1.037	.281	1.318
Block Pricing	3	1.094	.328	1.422
Non-Linear Pricing	N	1.152	.216	1.368
Perfect Price Discrimination		1.75	0	1.75

Table 1: Example of one-period complete information results: Optimal profit per good with $w = 10$, $N = 10$, and $\bar{k} = .7$

2.2.4 Summary of Results for Static Model

We calculate a numerical example to illustrate the relationships between the various pricing strategies. Suppose that $w = 10$, $N = 10$, and $\bar{k} = .7$. In Table 1 we report the resulting profits, consumers' surplus, and social welfare.⁵ We report in the last row the results if the producer knew how much each consumer valued each specific article, rather than just the values as a function of the number of articles. In this case of *perfect price discrimination* the producer would charge each consumer a different price for each article (thus the price schedule would have the number of consumers times N parameters); this provides a benchmark for the maximal profit possible if the producer could obtain perfect information.

3 Incomplete Information: Simulations of Agent Behavior

The analytical results we have just presented identify which price schedule a producer should choose given that it has complete information about the distribution of consumer preferences. In this setting, a profit-maximizing producer will set prices to maximally *exploit* consumers. In real life, however, a producer seldom knows everything about consumer preferences. With incomplete information, each period of pricing and consumer purchasing may reveal information that enables the producer to update and improve his estimate of the consumer preference distribution. Generally, the producer's one-period profits will be higher the more accurate is his estimate.

Therefore, pricing decisions now serve two functions: exploitation and *exploration*. In general, the prices that extract maximal expected profit from consumer in the current period (given on the producer's current beliefs) will not provide the maximal improvement in the estimate of the preference distribution for use in future pricing decisions. Thus, there will be a tradeoff between exploitation and exploration. For this reason, we now explore how producers might learn about preferences through a dynamic sequence of price-purchase interactions with consumers. In particular, we

⁵We include results for *block* pricing, an intermediate case with three parameters, though we did not include the derivation above to conserve space.

examine how much profit is accumulated over time when the producer follows different learning methods on price schedules of varying complexity. Different exploitation-exploration strategies will tend to converge at different rates, but also will yield different accumulations of profit during the learning phase.

One important consideration is how much the producer knows about the *structure* of preferences. In at least some situations, knowledge of the true preference model can be used for better and faster learning. For example, we can show that if the producer knows consumer preferences are generated according to Chuang and Sirbu [CS99], then it need take only one period of purchasing to discover the true parameters of the preference distribution, and thereafter the producer can set the optimal prices for exploitation without further investment in exploration.

In practice, a producer will not know the exact functional form of the consumers' preference model(s). In one extreme case, if the producer has no prior knowledge of the structure of preferences (other than that demand is decreasing in price), it will have to adjust the parameters of its pricing schedule based on the apparent correlation between prices and the profit signals it receives.

Whether the producer is trying to learn the structure of consumer preferences (to thereby derive the profit landscape) or to learn the profit landscape (or even just its peak) directly, there are any number of potential learning algorithms. In this paper we experiment with two off-the-shelf learning systems: a neural network trained using Quickprop (a back-propagation method) [Fah88] and a simplex linear programming approach to multi-dimensional hill-climbing known as amoeba [WHP92]. In a competitive environment, of course, a producer might find it advantageous to develop a learning algorithm customized to the problem at hand. Our goal is not to discover the best learning algorithm, but to understand the tradeoffs between learning and exploitation that occur across range of pricing structures, and to compare these results between two common algorithms.

However, learning can be confounded if the consumer population evolves (through exit and/or entry), or if consumer preferences change, especially if a producer is only able to sample the consumer population infrequently (for example, based on a journal that is issued quarterly). In such "moving target" multiagent learning problems [VD98b, HW98], some amount of residual error in what is learned is unavoidable. For such problems, we expect that strategies that learn faster, albeit ultimately less accurately or completely, may be favored. Likewise, strategies that perform well despite errors in the estimation of preferences should be favored. We hypothesize that pricing schedules with fewer parameters can generally be learned more quickly. It is less clear whether there is a systematic relationship between price schedule complexity and the robustness of its performance *while* learning. To explore these questions, we have designed experiments that assess learning speed and robustness while learning for a variety of learning methods and pricing structures. We measure both the convergence rate and the average profit per agent per period prior to convergence, in order to begin characterizing important design tradeoffs in developing pricing schedules and agent learning capabilities for automated agents representing producers in an information economy.

3.1 Simulation Method

In these experiments, we generated a population of 1000 consumers with Chuang-Sirbu valuations as described above. We constructed a monopolist producer which attempted to learn the optimal parameters for each of the five schedules described above using the two different learning algorithms. The producer had no explicit knowledge of the consumers, such as the items they purchased or the fact that they fit a Chuang-Sirbu valuation model. It merely chose a set of parameters and observed a resulting profit signal.

We make no claim that these learning methods are optimally efficient; it would seem reasonable to assume that gains in performance could be achieved by either introducing explicit knowledge about the construction of the consumer population or by tuning these algorithms for use the specific problems. However, the purpose of these experiments is not to show how quickly a given algorithm can learn a particular schedule, but to show how these schedules behave as they are being learned.

While amoeba is always attempting to hill-climb and therefore has no explicit decision to make about whether it needs to acquire more information, a producer using the neural network must decide whether to use its currently optimal solution or to explore further. This decision was made by exploring with probability $\frac{found}{possible}$, where *found* is the number of samples seen so far and *possible* is the number of possible integer price schedules. The producer could modify this by pruning known parts of the profit landscape; for example, if a set of parameters yielded zero profit, increasing one parameter while fixing the others will also yield zero profit, and so that part of the space need not be explored. This choice of exploration strategy is not necessarily optimal; the point was to choose a simple strategy and hold it constant across the different schedules.

In these experiments, as in the analysis above, k_i was drawn from $U[0, 0.7]$. w_0 was fixed at 10. Experiments were conducted for both $N = 10$ and $N = 100$ goods. The producer was given 200 iterations to attempt to learn the optimal pricing parameters for this static population.

In order to produce a balanced comparison of amoeba and the neural network on the different price schedules, a uniform search space was selected. For each pricing parameter, an upper bound much larger than the optimal value was chosen. For example, in linear pricing with $w_0 = 10$, the upper bound was set at 25, where the optimal value is approximately 6.3. The neural net randomly chose points from this space when exploring, reducing the bounds when areas of zero profit were found. The amoeba algorithm began with a simplex at the origin and at the point specified by the upper bound. This provided both algorithms with approximately equivalent search spaces.

3.2 Results: Learning Complexity and Transitional Profits

Figures 1 and 2 show our results for the neural network and the amoeba hill-climber, respectively, on the five price schedules when $N = 10$. Each line, for a different price schedule, indicates the average cumulative per-period (per-article, per-customer)

profit for that schedule. For example, the value of about 0.65 for mixed bundling at iteration 10 in Figure 1 mean that over the first 10 iterations, profit averaged 0.65 (so cumulative profit is approximately 6.5). Higher lines indicate schedules which have been more profitable to date.

With neural net learning, mixed bundling performs extremely well. It has the highest profits of any schedule for about 100 periods, and thereafter is close to the leader, two-part tariff. Contrary to our expectations, the one-parameter schedules didn't perform particularly well, even during the learning phase. Their initial explorations lead to a substantial profit drop for 20 periods; after period 10, either or both of the two-parameter schemes nearly always dominate. Nonlinear pricing, which has the highest profit potential, performs very badly, getting stuck at a unit profit of about 0.6, only about 40% of the maximum possible. Note that the rate of convergence is not a good measure for comparing price schedules, since it does not reflect the quality of the final solution. For example, nonlinear pricing converges quickly to a local optimum, and thereafter obtains very poor profits.

The amoeba results are notably different. The one-parameter schemes (linear pricing and pure bundling) perform quite well for the first 30-40 iterations. Then the two-parameter schemes overtake the simpler schemes and head to a much higher average profit level. Amoeba also does much better with nonlinear pricing. As predicted, learning for the more complicated scheme is slower, and its cumulative profits are much lower than the other schemes for the first 400-1000 periods, but once the algorithm begins to converge on a solution its performance is swift and durable. Since amoeba converges more quickly and finds peaks that are closer to the global optimum than the neural network does, its results seem more useful for this experiment. The qualitative results are clear with amoeba: simpler schemes with lower profit potential nevertheless outperform more complex schedules during the early stages of learning.

The experimental results for the neural network and amoeba when $N = 100$ are shown in Figures 3 and 4. Increasing the number of goods from 10 to 100 helps to smooth out the profit landscape and actually improved the solution quality for a number of the schedules. For example, as the number of goods increases, the two-part tariff landscape is smoothed out. Also, it contains an easy-to-climb hill, and many non-optimal values have relatively high profit, making exploration less costly. Additionally, there are no large chasms, making it easy to move between optima. Contrast this with mixed bundling, in which two large valleys separate the optimal solution from the solutions for linear pricing and pure bundling. Because of this, even though these two-parameter schedules have the same potential optimal profit, the two-part tariff performs better on average after learning is complete⁶.

For the most part, increasing the number of articles that may be transacted reinforces the results discussed above. Amoeba learning again finds higher profit levels, and the qualitative results are consistent across schedules.

In Figure 4, the one-parameter schedules are much more profitable during the learning phase, but after about fifty periods the more complex schedules dominate.⁷

⁶Recall that the plots show results averaged over several runs with randomized starting values. The global peak of the mixed bundling landscape was found less often than that of two-part tariff.

⁷The unconstrained nonlinear schedule has $N = 100$ free parameters now; amoeba was not able

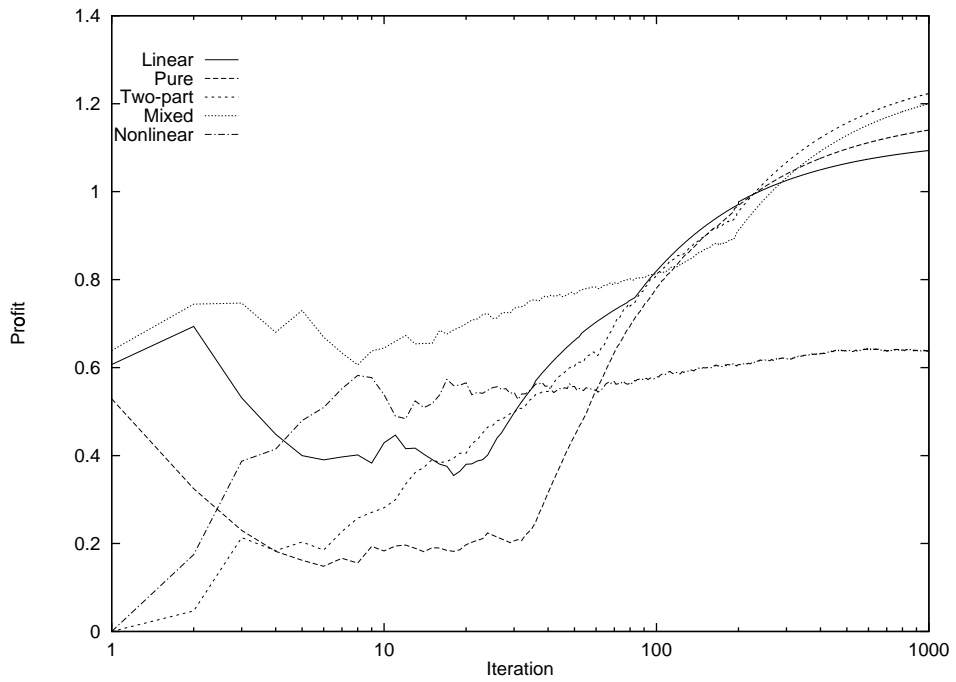


Figure 1: Cumulative profit per article, per period, per customer ($N = 10$, neural net)

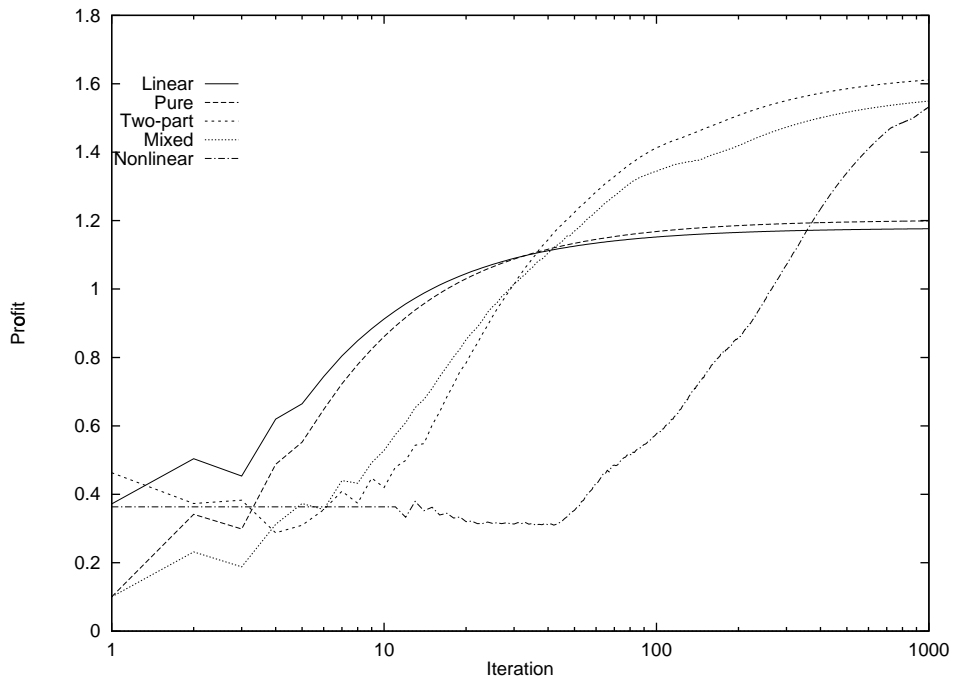


Figure 2: Cumulative profit per article, per period, per customer ($N = 10$, amoeba)

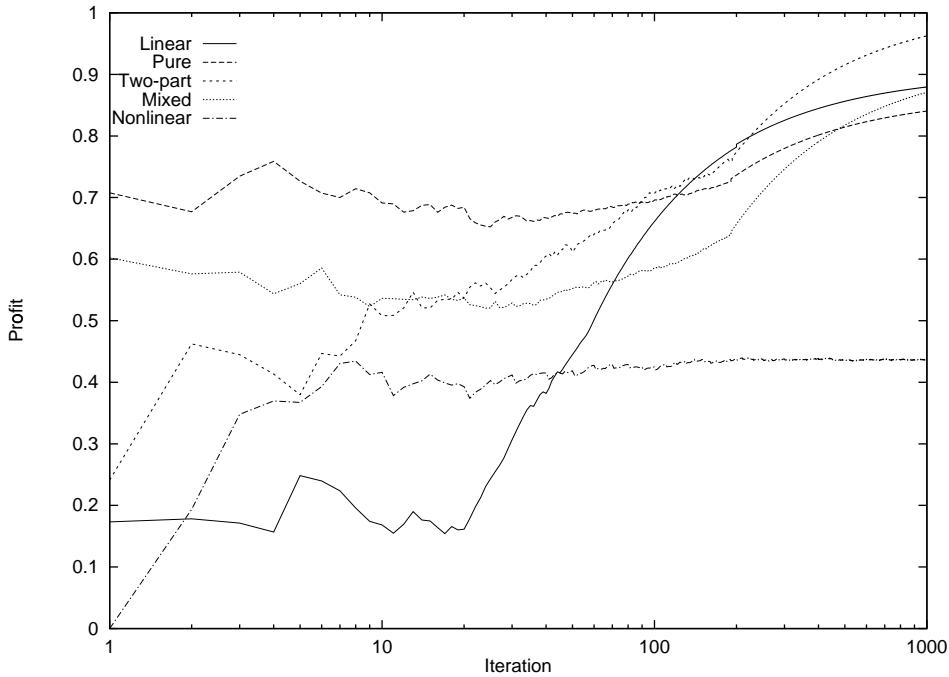


Figure 3: Cumulative profit per article, per period, per customer ($N = 10$, neural net)

The pattern is less systematic for neural net learning, but the results are qualitatively similar.

Note that the learner was able to find more optimal solutions for each of the schedules; this is presumably due to the flattening of discontinuities in the profit landscape.

3.3 Summary of Simulation Results

The amoeba learner experiments support our predictions: simpler pricing schemes perform better during the phase of high exploration; when learning is nearly complete, exploitation dominates and the higher-dimensional schemes perform better. However, the learning phase may be quite long, depending on the complexity of the schedule and the frequency of observable transactions. In particular, exploration is slow for the most general scheme, nonlinear pricing.

These results suggest that if an environment is changing significantly at moderate frequencies, a producer might be better off using a simpler scheme which has lower potential profit but is more robust to uncertainty. The results also suggest that more sophisticated agents might try to use statistical modeling to endogenously switch between simpler and more complex schemes, depending on indicators of stability and their progress in learning the current schedule. Although we have not modeled this phenomenon yet, the idea is suggestive of the variety of special promotions and

to search the resulting 100-dimensional space in an acceptable amount of time.

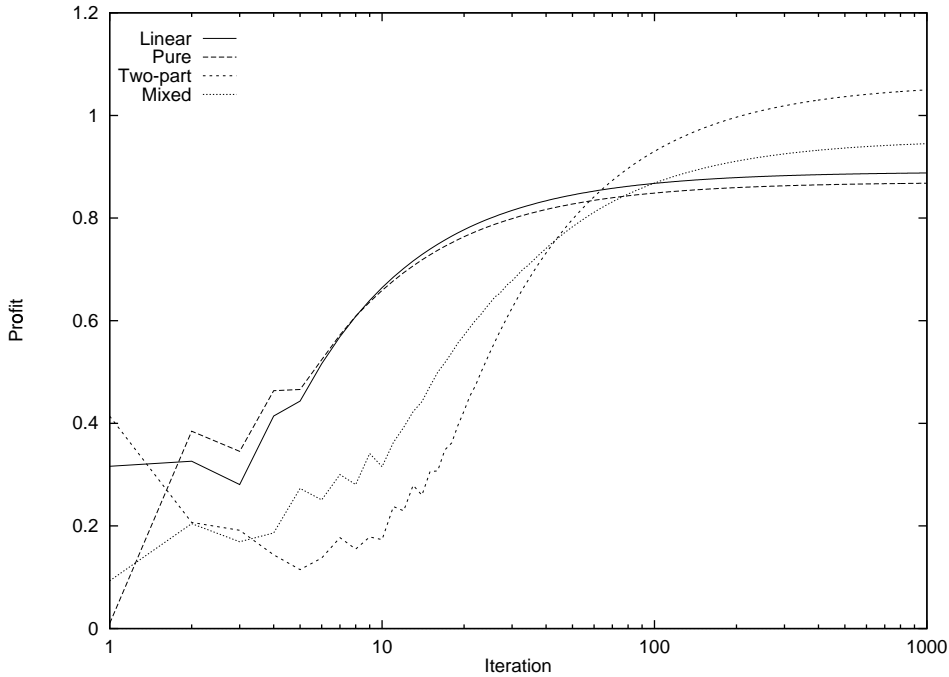


Figure 4: Cumulative profit per article, per period, per customer ($N = 100$, amoeba)

trial prices that are used when new products are introduced to speed up exploration without sacrificing too much profit.

Our results also indicate the sensitivity of performance to the choice of learning method. We discuss this more fully in the next section, noting now only that in these experiments amoeba typically found better results than the neural net. Further, although it is not directly obvious from the plots, we discovered that with rather craggy profit landscapes (especially for $N = 10$), the results are rather sensitive to the choice of initial conditions. This suggests that there is some value to incorporating more aggressive nonlocal exploration when a candidate solution seems to have been found.

One of the more interesting and less expected results is the performance gap between schedules with the same dimensionality and the same potential profit. In an environment which contains uncertainty and learning, the smoothness of the profit landscape is an important determinant of effective complexity; it is harder to get consistently good results on a craggy landscape; cf. two-part tariffs vs. mixed bundling in Figures 5 and 6.

A number of recent papers have ignored uncertainty regarding the distribution of consumer preferences and advocate forms of mixed bundling for its exploitative performance; our results suggest that when exploration is important, two-part tariffs (and perhaps other two or more parameter schedules with smooth landscapes) may be superior.

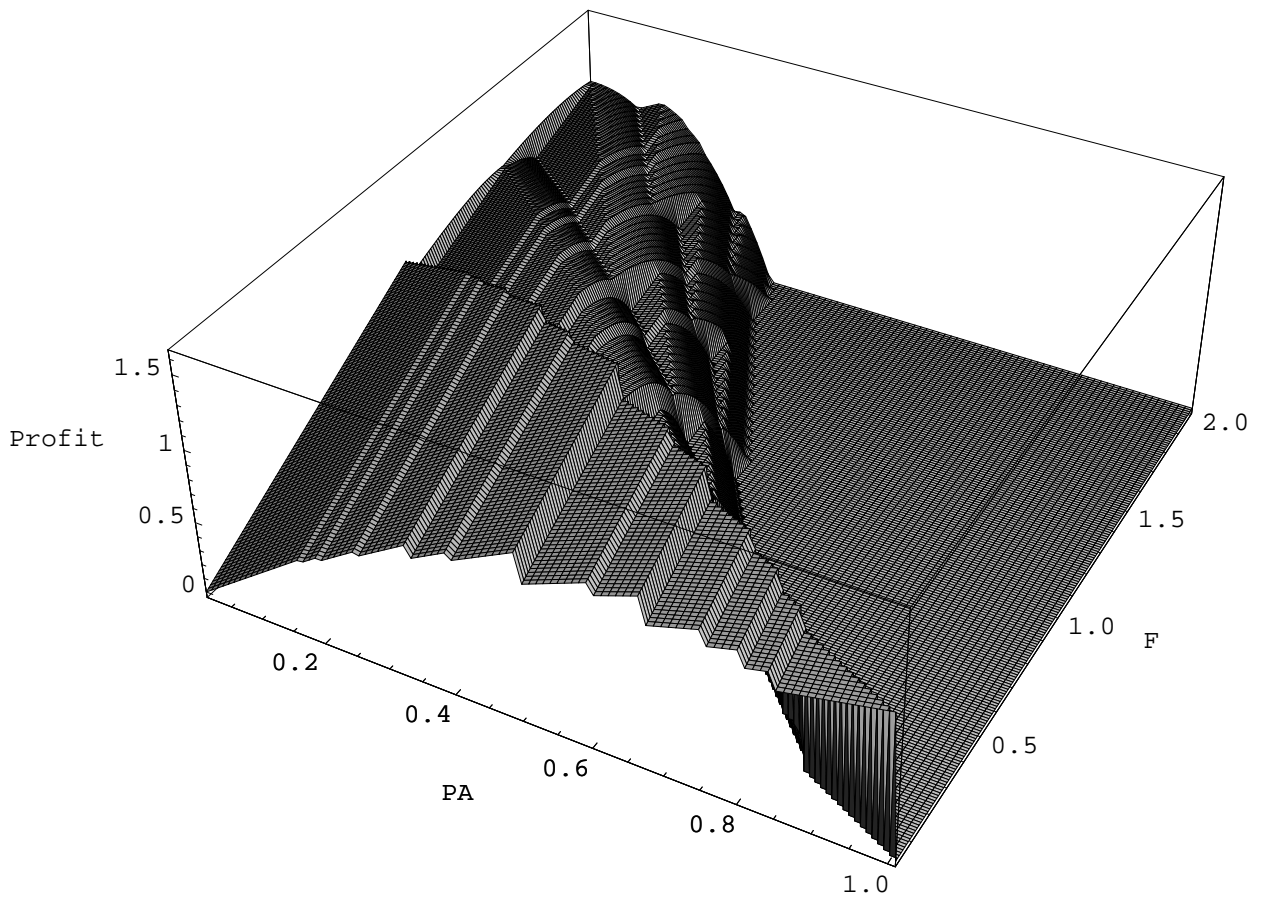


Figure 5: Exact profit landscape with $N = 10$ and two-part tariff pricing

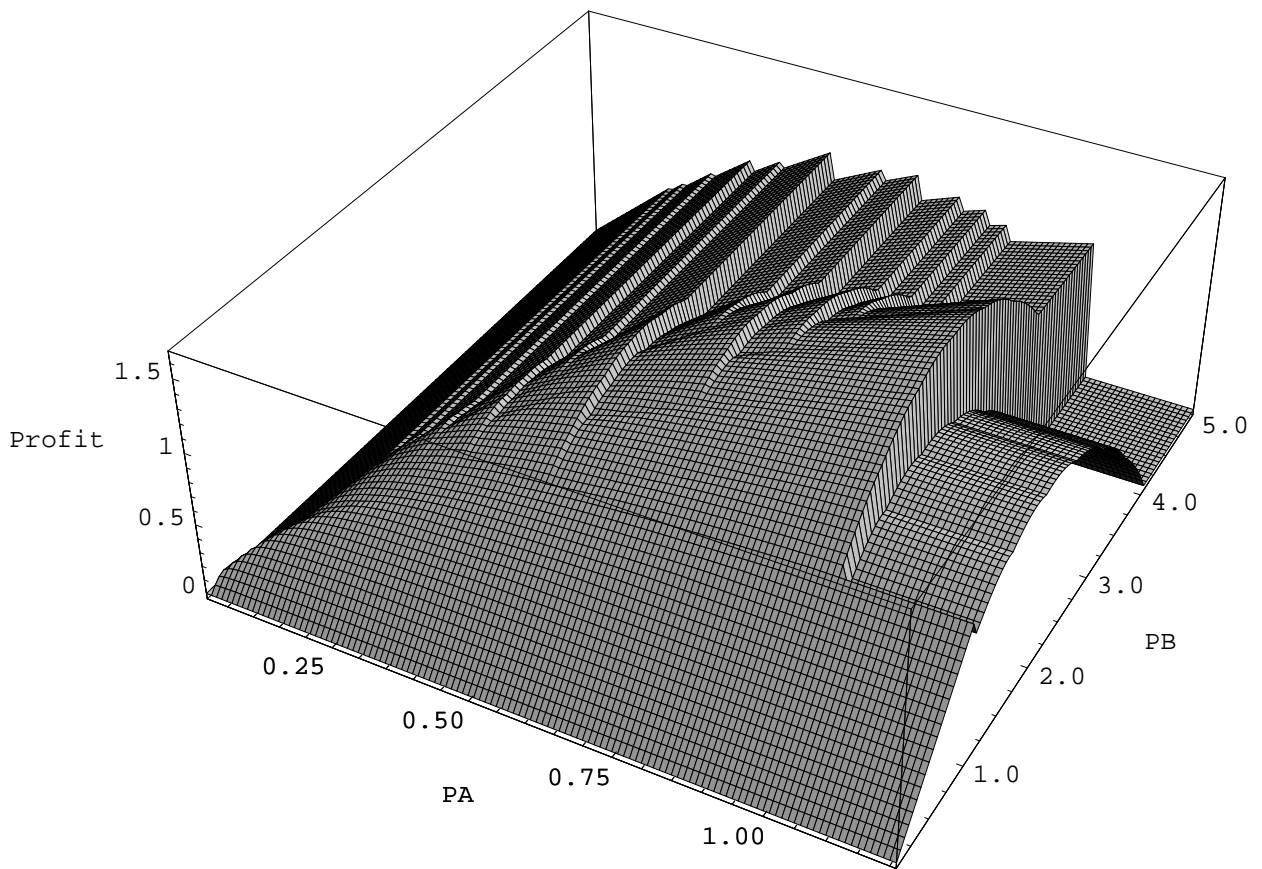


Figure 6: Exact profit landscape with $N = 10$ and mixed bundling pricing

Schedule	% of optimal reached (NN)	optimal profit reached (NN)	optimal profit possible (NN)	% of optimal reached (amoeba)	optimal profit reached (amoeba)	optimal profit possible (amoeba)
Linear	97.4	1.1228	1.1598	99.7	1.1763	1.1813
Pure	99.6	1.183	1.1886	99.9	1.2028	1.2035
Two-part	89.0	1.2638	1.4064	99.6	1.6374	1.6429
Mixed	84.0	1.2719	1.5124	96.8	1.5823	1.6344
Nonlinear	56.2	0.9643	1.7172	98	1.7634	1.7885

Table 2: A comparison of the performance of neural net (NN) and amoeba over the different pricing schedules for $N = 10$

4 Learning approaches in dynamic electronic commerce environments

The two optimization techniques studied in this section are instances of two fundamentally different approaches: model-based (exemplified by the neural net) and model-free (exemplified by amoeba).

Note that, in this section, we take the term “model-based” to mean that the learner constructs an approximation of the functional relationship between pricing parameters and profits. As in the rest of the paper, the learner makes no attempt to represent consumer preferences directly. In contrast, a “model-free” technique is one which alters pricing parameters so as to maximize profit without retaining any information about the structure of the function being optimized.

Model-based techniques such as that used by the neural net comprise two steps. First, a model of the entire profit landscape is estimated from an observed series of price schedules and resultant profits. Second, this model landscape is used to predict expected profits for a large number of proposed price schedules, and the schedule for which the model predicts the largest expected profit is chosen as the (approximately) optimal one. In general, a model-based approach would appear to have the advantage that optimization can be done in “virtual” time, i.e. once the model has been learned, the method can explore a vast number of possible price schedules without risking the loss of real time or real money. On the other hand, a large number of data samples may be required to learn the model in the first place, so overall the method may be expensive and slow. Furthermore, model-based optimization is likely to fare poorly if the learned model is insufficiently accurate in the vicinity of the peak — a problem to which the technique is vulnerable because it strives for a good global fit to the entire landscape rather than a good localized fit to the peak. This study provides some evidence for both of these effects: relative to the model-free amoeba method, the neural net took longer to reach a plateau, and this plateau was generally less optimal than what was attained by amoeba, as is seen in Table 2.

Also compare the craggy landscape shown in Figure 5 to the estimated landscape produced by the neural net in Figure 7. In Figure 7, we can see that the peaks and fissures that are a part of the two-part tariff price schedule have been replaced within

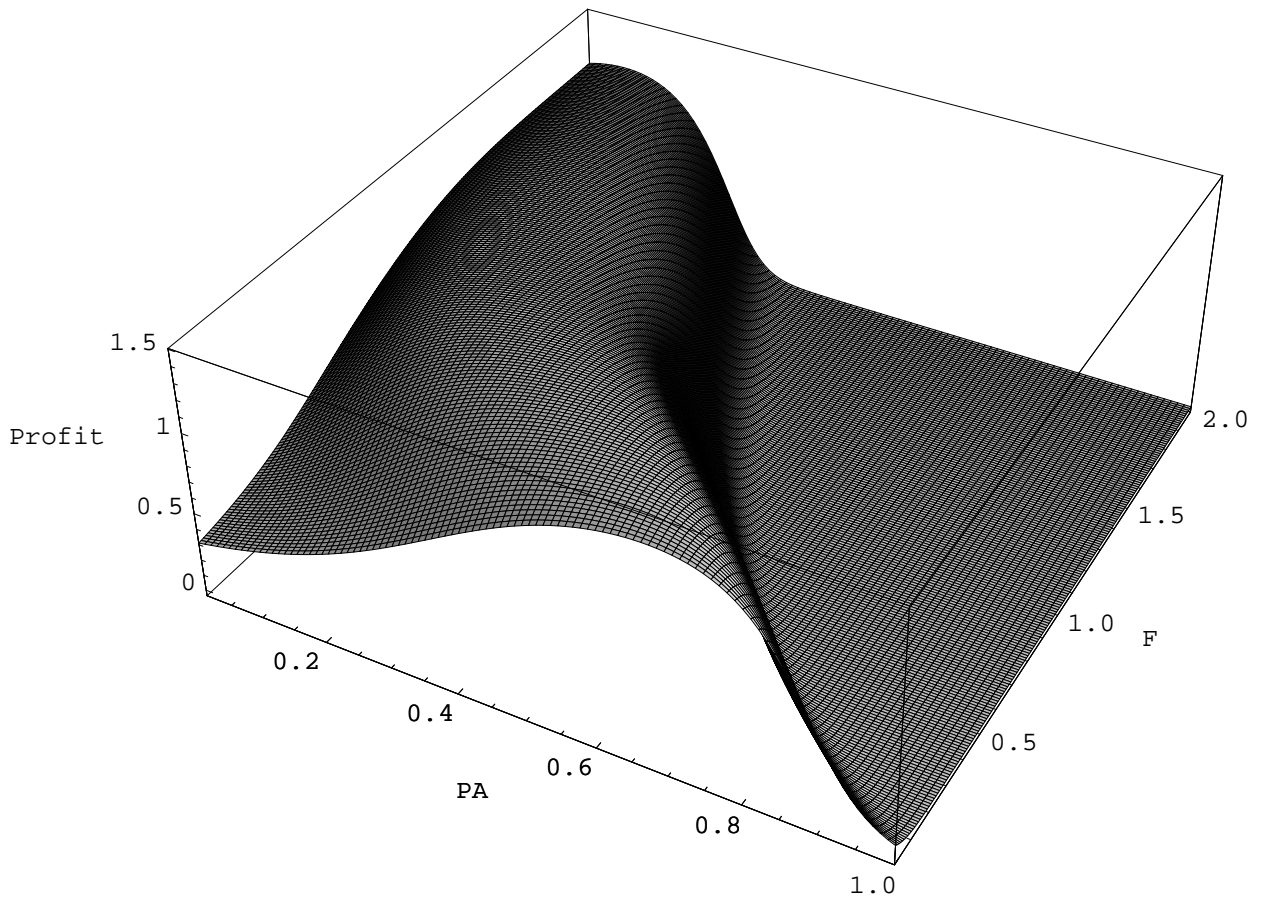


Figure 7: Neural net-learned profit landscape with $N = 10$ and two-part tariff pricing

the neural net by an overly smooth landscape, leading to a loss of some optimal solutions.

In contrast, model-free approaches such as amoeba make no attempt to establish a functional relationship between prices and profits. Amoeba single-mindedly seeks the optimal price schedule without retaining any information about the portions of the profit landscape that lie beyond the borders of its ever-shrinking simplex. Ultimately, amoeba comes to rest at or near a peak (almost always the global peak in these experiments), but knows nothing of its environs. In the problem studied here, this ignorance is not a liability because no other force can dislodge it from the peak: the landscape is static since consumers are fixed in their preferences and their behavior and there are no competitors which can shift the producer to a different point in the landscape.

While amoeba appears to outperform the neural network for this specific problem, it would be premature to conclude that model-based techniques are inferior to model-free ones. There are two basic reasons why further exploration of model-based approaches ought to be encouraged.

First, the neural network is only one of a great variety of model-based techniques, and has not been tuned for this particular problem. Training methods other than Quickprop might allow faster training with less data. Off-the-shelf Quickprop is very generic, and what we have been referring to as the “model” it learns is nothing more than function approximation using sigmoidal basis functions. These might be replaced with basis functions that are more appropriate to the application at hand. Also, there is no *a priori* reason why the algorithm must minimize the error over the entire landscape; it could be altered to minimize the error at the landscape’s peaks, at the cost of a loss of resolution of lower parts of the landscape.

A more sophisticated approach to modeling might apply domain knowledge and analysis to formulate a more useful and appropriate model. It might prove advantageous to model the consumer preferences rather than the profit landscape itself, particularly if the number of parameters that define consumer preferences is smaller than the number of parameters defining the price schedule. This requires inverting the profit signal and possibly some auxiliary information such as the distribution of the number of purchased items to estimate the consumer preference distributions.

Learning consumer preferences and then deriving the profit landscape might be faster than directly learning the profit landscape, perhaps due to the lower dimensionality and/or greater smoothness. In preliminary experiments, we have used such an approach to infer the distribution $f(k)$ very quickly and accurately, resulting in an excellent approximation to the profit landscape. After developing an accurate model of the profit landscape with a very small number of real data samples, we applied the amoeba technique to find a near-optimal point in the landscape in *virtual* time, followed by further iterations of amoeba in real time to quickly reach an absolute optimal price schedule. However, deriving a model of the profit landscape from a model of consumer preferences can be risky unless the consumer preferences (or at least their functional form) are characterized faithfully. If, for example, a producer believed that preferences took a negative exponential form, when in fact they took the linear Chuang-Sirbu form, the producer might be able to learn the best param-

eters and pricing structure for the assumed preference distribution quickly, but with exploration could discover that higher profits were possible, and thus might start a more lengthy process of exploration to discover the functional form of preferences, rather than just the correct parameters for the incorrect “known” form. As in the simpler model we studied, there would be a learning period, and a trade-off between exploitation and exploration.

A second reason to encourage further exploration of model-based approaches is that the situation studied here is not truly representative of the complex, dynamic environments in which we expect producer agents to operate, and model-based approaches may prove to be advantageous in such environments. A more global knowledge of the landscape is likely to be helpful in dealing with shifting consumer preferences, and is probably essential in dealing with competitors, who are continually trying to knock one another off of the peak in an endless game of *King of the Hill*.⁸

Finally, it should be noted that model-free and model-based approaches might be complementary. If model-free techniques prove to be generally faster in their convergence, but fragile in the face of competition that drives the producer off of a local peak, then one could quietly train a model of the entire landscape in parallel while the model-free optimizer is being used to seek a good price schedule. Once a reasonable model of the landscape has been developed, the model-free optimizer can start using the model rather than real-world data, resulting in a tremendous optimization speedup.

5 Conclusions and Future Research

We have presented a hierarchy of pricing schedules for the complex space of electronic goods and studied their performance using analytical methods for the steady state and simulations to determine their dynamic behavior. We find that most of the improvement in schedules comes with the move from one-parameter to two-parameter schedules. More complex schedules can provide some additional profit at equilibrium, but it is not clear that this small increase is balanced by the longer time needed to learn these schedules. The simpler schedules are also appealing from a more practical point of view; they are easier for a consumer to understand, and consequently it is easier for the consumer to participate. All of these results are based on the assumption that consumers will always act to maximize their surplus. Our continuing research questions the effect of consumers who are unable to determine how many goods to buy (because the schedule is too complex), or who act strategically to exploit the fact that the producer is exploring.

Another area of future research involves the introduction of multiple producers or changing consumers. In this case, the optimal set of prices will change over time, and both the pricing schedule and the producer’s learning algorithm will need to be

⁸For example, each time a competitor changes its price schedule and attracts a different set of consumers, or consumers switch between producers to do their own exploration, the effective population of consumers (and thus the distribution of preferences and the resulting profit landscape) facing the first producer will change.

robust to these changes.

A third area of future research includes the introduction of consumer modeling into the learning process. If a producer has access to some information about the consumer's valuations, it may be able to learn the optimal set of prices more quickly and therefore increase its profits. The question remains open as to what sorts of consumer knowledge are realistic for a producer to be able to infer, given that real-world producers seem to need large amounts of data to do a reasonable job of modeling consumers.

References

- [BB99] Y. Bakos and E. Brynjolfsson. Bundling information goods: Pricing, profits and efficiency. In D. Hurley, B. Kahin, and H. Varian, editors, *The Economics of Digital Information Goods*. MIT Press, Cambridge, Massachusetts, 1999.
- [Ber72] D.A. Berry. "a bernoulli two armed bandit.". *Annals of Mathematical Statistics*, 43:871–897, 1972.
- [BO94] David J. Braden and Shmuel S. Oren. Nonlinear pricing to produce information. *Marketing Science*, 13(3):310–326, 1994.
- [CS99] J. C. Chuang and M. A. Sirbu. Network delivery of information goods: Optimal pricing of articles and subscriptions. In D. Hurley, B. Kahin, and H. Varian, editors, *The Economics of Digital Information Goods*. MIT Press, Cambridge, Massachusetts, 1999.
- [Fah88] Scott E. Fahlman. An empirical study of learning speed in back-propagation networks. Technical report, Carnegie Mellon University, 1988.
- [GLS84] Barry Goldman, Hayne E. Leland, and David S Sibley. Optimal nonuniform prices. *Review of Economic Studies*, 51:305–319, 1984.
- [HW98] Junling Hu and Michael P. Wellman. Online learning about other agents in dynamic multiagent systems. In *Proceedings of the Second International Conference on Autonomous Agents (Agents '98)*, 1998.
- [JOKMMng] Rajarshi Das Jeffrey O. Kephart and Jeffrey K. MacKie-Mason. Two-sided learning in an agent economy for information bundles. In *Agent-mediated Electronic Commerce*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin, Forthcoming.
- [MR84] Eric Maskin and John Riley. Monopoly with incomplete information. *Rand Journal of Economics*, 15(2):171–196, 1984.

- [Oi71] W. Oi. A disneyland dilemma: Two-part tariffs for a mickey mouse monopoly. *The Quarterly Journal of Economics*, 85(1):77–96, February 1971.
- [Rot74] Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9:185–202, 1974.
- [SB98] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.
- [Sen96] Sandip Sen, editor. *Adaptation, Coevolution and Learning in Multiagent Systems: Papers from the 1996 AAAI Spring Symposium*. AAAI Press, March 1996.
- [VD98a] Jose M. Vidal and Edmund H. Durfee. Learning nested agent models in an information economy. *Journal of Experimental and Theoretical Artificial Intelligence*, 10, 1998.
- [VD98b] Jose M. Vidal and Edmund H. Durfee. The moving target function problem in multi-agent learning. In *Proceedings of the Third Annual Conference on Multi-Agent Systems*, Paris, France, 1998.
- [Wei79] Martin L. Weitzman. Optimal search for the best alternative. *Econometrica*, 47(3):641–654, 1979.
- [WHP92] et al. William H. Press. *Numerical Recipes*. Cambridge University Press, 1992.
- [Wil93] Robert B. Wilson. *Nonlinear Pricing*. Oxford University Press, 1993.
- [WS96] G. Weiss and S. Sen, editors. *Adaptation and Learning in Multiagent Systems*. Springer-Verlag, Berlin, 1996.