This work is distributed as a Discussion Paper by the

STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

SIEPR Discussion Paper No. 06-18

Selling to Overconfident Consumers

By
Michael D.Grubb
Stanford University

November 2006

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

Selling to Overconfident Consumers*

Michael D. Grubb
Graduate School of Business
Stanford University
Stanford, CA 94305
mgrubb@stanford.edu
www.stanford.edu/~mgrubb

November 7, 2006

Abstract

Consumers may overestimate the precision of their demand forecasts. This overconfidence creates an incentive for both monopolists and competitive firms to offer tariffs with included quantities at zero marginal cost, followed by steep marginal charges. This matches observed cell-phone service pricing plans in the US and elsewhere. An alternative explanation with common priors can be ruled out in favor of overconfidence based on observed customer usage patterns for a major US cellular phone service provider. The model can be reinterpreted to explain the use of flat rates and late fees in rental markets, and teaser rates on loans. Nevertheless, firms may benefit from consumers losing their overconfidence.

*A previous version of this paper circulated under the title "Screening Overconfident Consumers" (2005). I am very grateful to Jeremy Bulow, Jonathan Levin, and Andrzej Skrzypacz for many valuable discussions of the issues in the paper and to Katja Seim for help and advice especially in obtaining data. For helpful comments and suggestions, I would also like to thank three anonymous referees, Susan Athey, Lawrence Ausubel, Doug Bernheim, Simon Board, Carlos Corona, Liran Einav, Erik Eyster, David Laibson, Edward Lazear, Peter Lorentzen, Muriel Niederle, Peter Reiss, John Roberts, Illya Segal, Enrique Seira, Steven Tadelis, and Brian Viard. I am thankful for financial support from the Taube Scholarship Fund Fellowship through a grant to the Stanford Institute for Economic Policy Research, and from the State Farm Companies Foundation Doctoral Award.

1 Introduction

Firms commonly offer three-part tariffs, or menus of three-part tariffs, in a variety of contexts. A three-part tariff consists of a fixed fee, an included allowance of units for which marginal price is zero, and a positive marginal price for additional usage beyond the allowance. A prime example is the US cellular phone services market in which firms typically offer consumers a menu of tariffs each consisting of a fixed monthly fee, an allowance of minutes, and an overage rate for minutes beyond the allowance. Pricing of internet service is similar in many European countries, where usage is billed per megabyte (Lambrecht and Skiera 2006). Other examples of three-part tariffs include car leasing contracts, which typically include a mileage allowance and charge per mile thereafter.[1] In a variety of rental markets, contracts often charge a flat rate for a specified period followed by steep late fees. Finally, introductory credit card offers are often three part tariffs. For instance a $1,000 balance might be charged an initial balance transfer fee, zero marginal charge per month for the first six months, and a high marginal charge per month thereafter.

The existing literature on non-linear pricing does not provide a compelling explanation for such pricing patterns. For perfect competition one expects prices to be driven down to cost, while typical non-linear pricing models suggest that we should observe quantity discounts so that the highest demand consumer pays the lowest marginal price. Instead, a tendency of consumers to underestimate the variance of their future demand when choosing a tariff provides a more plausible explanation of observed menus of three-part tariffs. Two important biases lead to this tendency: forecasting overconfidence, which has been well documented in the psychology literature, and projection bias, which is described by Loewenstein, O'Donoghue and Rabin (2003).

Intuitively, underestimating variance of future demand may lead to tariffs of the form observed because consumers do not take into account the risk inherent in the convexity of the tariffs on the menu. This is because although the tariffs have a high average cost per unit for consumers who consume far above or far below their allowance, consumers are overly certain that they will choose a tariff with an allowance that closely matches their consumption. Thus consumers expect to pay a low average price per unit, but sellers profit ex post when consumers make large revisions in either direction.[2] This intuition is illustrated with a simple example in Section 3.

I develop a model of firm pricing when consumers are overconfident. I begin by assuming that consumers are homogenous ex ante, so that there is no screening at the contracting stage and firms

[1] My thanks to an anonymous referee for suggesting this example.

[2] According to a pricing manager at a top US cellular phone service provider, "people absolutely think they know how much they will use and it's pretty surprising how wrong they are."

only offer a single tariff. In this context, I show that prices will be qualitatively similar to three-part tariffs: Given consumer overconfidence, free disposal, and low marginal costs, consumers will be offered a tariff which involves a range of minutes offered at zero marginal price, followed by positive marginal prices for additional units. This result holds not only under monopoly, but is also robust to perfect competition. Furthermore, while overconfidence always reduces total surplus, it may increase consumer welfare and reduce monopoly profits.

I extend the model to allow for ex ante heterogeneity and screening at the contracting stage via a menu of multiple tariffs. In particular, I characterize a monopolist's optimal two-tariff menu given two ex ante types at the same level of generality as the single tariff model, and a monopolist's optimal menu of a continuum of tariffs given a continuum of ex ante types under more restrictive conditions. In both cases, the qualitative pricing results of the single tariff model are robust as long as overconfidence is sufficiently high relative to ex ante heterogeneity. A general characterization of multi-tariff menus given perfect competition is more difficult. However, specific examples of multi-tariff menus under perfect competition illustrate the same qualitative tariff features as in the single-tariff model.

I consider several potential alternative explanations for three-part tariff pricing. First, I consider the flat-rate tariff bias, which encompasses demand overestimation, risk aversion, and the taxi-meter effect (Lambrecht and Skiera 2006). Second, I consider demand underestimation, which is related to quasi-hyperbolic discounting. Third, I develop a monopoly price discrimination explanation which is closely related to Courty and Li (2000). Although the first two potential alternatives may have important effects on pricing, neither can explain three-part tariff offerings. However, the monopoly price discrimination model does predict three-part tariff pricing given the right type distribution. Since the overconfidence and price discrimination models cannot be distinguished based on observed (monopoly) prices, I compare the two explanations using both observed prices and tariff and quantity choices in a particular setting: cellular phone services.

I have obtained billing records for 2,332 student accounts managed by a major US university for a national US cellular phone service provider. The data span 40 of the 41 months February 2002 through June 2005 (December 2002 is missing), and include 32,852 individual bills. I find that customer tariff choices and subsequent usage decisions are not only consistent with the model of overconfidence, but just what would be expected from overconfident consumers. Moreover, usage patterns suggest that the overconfidence explanation is more appropriate than the price discrimination explanation in this particular application. Specifically, the distribution of usage by customers on a plan with a large number of included minutes strictly first order stochastically dominates (FOSD) the distribution of usage by customers on a plan with a small number of included minutes. This is inconsistent with the price discrimination model given three-part tariff pricing.

The paper proceeds as follows. Section 2 discusses related literature. Section 3 illustrates the intuition for the results with a simple example. Sections 4 and 5 present and analyze the single-tariff model, and Section 6 extends the analysis to multiple-tariff menus. Potential alternative explanations are described in Section 7, and then tested empirically in Section 8. Finally, Sections 9 and 10 describe extensions and conclude. (Supplementary material referred to in the paper is contained in a Web Appendix that is available for download from my website: www.stanford.edu/~mgrubb/)

2 Related Literature

2.1 Nonlinear Pricing

Any model which explains the use of three-part tariffs should capture their primary qualitative feature: included quantities at zero marginal price followed by positive marginal charges. While there is a large literature on non-linear pricing (see Wilson (1993)), there are no compelling explanations for three-part tariffs.

The initial monopoly model of non-linear pricing developed by Mussa and Rosen (1978) explains the use of quantity discounts, implemented either via a single nonlinear tariff or via a menu of two-part tariffs. The two implementations are equivalent when the optimal nonlinear tariff is concave, because Mussa and Rosen's (1978) model assumes that buyers do not to learn more about their demand over time, so consumers simultaneously choose tariff schedules and consumption quantities. The model predicts marginal cost pricing for the last unit, and higher marginal prices for all lower quantities. It cannot explain marginal charges which are at or below marginal cost for low quantities, but are significantly more expensive at higher quantities.[3]

While standard screening models are static, reality is dynamic. Consumers first choose from a menu of offered tariffs, and then later choose how much to consume. In the intervening period, consumers may acquire more private information about their demand. The importance of information arrival between tariff choice and usage choice can be inferred from the substantial empirical evidence documenting ex post "mistakes" in which consumers would have paid less for the same usage had they initially selected a different tariff. Such ex post "mistakes" are prevalent

[3] Of course, prices on a particular tariff for quantities that are never chosen may be somewhat arbitrary. In a static screening model, all that matters in a tariff menu is the lower envelope of tariffs on the menu. Segments of tariffs which are above that minimum may be set arbitrarily, for instance to include regions of zero marginal price. This is not a compelling explanation of the structure of cell phone tariffs, however. First, zero marginal price regions are typically part of the lower envelope of tariffs on the menu. What is more, customer billing data shows that usage falls within the zero marginal price regions of tariffs approximately 80% of the time, and then on average reaches only half of the included allowance (See Section 8).

in the usage data described in Section 8, and have been documented by others in similar contexts (Miravete 2003, Miravete 2005, Lambrecht, Seim and Skiera 2005).[4]

There are several papers which explicitly model two-stage screening by a monopolist in which consumers choose a tariff after receiving a signal about their type, and later make a consumption decision given the chosen tariff once they learn their true type. Although none of this research specifically addresses three-part tariff pricing, this branch of the nonlinear pricing literature is a natural place to begin thinking about three-part tariff pricing.

Miravete (1996) explicitly models two-stage screening by a monopolist who offers a menu of two-part tariffs. Courty and Li (2000) model a two-stage screening problem without restricting tariffs to any particular format. Their motivating example relates to airline ticket pricing, so assumes unit demand. There is a continuous quantity variable in the model because the monopolist is able to commit to deliver the good with probability $q \in [0, 1]$. In a model with a continuum of ex ante types, Courty and Li (2000) show that a menu of deterministic refund contracts, which are equivalent to two-part tariffs, is optimal.[5] They characterize the optimal menu of refund contracts for two ex ante types as well as for a continuum of ex ante types. Rochet and Stole (2003) point out that Courty and Li's (2000) model can be adapted to allow for sales of multiple units for which consumers have declining marginal valuations.[6] In this spirit, Miravete (2005) characterizes an optimal menu of nonlinear tariffs in a two-stage screening problem for a particular utility function and class of type distributions.

Both the standard screening model, and Courty and Li (2000) are closely related to this paper. The single-tariff model presented in Section 4 is closely related to the standard monopoly screening problem. The important differences in this paper are the incorporation of consumer overconfidence, free disposal, and an ex ante participation constraint, which together predict pricing qualitatively similar to a single three-part tariff. Similarly, both the multi-tariff monopoly model with overconfidence presented in Section 6.1 and the price discrimination model with common priors discussed in Section 7 are closely related to Courty and Li (2000). After incorporating continuous demand with declining marginal valuations and free disposal, I provide two alternative explanations for menus of three-part tariffs, by incorporating either overconfidence (Section 6.1) or considering alternative type distributions (Section 7).

---

[4]Using data from a 1986 local telephone tariff experiment, Miravete (2005) finds that the average correlation between expected and actual calls is only 0.34.

[5]Since the quantity variable is the probability of delivery, the marginal value of a unit of quantity is constant over the feasible range $q \in [0, 1]$. This produces bang-bang results in which the optimal allocation is either 0 or 1.

[6]See Rochet and Stole (2003) Section 8.

There are a number of related papers which consider the affect of consumer biases or non-standard preferences on optimal nonlinear pricing. I discuss these as potential alternatives to overconfidence in Section 7.

2.2 Overconfidence

Loewenstein et al. (2003) present a variety of evidence demonstrating the prevalence of projection bias. Individuals who exhibit this bias overestimate the degree to which their future tastes will resemble their current tastes, and therefore tend to underestimate the variance of their future demand. Moreover, a significant body of literature shows that individuals are overconfident about the precision of their own predictions when making difficult[7] forecasts. In other words, individuals tend to set overly narrow confidence intervals relative to their own confidence levels.

Lichtenstein, Fischhoff and Phillips (1982) and Arkes (2001) provide surveys of the experimental literature concerning forecasting overconfidence. A typical study in this literature might pose the following question to a group of subjects: "What is the shortest distance between England and Australia?" Subjects would then be asked to give a set of confidence intervals centered on the median. Lichtenstein et al. (1982) tabulate the results of 13 such studies. A typical finding is that the true answer lies outside a subject's 98% confidence interval about 30% to 40% of the time. The literature provides evidence that overconfidence diminishes with appropriate feedback (Bolger and Onkal-Atay 2004), but also that professionals are often overconfident within the realms of their expertise (Griffin and Tversky 1992). Experimental evidence therefore suggests that, at a minimum, new cell-phone users will be overconfident about their usage predictions when they initiate service and choose a calling tariff. Moreover, ex post tariff-choice "mistakes" made by cellular phone customers are consistent with such overconfidence, as documented in Section 8.

3 Illustrative Example

A simple example illustrates my main results. Assume that a supplier has a constant marginal cost of 5 cents per minute and a fixed cost of $50 per customer.[8] Consumers value each additional minute of consumption at 45 cents up to some satiation point, beyond which they value further minutes at 0 cents. When consumers sign up for a tariff in period one, they are homogeneously

[7] Predicting one's future demand for minutes is a relatively difficult task, at least for new cell-phone users. Consumers must predict not only the volume of outgoing calls they will make, but also the number of incoming calls they will receive.

[8] Fixed costs per customer may arise due to billing costs, a subsidy for a new phone, or customer acquisition fees paid to retailers.

uncertain about their satiation points. Then in period two, consumers learn their satiation points, and use this information to make their consumption choices. In particular, assume that one third of consumers learn that they will be satiated after 100 minutes, one third after 400 minutes, and the remaining third after 700 minutes.

If consumers and the supplier share this prior belief, then it is optimal for the firm to charge a marginal price equal to the marginal cost of 5 cents per minute.[9] Under monopoly the firm extracts all the surplus via a fixed fee of $160, earning profits of $110 per customer. Under perfect competition, the firm charges a fixed fee of $50, leaving $110 in surplus to consumers.

If consumers are overconfident, however, marginal cost pricing is no longer optimal. For instance, if all consumers are extremely overconfident and believe that they will be satiated after 400 minutes with probability one, then it is optimal to charge 0 cents per minute for the first 400 minutes, and 45 cents per minute thereafter. In other words it is optimal to have 400 "included" minutes in the tariff.

Under monopoly the firm charges a fixed fee of $180, earning expected profits of $155 per customer. Ex ante consumers expect to receive zero surplus, but on average ex post realize a loss of $45. Under perfect competition, the firm charges a fixed fee of $25, and consumers expect to receive $155 in surplus, but actually only realize $110. Consumer overconfidence allows the creation ex ante of an additional $45 in perceived consumer surplus, which is never realized ex post.

To see why this tariff is optimal, consider the pricing of minutes 100-400 and 400-700 separately. On the one hand, overconfident consumers believe that they will consume minutes 100-400 with probability 1, while the firm knows that they will actually consume them only with probability $\frac{2}{3}$. As a result, reducing the marginal price of minutes 100-400 from 5 cents to 0 cents is perceived differently by the firm and consumer. The consumer views this as a $15 price cut and will be indifferent if the fixed fee is increased by $15. The firm, however, recognizes this as only a $10 revenue loss, and will be better off by $5 if the fixed fee is raised by $15.

On the other hand, overconfident consumers believe that they will consume minutes 400-700 with probability 0, while the firm knows that they will actually consume them with probability $\frac{1}{3}$. Therefore from the consumer's perspective, increasing the marginal price of minutes 400-700 from 5 cents to 45 cents does not impact the expected price paid. The firm, however, views this as an increase in expected revenues of $40.

Essentially, the firm finds it optimal to sell the first 400 minutes upfront to overconfident

[9] Note that this is only one of a continuum of optimal pricing structures which all implement the efficient allocation. Were demand curves not rectangular and were there a continuum of types, then marginal cost pricing would be uniquely optimal.

consumers. Then in the second period, the firm buys back minutes 100-300 from the low demand consumers at the monopsony price of 0 cents per minute, and sells minutes 400-700 to high demand consumers at the monopoly price of 45 cents per minute.

Note that in this example, a monopolist earns higher profits from overconfident consumers, making them worse off than consumers with correct priors. Under competition, however, overconfident consumers are equally as well off as consumers with correct priors. Neither result is true in general, rather both follow from the specific form of preferences assumed (see Section 5.6).

4 Single-Tariff Model

Game players are a firm, or multiple firms in the case of perfect competition, and a continuum of consumers. Timing of the game (Figure 1) differs from a standard screening model. At the contracting stage ($t = 1$) consumers are homogeneous and do not know their future demand type θ. The firm offers tariff $\{q(\theta), P(\theta)\}$, which describes a purchase quantity and payment pair intended for each type θ. Consumers accept or reject based on their prior belief over θ at $t = 1$. Finally, consumers privately learn θ and choose to purchase quantity $q(\hat{\theta})$ at $t = 2$.

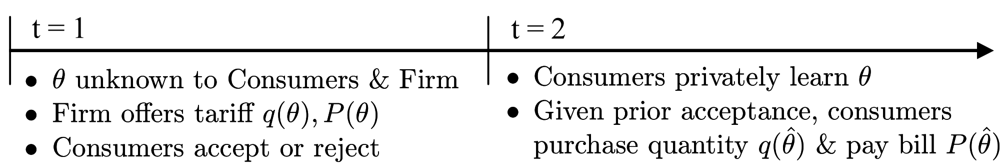

Figure 1: Time Line

The base assumptions about production and preferences match those of a standard screening model. A firm's profits Π are given by revenues P less production costs $C(q)$, which are increasing and convex in quantity delivered q. Consumers' utility U is equal to their value of consuming q^c units, $V(q^c, \theta)$, less their payment to the firm, P.

Consumers' marginal value of consumption V_q is strictly decreasing in consumption q^c, and strictly increasing in consumers' type θ. The outside option of all consumers is the same and normalized to zero: $V(0, \theta) = 0$. The partial derivative $V_{qq\theta}$ is assumed to be equal to zero, and with this additional assumption it is then without further loss of generality to set $V_{q\theta\theta} = 0$ by appropriate normalization of θ. The consumers' value function may then be written as $V(q^c, \theta) = v(q^c) + q^c\theta$.[10]

[10] The assumption $V_{qq\theta} = 0$ implies that demand curves of all types are parallel. As illustrated by the example in Section 3, this assumption is not necessary for my results. I make the assumption to guarantee that the virtual

I make an additional assumption concerning consumer preferences, which would not be relevant in a standard model: Consumers have a finite satiation point, $q^S(\theta) \equiv \arg\max_{q^c \geq 0} V(q^c, \theta)$, beyond which they may freely dispose of unwanted units. Hence, given free disposal, consumer type θ who purchases $q(\hat{\theta})$ units will only consume the minimum of $q(\hat{\theta})$ and $q^S(\theta)$, and will receive consumption value $V(\min\{q^S(\theta), q(\hat{\theta})\}, \theta)$.[11]

The key assumption of the model, which deviates sharply from a standard model, is that consumers underestimate the variance of their future demand θ. This is either because they are overconfident about the accuracy of their forecasts of θ, or because they are subject to projection bias. Thus while the firm knows[12] that consumer demand θ follows cumulative distribution $F(\theta)$, consumers have the prior belief that θ follows $F^*(\theta)$. Moreover, the firm knows that consumers are overconfident, so will take this into account when designing its tariff offering. Finally, the disagreement between the firm and consumers is captured by assumption A*:

Assumption A*:[13] *$F^*(\theta)$ crosses $F(\theta)$ once from below at θ^*.*

An interesting special case of A* is where consumers and the firm agree on the mean of θ, in which case $F(\theta)$ is a mean preserving spread of $F^*(\theta)$ and consumers underestimate the variance of their future demand. Moreover, it implies that consumers correctly predict their mean value of

---

surplus function described in Proposition 1 is strictly concave, and therefore that a first order condition uniquely identifies its maximum. This is important in the proof of Lemma 2. A weaker, but still not necessary, sufficient condition is: $V_{qq\theta}(q,\theta)\frac{F(\theta)-F^*(\theta)}{f(\theta)} < C_{qq}(q) - V_{qq}(q,\theta)$. The analogous assumption in a standard screening model is $V_{qq\theta} \leq 0$ (e.g. see Fudenberg and Tirole (1991) Chapter 7).

[11] The satiation point $q^S(\theta)$ is the point at which type θ's (strictly decreasing) marginal value of consumption becomes negative. Rather than explicitly allowing for free disposal, it would have been equivalent to assume directly that consumers have value function $\hat{V}(q,\theta) = V\left(\min\left\{q, q^S(\theta)\right\}, \theta\right)$ for which the marginal value of consumption is zero beyond the finite point $q^S(\theta)$.

[12] Strictly speaking there is no need to assume that either the firm's prior or the consumer's prior is correct, except in order to make statements about welfare. The interpretation maintained throughout this paper is that the firm's beliefs are correct and the consumers' beliefs are incorrect. A larger game is imagined in which the firm quickly learns the true distribution of types of new consumers by observation of its large number of existing customers. New consumers, however, are overconfident and believe they know more about their own type than they really do, as described in (A*).

[13] Note that assumption A* corresponds closely to the two documented biases, forecasting overconfidence and projection bias, from which it is motivated. For instance, the special case of assumption A* where $F^*(\theta)$ is given by the equation below for some $\alpha \in (0,1)$ exactly matches Loewenstein et al.'s (2003) formalization of projection bias.

$$F^*(\theta) = \begin{cases} (1-\alpha)\cdot F(\theta) & \theta < \theta^* \\ (1-\alpha)\cdot F(\theta) + \alpha & \theta \geq \theta^* \end{cases}$$

In this case θ^* would be interpreted as a consumer's current taste for consumption when making his or her participation decision at $t = 1$. (This is not how Loewenstein et al. (2003) present their model, but it is straightforward to show the equivalence, as they hint in their Footnote 8.)

Further, assumption A* guarantees that any confidence interval drawn by an individual that includes θ^* will be overly narrow. Furthermore, if all of an individual's perceived confidence intervals which include θ^* are strict subsets of the true confidence intervals, assumption A* must hold. If we think of θ^* as a central point such as the median, this provides a strong link to the studies of forecasting overconfidence.

each unit.

Within the context of this model, the equilibrium tariff, or allocation and payment pair $\{q^* (\theta), P^* (\theta)\}$, will be characterized under both monopoly and perfect competition. This analysis requires several more technical assumptions. As is standard, it is assumed that $V(q, \theta)$ is thrice continuously differentiable, $C(q)$ and $F(\theta)$ are twice continuously differentiable, $F^* (\theta)$ is continuous and piecewise smooth, consumption is non-negative, and total surplus is initially strictly increasing in q. The firm's prior $F(\theta)$ has full support over $\left[\underline{\theta}, \bar{\theta}\right]$, a range which includes the support of consumers' prior $F^* (\theta)$.

5 Single-Tariff Pricing

5.1 Defining the Problem

Invoking the standard revelation principle, the equilibrium monopoly tariff $\{q^M (\theta), P^M (\theta)\}$ must solve the following constrained profit maximization problem:

$$\max_{\substack{P(\theta) \\ q(\theta) \geq 0}} E\left[P(\theta) - C(q(\theta))\right]$$

such that:

1. Global IC $\quad U(\theta, \theta) \geq U(\theta, \hat{\theta}) \quad \forall \theta, \hat{\theta} \in \left[\underline{\theta}, \bar{\theta}\right]$
2. Consumer Participation[14] $\quad E^* \left[U(\theta)\right] \geq 0$

The monopolist's problem is similar to that in a standard screening model. The monopolist's objective is the same: to maximize expected profits. Moreover, at $t = 2$ when consumers privately learn their types, it must be optimal for consumers to truthfully reveal their types by self-selecting appropriate quantity - payment pairs from the tariff. Thus the standard incentive compatibility constraint applies: the utility $U(\theta, \hat{\theta}) \equiv V(\min\{q^S (\theta), q(\hat{\theta})\}, \theta) - P(\hat{\theta})$ of a consumer of type θ who reports $\hat{\theta}$ at $t = 2$ must be weakly below the utility $U(\theta) \equiv U(\theta, \theta)$ of a consumer of type θ who reports truthfully at $t = 2$.

There are two important deviations from a standard screening model. First, free disposal is explicitly incorporated through consumer preferences, which depend on the consumed quantity $\min\{q^S (\theta), q(\hat{\theta})\}$ rather than the purchased quantity $q(\hat{\theta})$.[15] Second, consumers' ex ante prior over

[14] Expectations taken with respect to the consumers' prior $F^* (\theta)$ are denoted by a superscript * on the expectations operator.

[15] This alone would have no impact on a standard monopoly screening model. The assumption will be important

types $F^{*}(\theta)$ differs from that of the firm $F(\theta)$. Thus the ex ante participation constraint requires that consumers' perceived expected utility $E^{*}[U(\theta)]$ must be positive, but puts no constraint on their true expected utility $E[U(\theta)]$. The difference in priors between consumers and the firm creates a wedge separating the expected utility consumers believe they are receiving from the expected utility the firm believes it is actually providing.

Invoking the revelation principle a second time, the equilibrium tariff $\{q^{C}(\theta), P^{C}(\theta)\}$ under perfect competition must solve the following closely related constrained maximization problem:

$$\max_{\substack{P(\theta)\\ q(\theta)\geq 0}} E^{*}[U(\theta)]$$

such that:

1. Global IC $\quad U(\theta,\theta) \geq U(\theta,\hat{\theta}) \;\; \forall \theta, \hat{\theta} \in [\underline{\theta}, \overline{\theta}]$
2. Producer Participation $\quad E[P(\theta) - C(q(\theta))] \geq 0$

As under monopoly, the equilibrium tariff must satisfy incentive compatibility constraints. The difference is that the objective function and participation constraints are reversed. Under perfect competition the equilibrium tariff maximizes consumers' perceived expected utility subject to firm participation,[16] whereas under monopoly firm payoff is maximized subject to consumer participation.

5.2 Simplifying the Problem

The initial step in simplifying the problem is to recognize that there will never be any reason for firms to induce a consumer to purchase beyond her satiation point, since she would simply dispose of unwanted additional units. By initially selling the consumer her satiation quantity at the same price, the consumer would have been equally well off and the firm could have reduced production costs.

Lemma 1 *If the pair $\{q^{*}(\theta), P^{*}(\theta)\}$ is an optimal tariff under either monopoly or perfect competition, then the pair $\left\{\min\left\{q^{*}(\theta), q^{S}(\theta)\right\}, P(\theta)\right\}$ is also optimal. Moreover, if production costs are strictly increasing, then $q^{*}(\theta) = \min\left\{q^{*}(\theta), q^{S}(\theta)\right\}$ almost everywhere.*

Proof. See Appendix A. ■

here, however, because consumers and the firm have different priors over θ.

[16] Otherwise there would be an opportunity for profitable entry.

I will focus on equilibria in which firms offer allocations no higher than consumers' satiation points. Given Lemma 1, this is without loss of generality when marginal costs are strictly positive. When marginal costs are zero, this refinement simply selects the limiting equilibrium as marginal costs approach zero. As a result, rather than separately tracking equilibrium purchases $q(\theta)$ and consumption $\min\left\{q(\theta), q^S(\theta)\right\}$, knowing that they will be the same I can work with a single quantity $q(\theta)$ by imposing a satiation constraint $q(\theta) \leq q^S(\theta)$.

Having reduced equilibrium purchase and consumption quantities to a single function $q(\theta)$, the problem can be further simplified following the standard approach. The first step, introduced by Mirrlees (1971), is to replace the global incentive compatibility constraint with the joint constraints of local incentive compatibility and monotonicity. The second step is to recognize that under either monopoly or perfect competition, the relevant participation constraint must bind. Now, for every allocation $q(\theta)$ there is a unique payment function $P(\theta)$ which satisfies local incentive compatibility, and meets the relevant participation constraint with equality.[17] Both monopoly and perfect competition problems may then be simplified by substituting these two constraints in place of payments $P(\theta)$ in the objective function.

Completing the described substitution for both monopoly and perfect competition reveals a beautiful aspect of the two problems. The transformed objective functions, now expressed solely as a function of the allocation $q(\theta)$, are identical under both monopoly and perfect competition. In particular, in both market scenarios $q(\theta)$ maximizes an expected virtual surplus $E[\Psi(q(\theta), \theta)]$, subject to the remaining non-negativity, monotonicity, and satiation constraints. Expected virtual surplus is equal to the sum of expected true surplus, $S(q(\theta), \theta) \equiv V(q(\theta), \theta) - C(q(\theta))$, and a "fictional surplus", which is the difference between the expected utility $E^*[U(\theta)]$ consumers believe they are receiving and the expected utility $E[U(\theta)]$ the firm believes it is delivering (equation 1). Moreover, having substituted local incentive compatibility and participation constraints in place of payments, fictional surplus is given by equation (2).

$$E[\Psi(q(\theta), \theta)] = E[S(q(\theta), \theta)] + E^*[U(\theta)] - E[U(\theta)] \tag{1}$$

$$E^*[U(\theta)] - E[U(\theta)] = E\left[V_\theta(q(\theta), \theta)\, \frac{F(\theta) - F^*(\theta)}{f(\theta)}\right] \tag{2}$$

When consumers and the firm share the same prior ($F^*(\theta) = F(\theta)$) fictional surplus is zero, so the equilibrium tariff maximizes expected surplus $E[S(q(\theta), \theta)]$. This implies first best allocation

[17] This payment function can be found first by expressing payments in terms of consumer utility: $P(\theta) = U(\theta) - V\left(\min\left\{q(\theta), q^S(\theta)\right\}, \theta\right)$. Next, local incentive compatibility requires that $U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} V_\theta(q^c(z), z)\, dz$, which pins down payments up to a constant $U(\underline{\theta})$. Finally, binding participation constraints determine the constant $U(\underline{\theta})$.

$q^{FB}(\theta)$ and marginal payment equal to marginal cost.

$$q^{FB}(\theta) \equiv \arg\max_{q} \left[V(q,\theta) - C(q)\right]$$

When consumers are overconfident, however, fictional surplus need not be zero, and may distort the equilibrium allocation away from first best, and marginal pricing away from marginal cost. These distortions, and thus the equilibrium allocation $q^*(\theta) = q^M(\theta) = q^C(\theta)$, will be identical under monopoly and perfect competition. As a result, marginal pricing, which is pinned down jointly by the allocation and local incentive compatibility, will be the same across market conditions. The only variation in pricing will be a higher fixed fee under monopoly, due to the difference in participation constraints across market conditions.[18] Proposition 1 summarizes these results precisely.

Proposition 1 *Under both monopoly and perfect competition:*

1. *Equilibrium allocations are identical, and maximize expected virtual surplus:*

$$q^*(\theta) = \arg \max_{\substack{q(\theta)\in\left[0,q^S(\theta)\right] \\ q(\theta) \text{ non-decreasing}}} E\left[\Psi(q(\theta),\theta)\right]$$

$$\Psi(q,\theta) \equiv V(q,\theta) - C(q) + V_\theta(q,\theta)\frac{F(\theta) - F^*(\theta)}{f(\theta)} \tag{3}$$

2. *Payments differ only by a fixed fee and are given by:*

$$P^C(\theta) = V(q^*(\theta),\theta) - \int_{\underline{\theta}}^{\theta} V_\theta(q^*(z),z)\,dz - E\left[S(q^*(\theta),\theta)\right] \tag{4}$$

$$P^M(\theta) = P^C(\theta) + E\left[\Psi(q^*(\theta),\theta)\right] \tag{5}$$

3. *At quantities for which there is no pooling, marginal price is given by equation (6) as a function of the inverse equilibrium allocation $\theta(q)$:*

$$\frac{dP^*(q)}{dq} = V_q(q,\theta(q)) \tag{6}$$

[18] The main results are easily extended to imperfect competition in which firms are differentiated by location and consumers' transportation costs d are independent of consumption or type θ. (For example $V(q,\theta,d) = V(q,\theta) - d$). Equilibrium allocations and marginal prices would be identical to those in the current model, which maximize expected virtual surplus. Firms would compete with each other through the fixed fees, which would drop with the level of competition. (In contrast, distortions of price away from marginal cost in a standard price discrimination model disappear with increasing competition (Stole 1995).)

Proof. Outlined in the text above. For further details see Appendix A. ■

5.3 Equilibrium Allocation

Further characterization of the equilibrium allocation follows the standard approach. First, the solution $q^R(\theta)$ to a relaxed problem (equation 7) that ignores the monotonicity constraint is characterized.

$$q^R(\theta) \equiv \arg \max_{q \in [0, q^S(\theta)]} \Psi(q, \theta) \tag{7}$$

Second, any non-monotonicities in $q^R(\theta)$ are "ironed out." Implications about pricing can then be drawn based on the result in Proposition 1 that marginal price is equal to $V_q(q, \theta(q))$.

Lemma 2 *1. The relaxed solution $q^R(\theta)$ is a continuous and piecewise smooth function characterized by the first order condition $\Psi_q(q, \theta) = 0$ except where satiation or non-negativity constraints bind.*

2. The equilibrium allocation $q^*(\theta)$ is continuous and piecewise smooth. On any interval over which the monotonicity constraint is not binding, the equilibrium allocation is equal to the relaxed allocation: $q^*(\theta) = q^R(\theta)$.

Proof. Part 1: See Appendix A. Part 2: The proof of part 2 is omitted as it closely follows ironing results for the standard screening model. It follows from the application of standard results in optimal control theory (Seierstad and Sydsæter 1987, Leonard and Long 1992) and the Kuhn-Tucker theorem. ■

Lemma 2 closely parallels analogous results in standard screening models. The important point is that the equilibrium allocation $q^*(\theta)$ is continuous and equal to the relaxed allocation $q^R(\theta)$ where the monotonicity constraint is not binding. This fact is useful since it implies that the relaxed solution $q^R(\theta)$ determines marginal prices (Proposition 2).

When consumers are extremely overconfident, the relaxed solution will violate the monotonicity constraint (Web Appendix C Proposition 6). Thus to avoid excluding interesting cases, Web Appendix C characterizes an ironed solution (Proposition 4) and provides details of pooling in equilibrium.

5.4 Pricing Implications

Having characterized the equilibrium allocation $q^*(\theta)$, it is now possible to draw implications about pricing using Proposition 1.

Proposition 2 *The equilibrium payment $P^*(q) = P^*(\theta(q))$ is a continuous and piece-wise smooth function of quantity. There may be kinks in the payment function where marginal price increases discontinuously. These kinks occur where the monotonicity constraint binds and an interval of types "pool" at the same quantity. For quantities at which there is no pooling, marginal price is given by equation (8):*

$$\frac{dP^*(q)}{dq} = \max\left\{0, C_q(q) + V_{q\theta}(q, \theta(q)) \frac{F^*(\theta(q)) - F(\theta(q))}{f(\theta(q))}\right\} \tag{8}$$

Proof. See Appendix A. ■

Since it is assumed that $V_{q\theta}$ is strictly positive and $f(\theta)$ is finite, Proposition 2 allows marginal price to be compared to marginal cost based on the sign of $[F^*(\theta) - F(\theta)]$. In particular, the sign of $\left[P_q^*(q) - C_q(q)\right]$ is equal to the sign of $[F^*(\theta) - F(\theta)]$ except when $F^*(\theta) < F(\theta)$ and marginal cost is zero, since then marginal price is also zero. This is informative about equilibrium pricing, since assumption A* dictates the sign of $[F^*(\theta) - F(\theta)]$ above and below θ^*.

Define $\underline{q}$, Q, and $\bar{q}$ to be the equilibrium allocations of types $\underline{\theta}$, θ^*, and $\bar{\theta}$ respectively:

$$\{\underline{q}, Q, \bar{q}\} \equiv \{q^*(\underline{\theta}), q^*(\theta^*), q^*(\bar{\theta})\}$$

Relevant implications of Proposition 2 are then summarized in Corollary 1.

Corollary 1 *Given A^*, for quantities at which there is no pooling: (1) If marginal cost is zero for all q then:*

$$\begin{aligned} P_q^*(q) &= 0 \quad , \; q \in (\underline{q}, Q) \cup \{\underline{q}, Q, \bar{q}\} \\ P_q^*(q) &> 0 \quad , \; q \in (Q, \bar{q}) \end{aligned}$$

(2) If marginal cost is strictly positive for all q then:

$$\begin{aligned} P_q^*(q) = C_q(q) &> 0 \quad , \; q \in \{\underline{q}, Q, \bar{q}\} \\ C_q(q) > P_q^*(q) &\geq 0 \quad , \; q \in (\underline{q}, Q) \\ P_q^*(q) > C_q(q) &> 0 \quad , \; q \in (Q, \bar{q}) \end{aligned}$$

Proof. Follows directly from Proposition 2, assumption A*, and $q^*(\theta)$ non-decreasing. ■

Corollary 1 shows that when marginal costs are zero, marginal price will be zero below some included allowance Q, and positive thereafter. When marginal costs are strictly positive, marginal price will initially be positive, but will fall below marginal cost and may be zero for some early range of consumption. The following section illustrates these results with numerical examples.

It is reasonable to assume that the marginal cost of providing an extra minute of call time to a cell phone customer is small. Therefore, given overconfident consumers, the equilibrium tariff bears a striking qualitative resemblance to those offered by cell-phone service providers. Both predicted equilibrium tariffs and observed tariffs involve zero marginal price up to some included minute limit Q and become positive thereafter.

The primary difference is that beyond the included limit Q, marginal price is constant for observed tariffs. I conjecture that this simpler pricing structure approximates optimal pricing and is more practical to implement. The fact that marginal price does not fall to marginal cost at the top may also be due to binding period-one incentive compatibility constraints relevant to the un-modeled self-selection among tariffs at time one (see Section 6).

The intuition for the result is as follows. If consumers are overly confident that their future consumption will be near Q minutes, they will underestimate both the probability of extremely low and extremely high consumption. Thus a firm cannot charge these consumers for extremely high consumption through a fixed fee ex ante. Instead, the firm must wait until consumers learn their true values and charge a marginal fee for high consumption above Q. A firm can, however, charge consumers for low levels of consumption through a fixed fee ex ante. By setting a zero marginal price, the firm avoids paying a refund to those consumers who are later surprised by a low level of demand below Q. (Web Appendix B provides additional intuition based on option pricing.)

5.5 Example

The implications of Proposition 2 that are summarized in Corollary 1 are best illustrated with figures from specific examples. Consider the following example that satisfies the model assumptions outlined in Section 4.

Example 1 *Firms have a fixed cost of* \$25 *and a constant marginal cost of* $c \geq 0$ *per unit:* $C(q) = 25 + q \cdot c$. *Consumers' inverse demand function is linear in* q *and* θ*. In particular,* θ *simply shifts the consumers inverse demand curve up and down (Figure 2):*

$$V(q,\theta) = \frac{3}{2} q \left[1 + \theta - \frac{1}{1000} q\right]$$

$$V_q(q,\theta) = \frac{3}{2} \left[1 + \theta - \frac{2}{1000} q\right]$$

The firm and consumers' priors are uniform, centered on 0*:* $F : U\left[-\frac{1}{2}, \frac{1}{2}\right]$ *and* $F^* : U\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$. *Consumers and the firm both agree that the mean of* θ *is equal to* 0*:* $E^*[\theta] = E[\theta] = 0$. *The parameter* $\Delta \in [0,1]$ *is a measure of consumer overconfidence. For* $\Delta = 1$*, consumers are not overconfident at all, and share the firm's prior. For* $\Delta = 0$*, consumers are extremely overconfident and believe* $\theta = 0$ *with probability one (Figure 2).*

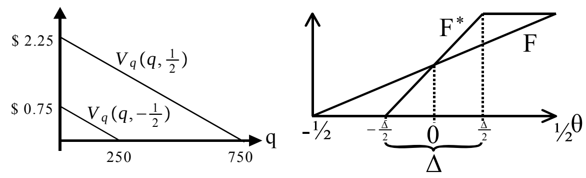

Figure 2: Inverse demand curves and priors in example 1.

Satiation and first best allocations are given by:

$$q^S(\theta) = 500(1+\theta)$$

$$q^{FB}(\theta) = 500(1+\theta-c)$$

The equilibrium allocation $q^*(\theta)$ and pricing $P^*(q)$ depend on the size of marginal cost c and the level of overconfidence Δ.

Figure 3 illustrates Corollary 1 given zero marginal costs, using the example described above. In the top row, plots A and B show total equilibrium payment $P^C(q)$ and total cost $C(q)$ versus quantity under perfect competition. In the bottom row, plots C and D show marginal equilibrium payment $P_q^*(q)$ and marginal cost $C_q(q)$ versus quantity, under either perfect competition or monopoly. In the left hand column, plots A and C assume low overconfidence $\Delta = 0.75$ for which there is no pooling. In the right hand column, plots B and D assume high overconfidence $\Delta = 0.25$ for which there is pooling at Q.

Figure 3 shows that total payment is constant and marginal price is zero up to some quantity Q. Beyond Q, marginal price is positive. When there is no pooling at Q, total payment increases smoothly beyond Q. When there is pooling at Q, however, the total payment has a kink at Q where marginal price jumps upwards discretely. In both cases marginal price falls to zero at the highest quantity $\bar{q}$.

Figure 4 shows the same plots given in Figure 3 except that equilibrium payments are plotted for strictly positive marginal cost $c = \$0.035$ rather than zero marginal cost. The plots are similar to those in Figure 3 for quantities above Q. However, since marginal cost is strictly positive, marginal price is not zero everywhere below Q. In particular, below Q marginal price is strictly positive near $\underline{q}$ and Q. In the example shown the satiation constraint does bind and marginal price is zero over some subset of the interval $[\underline{q}, Q]$. However, were marginal cost higher, the satiation constraint

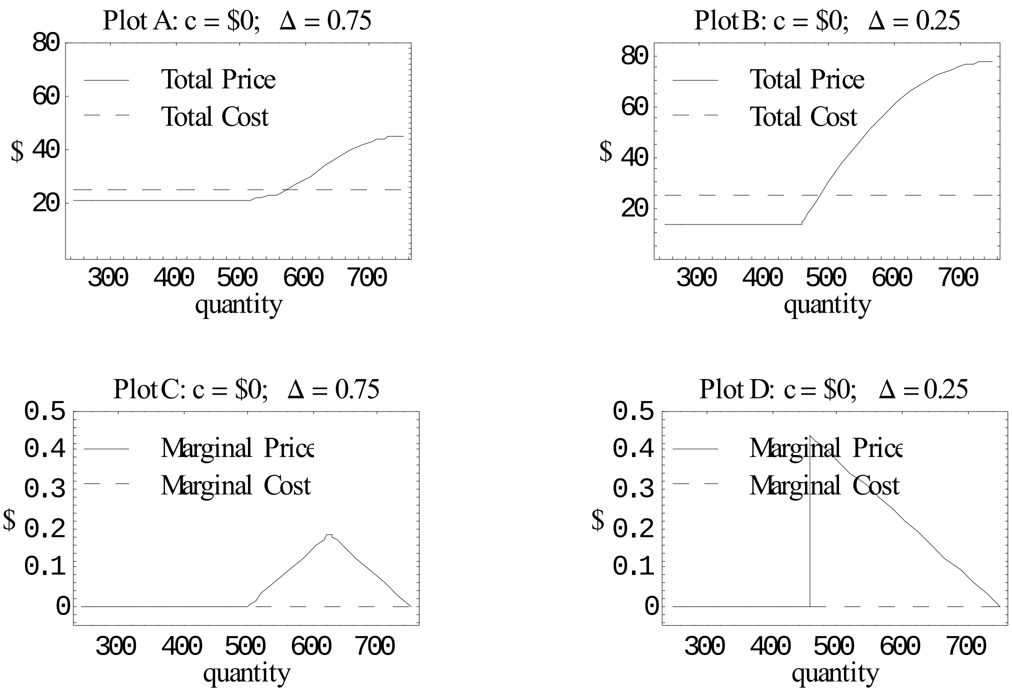

Figure 3: Equilibrium pricing under perfect competition and zero marginal cost is depicted for low overconfidence ($\Delta = 0.75$) in the left hand column and for high overconfidence ($\Delta = 0.25$) in the right hand column.

might never bind, and marginal price could be strictly positive at all quantities.

5.6 Welfare

To evaluate welfare I assume that the firm's prior $F(\theta)$ is correct.[19] Therefore consumers' expected surplus is evaluated with respect to the firm's prior $F(\theta)$, as are expected firm profits and total surplus. Under perfect competition, welfare conclusions are straightforward. Consumers receive all the surplus generated. However, while consumers with correct priors receive the efficient allocation, overconfident consumers receive an allocation that is distorted away from first best. As a result, overconfident consumers must be worse off. This suggests that educating consumers or regulating constant marginal prices could potentially improve consumer welfare, and therefore total welfare, since firm profits are always zero.

Under monopoly, total welfare is also lower when consumers are overconfident, but in general

[19] See Footnote 12.

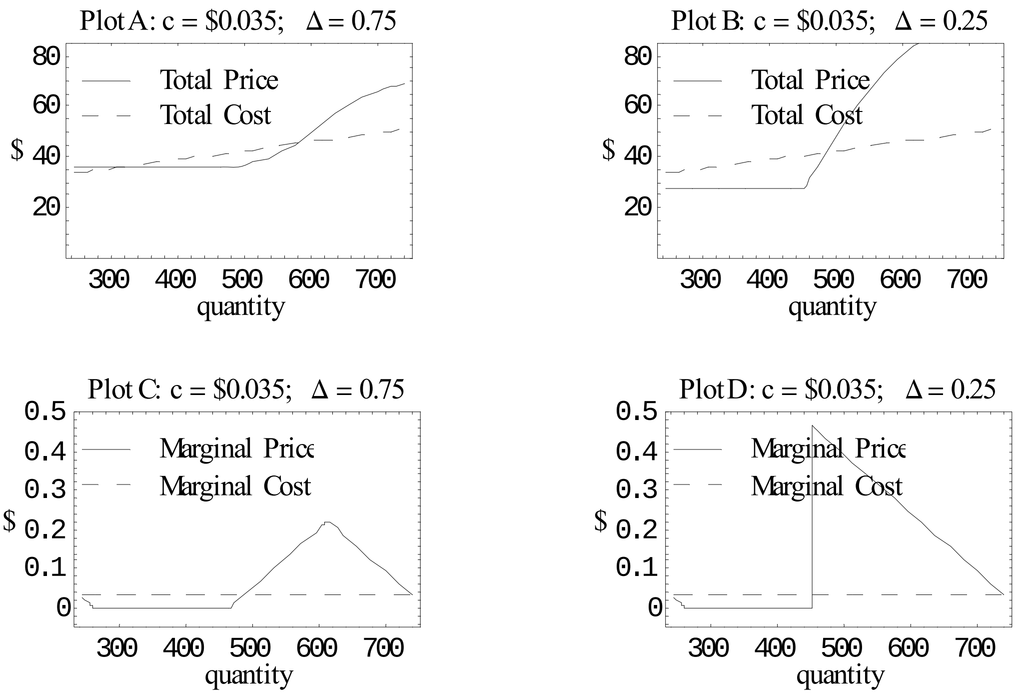

Figure 4: Equilibrium pricing under perfect competition and positive marginal cost $c = \$0.035$ is depicted for low overconfidence ($\Delta = 0.75$) in the left hand column and for high overconfidence ($\Delta = 0.25$) in the right hand column.

it is ambiguous as to whether consumers or the firm are better or worse off. The firm earns expected profits equal to expected virtual surplus, the sum of true surplus and fictional surplus, and therefore benefits from consumer overconfidence if and only if this is higher than first best surplus: $E[\Psi^*] \geq E\left[S^{FB}\right]$. Overconfident consumers' expected payoff under the correct prior is the remaining surplus $E[S^*] - E[\Psi^*]$, which is negative fictional surplus. They are therefore worse off if and only if: $-E\left[V_\theta\left(q^*(\theta), \theta\right) \frac{F(\theta) - F^*(\theta)}{f(\theta)}\right] \leq 0$. Neither condition is very helpful since both are in terms of the equilibrium allocation, but Lemma 3 gives a simple sufficient condition for both to be true.

Lemma 3 *Under monopoly, whenever overconfident consumers weakly overestimate the surplus created by the first best allocation,*[20] $E^*\left[S^{FB}\right] \geq E\left[S^{FB}\right]$, *the firm is better off and consumers are worse off due to their overconfidence.*

---

[20] *Note that under zero marginal costs, assuming that* $E^*\left[S^{FB}\right] \geq E\left[S^{FB}\right]$ *is equivalent to assuming that overconfident consumers overestimate their expected value of consuming up to their satiation points.*

Proof. See Appendix A. ■

The tables may be turned if overconfident consumers underestimate the expected surplus generated by the first best allocation, because this under estimation creates bargaining power. The firm cannot extract all surplus ex ante, and to extract it ex post the firm must give away information rents since the customer is privately informed about θ in period two. This is the case in the examples discussed in Section 5.5. There, although it is assumed that consumers estimate the mean of θ correctly, since the value of first best allocation is proportional to θ^2 and overconfident consumers underestimate the spread of θ, $E^* \left[S^{FB}\right]$ is strictly below $E\left[S^{FB}\right]$. Moreover, the underestimation of surplus is great enough that consumers are strictly better off when overconfident. Of course this also implies that the firm is worse off and would prefer customers to have correct priors.

The discussion of welfare has thus far assumed that consumers are homogeneously overconfident. If there are both correct-prior and overconfident types served in the marketplace, overconfident consumers must be weakly worse off than their counterparts with correct priors because correct beliefs lead to better decisions. That being said, it is possible that the presence of overconfident types in the marketplace improves the outcome for both types. Since types with correct priors can always choose any tariff offered to overconfident types, serving overconfident types also limits the rents which can be extracted from types with correct priors.

When there is ex ante heterogeneity in average demand (see Section 6), overconfidence causes consumers to believe that they are more different than they really are, and a monopolist must give up more information rents to screen them. This second effect suggests that consumer overconfidence is more likely to lower monopoly profits when initial screening of consumers between separate tariffs is important.

6 Multi-Tariff Menus

The model presented in Sections 4-5 assumes that consumers have homogeneous priors ex ante, and therefore firms offer only a single tariff. In reality consumers have heterogenous priors and are offered menus of multiple tariffs.

Rather than assuming that ex ante all consumers have homogenous prior $F^*(\theta)$, assume instead that each consumer receives a private signal $s \sim G(s)$ prior to choosing a tariff. The signal s does not enter payoffs directly, but is informative about $\theta \sim F^*(\theta|s)$. At $t = 1$, a firm (or multiple firms) will offer a menu of tariffs $\{q(s,\theta), P(s,\theta)\}$ from which consumers will choose based on their signal s. Then at $t = 2$, consumers privately learn θ, and choose a quantity based on both θ and their prior choice of tariff (Figure 5).

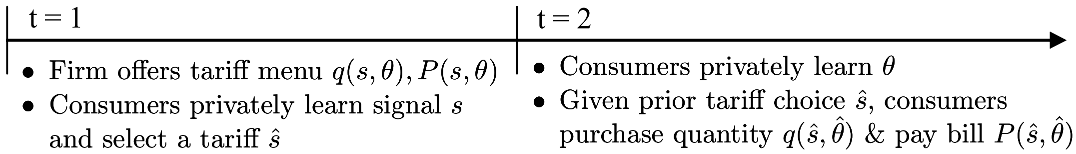

Figure 5: Time line for multi-tariff model.

Preferences are identical to those in the single-tariff model. Consumers are overconfident in the sense that consumers' conditional priors $F^*(\theta|s)$ cross firms' conditional priors $F(\theta|s)$ once from below at $\theta^*(s)$. The simplest ordering to consider is that in which signals are ordered by FOSD so that $F^*(\theta|s) \leq F^*(\theta|\hat{s})$ for all $s \geq \hat{s}$.[21]

Extending the model in this fashion now requires separate treatment for the monopoly and perfect competition market conditions.

6.1 Monopolist's Multi-Tariff Menu

The monopolist's problem is to maximize expected profits $E[P(s,\theta) - C(q(s,\theta))]$ subject to incentive and participation constraints. There are now two sets of incentive constraints which must be satisfied. In the second period, given any initial choice of tariff s, it must be incentive compatible for consumers to choose the intended allocation $q(s,\theta)$. Moreover, in the first period it must be incentive compatible for consumers to reveal their signal s by choosing the intended tariff. The monopolist's problem is formally stated and analyzed in Web Appendix D both for two first-period signals $s \in \{L, H\}$ and for a continuum of first-period signals.

6.1.1 Two-Tariff Menu

If there are only two first period signals, $s \in \{L, H\}$, a monopolist will offer a menu of two tariffs $\{q_L(\theta), P_L(\theta)\}$ and $\{q_H(\theta), P_H(\theta)\}$. I characterize the optimal menu in Web Appendix D.1. The approach is similar to that for the single-tariff menu. After formally stating the monopolist's profit maximization problem, I substitute incentive and participation constraints in place of payments. As in the single-tariff case, second period local-incentive compatibility pins down marginal prices. Fixed fees are given by two binding constraints: type L's participation constraint and the downward incentive constraint that type H should not want to deviate and choose the tariff L.

[21] Note that the assumed ordering applies to the consumers' beliefs F^*, rather than the firm's beliefs F. The ordering is required to make incentive constraints tractable, and these depend on consumer beliefs.

Beyond the familiar non-negativity, satiation, and monotonicity constraints, substituting constraints for payments leaves only the upward incentive constraint that type L should not want to deviate and choose tariff H. In their analysis of optimal refund contracts, Courty and Li (2000) solve a related problem by first relaxing the upward incentive constraint, and then checking that it is satisfied. Unfortunately, overconfidence implies that the upward incentive constraint is likely to bind. A more productive approach in this context is to incorporate the upward incentive constraint as a state variable and endpoint constraint in an optimal control framework. Doing so leads to the following characterization of marginal prices.

Proposition 3 *The equilibrium payment functions* $P_L^*(q) = P_L^*(\theta(q, L))$ *and* $P_H^*(q) = P_H^*(\theta(q, H))$ *are continuous and piece-wise smooth functions of quantity. There may be kinks in the payment functions where marginal price increases discontinuously. These kinks occur where a monotonicity constraint binds and an interval of types "pool" at the same quantity on the given tariff. For quantities at which there is no pooling, marginal prices are given by equations (9-10) where* $\gamma \geq 0$ *is the shadow price of the upward incentive constraint and* $\alpha = \Pr(H)$*:*

$$\begin{aligned} \frac{dP_L^*(q)}{dq} &= \max\left\{0, C_q(q) + V_{q\theta}(q,\theta)\frac{F_L^*(\theta) - F_L(\theta)}{f_L(\theta)} + \frac{\gamma+\alpha}{1-\alpha}V_\theta(q_L,\theta)\frac{F_L^*(\theta) - F_H^*(\theta)}{f_L(\theta)}\right\} \quad (9) \\ \textit{where } \theta &= \theta(q, L) \end{aligned}$$

$$\begin{aligned} \frac{dP_H^*(q)}{dq} &= \max\left\{0, C_q(q) + V_{q\theta}(q,\theta)\frac{F_H^*(\theta) - F_H(\theta)}{f_H(\theta)} - \frac{\gamma}{\alpha}V_\theta(q_H,\theta)\frac{F_L^*(\theta) - F_H^*(\theta)}{f_H(\theta)}\right\} \quad (10) \\ \textit{where } \theta &= \theta(q, H) \end{aligned}$$

Proof. Given results in Web Appendix D.1 the proof is analogous to the proof of Proposition 2 and hence omitted. ■

Equations (9-10) show that when the upward incentive constraint is not binding ($\gamma = 0$), tariff-H marginal prices are identical to those in the single tariff case. On the other hand, given FOSD, tariff-L marginal prices are distorted upwards from the single-tariff benchmark. This matches the standard price discrimination intuition in which there is no distortion at the top, but the allocation of lower types is distorted downwards to increase rent extraction from high-types. If the upward incentive-constraint binds ($\gamma > 0$), the upward distortion of tariff-L marginal prices is exacerbated. Further, tariff-H marginal prices are distorted downwards.

When overconfidence, measured by $|F_s^* - F_s|$, is high relative to the ex ante heterogeneity, measured by $|F_L^* - F_H^*|$, the qualitative predictions of the single tariff model are robust. The distortions from first-period screening will tend to reduce the number of included units and increase overage rates for the low-tariff, but have the opposite affect on the high-tariff. (This fits observed cellular phone service tariff menus which involve increasing allowances and declining overage rates across tariffs.) If ex ante heterogeneity is sufficiently large relative to overconfidence, it should be expected that included units would be eliminated altogether from the low-tariff, that the upward incentive constraint would not bind, and hence the high-tariff would set marginal price equal to the single-tariff benchmark.

6.1.2 A Continuum of Tariffs

When the first-period signal has a continuous distribution and the monopolist offers a continuum of tariffs, it can be shown that the monopolist's problem reduces to maximization of an expected virtual surplus (equation 11) subject to a reduced set of constraints (Web Appendix D.2, Proposition 9).

$$\Phi(s,q,\theta) = V(q,\theta) - C(q) - V_\theta(q,\theta)\left\{\frac{1-G(s)}{g(s)}\frac{\frac{\partial}{\partial s}\left[1-F^*(\theta|s)\right]}{f(\theta|s)} + \frac{F^*(\theta|s)-F(\theta|s)}{f(\theta|s)}\right\} \quad (11)$$

The bracketed term in the virtual surplus function now includes the information rent term $\frac{1-G(s)}{g(s)}\frac{\frac{\partial}{\partial s}[1-F^*(\theta|s)]}{f(\theta|s)}$ which arises in Courty and Li's (2000) model, and the fictional surplus term $\frac{F^*(\theta|s)-F(\theta|s)}{f(\theta|s)}$ which arises in the single tariff model with overconfidence. The optimal tariff maximizes virtual surplus subject to non-negativity, satiation, monotonicity ($q(s,\theta)$ non-decreasing in θ), and first period global incentive compatibility. As in Courty and Li's (2000) model, allocation $q(s,\theta)$ non-decreasing in signal s is sufficient but not necessary for first period global incentive compatibility.

Now, however, general examples for which Courty and Li (2000) were able to show the relaxed solution was non-decreasing in signal s may violate this sufficient condition given high enough levels of overconfidence. This is because for high levels of overconfidence, the relaxed solution involves marginal prices near the monopoly price for each particular minute as discussed in Web Appendix B. Since monopoly price increases with demand, this implies types with higher signals should face higher marginal prices at a given quantity, and therefore consume less for a given θ. This is unfortunate, because in these cases it is not known what the optimal tariff will look like.[22]

For specific cases in which the allocation of the relaxed solution is non-decreasing in signal s, marginal price is given by equation (12) if it is well defined, where $\theta = \theta(s,q)$.

[22] Global incentive compatibility would need to be checked directly, and if it failed an ironing procedure could not be used to find the optimal tariff because monotonicity is not necessary.

$$P_q(s,q) = Max\left\{0, C_q\left(q\right) + V_{q\theta}\left(q,\theta\right)\left\{\begin{array}{c} \frac{1-G(s)}{g(s)}\frac{\frac{\partial}{\partial s}[1-F^*(\theta|s)]}{f(\theta|s)} \\ +\frac{F^*(\theta|s)-F(\theta|s)}{f(\theta|s)} \end{array}\right\}\right\} \tag{12}$$

Marginal pricing for the top tariff intended for consumers with signal $\bar{s}$ is the same as in the single tariff model. Marginal prices for all lower tariffs on the menu are distorted upwards by the information rent term. Whether these tariffs continue to offer initial quantities at zero marginal price depends on the relative size of the information rent and fictional surplus. Initial quantities are more likely to be offered at zero marginal price for all tariffs on the menu when overconfidence is high, and ex ante heterogeneity is low.

6.2 Perfectly Competitive Multi-Tariff Menu

Characterizing the equilibrium tariff menu under perfect competition in general is a more difficult problem than the monopoly case, because it is insufficient to solve a single constrained maximization problem. In fact, there may be no set of tariffs which yield non-negative profits such that entry is not profitable. In other words, if firms simultaneously set prices there may not be a pure strategy equilibrium. This is similar to Rothschild and Stiglitz's (1976) result about competitive insurance markets.

In practice, finding the equilibrium set of tariffs for natural examples with strictly increasing costs is often no more difficult than solving the single-tariff model. When costs are strictly increasing, incentive constraints need not bind. In this case pricing of each tariff matches that in the single-tariff model: consumers who receive signal s are offered the tariff described by the single-tariff model as if they were the only type in the market.

The equilibrium tariff menu for one such example is illustrated by Figure 6. This is a variation of the example presented in Section 5.5: As in column 2 of Figure 4, marginal cost c is \$0.035, and consumers are highly overconfident ($\Delta = 0.25$). However, here consumers receive one of three signals ex ante, low, medium, or high, which correspond to future θ being distributed uniformly over the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$, $[0,1]$, or $\left[\frac{1}{2}, \frac{3}{2}\right]$ respectively. This example yields a tariff menu qualitatively similar to cellular phone service tariff menus. Moreover, the predicted usage distributions of customers on each tariff are ordered by strict first order stochastic dominance, which matches actual usage patterns described in Section 8.

In general incentive constraints may bind, and distort marginal prices away from the single-tariff benchmark. A general characterization of multi-tariff menus under perfect competition is left for future research.

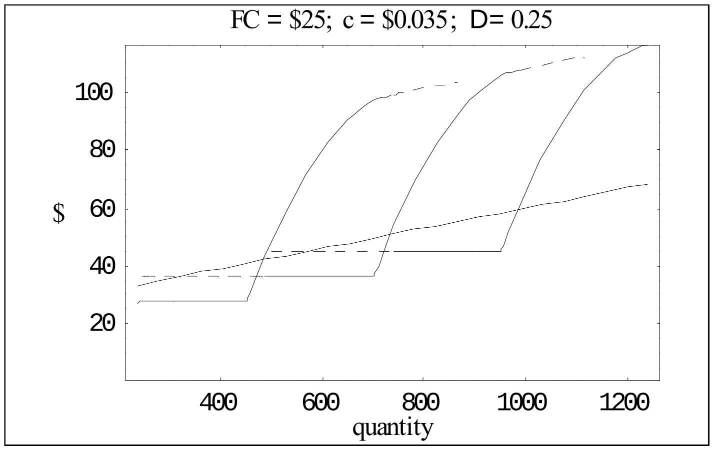

Figure 6: Total pricing for a 3-tariff menu under perfect competition. Solid portions of the tariffs are uniquely optimal. Dashed portions of the tariffs are illustrative extensions where no consumption takes place. The straight line shows total costs.

7 Potential Alternatives

There are several potential alternatives to the model of overconfidence which are worth considering. First, existing literature considers the implications of a number of biases for optimal nonlinear pricing. These include a flat-rate bias as well as biases related to systematically underestimating usage. Moreover, by considering alternative type distributions in the context of the multi-tariff monopoly model of the previous section, I am able to develop an alternative explanation of three-part tariffs based on price discrimination with common priors. I explore each of these three alternatives below.

7.1 Flat Rate Bias

Several authors have documented a "flat-rate bias". This term refers to a tendency of consumers to choose a flat-rate tariff despite the availability of a metered tariff which would be cheaper given their usage levels (Train 1991). Lambrecht and Skiera (2006) provide an overview of the work on flat-rate bias, document the bias among internet service customers in Germany, and identify three significant causes: risk aversion, demand overestimation, and the "taxi-meter" effect.[23]

Although the existence of a flat-rate bias may influence the terms of three-part tariffs, it does not provide a good explanation for their use. It is true that three-part tariffs are locally flat over the allowance of included units, but globally three-part tariffs are not flat. In fact, for the 1,484 cellular

[23] The "taxi-meter" explanation is that prices directly enter consumer preferences: Consumers derive less pleasure from units of consumption that accrue marginal charges, than those that are prepaid with a fixed fee.

customers in my sample who chose a three-part tariff, overages occur on 19% of bills. Conditional on occurrence, this leads to average overage charges more than twice the average monthly fixed-fee (See Section 8).

If consumers are risk averse to such variability in their monthly bill, the use of three-part tariffs with steep overage charges is more surprising rather than less surprising. Similarly, a tendency of consumers to overestimate their future demand does not explain three-part tariff pricing. Proposition 2 was derived without any assumption on the relationship between consumer and firm priors, hence it can be used to analyze biases other than overconfidence. Corollary 2 confirms that demand overestimation conflicts with steep overage rates observed on three-part tariffs.

Corollary 2 *If consumers over-estimate their future demand in a FOSD sense ($F^*(\theta) \leq F(\theta)$) then marginal price is between zero and marginal cost at all quantities.*

Proof. Follows from Proposition 2. ■

7.2 Underestimates

Gabaix and Laibson (2006) examine optimal pricing of goods with add-ons (e.g. printers and printer cartridges) when some consumers are myopic or are unaware of the need to purchase the add-on. When sophisticated consumers can substitute away from the add-on through advanced planning, myopic consumers subsidize low prices of the primary good by paying high add-on fees. In related work, DellaVigna and Malmendier (2004) examine optimal two-part tariff pricing when consumers are quasi-hyperbolic discounters. They show that marginal prices should be set below marginal cost for investment goods (such as health clubs) and above marginal cost for leisure goods (such as cellular phone service). DellaVigna and Malmendier (2004) mention that this may explain why cell phone tariffs include marginal prices above marginal cost, but this theory does not explain why marginal prices are initially zero.

Aside from the different welfare implications, in the context of the model in this paper, both naive beta-delta discounters and consumers unaware of future add-on purchases are essentially consumers who systematically underestimate their demand at the time of contracting. By Corollary 3, marginal price would therefore be predicted to be above marginal cost for all q.

Corollary 3 *If consumers under-estimate their future demand in a FOSD sense ($F^*(\theta) \geq F(\theta)$) then marginal price is weakly greater than marginal cost for all quantities. Moreover, where the FOSD is strict, marginal price will be strictly greater than marginal cost.*

Proof. Follows from Proposition 2. ■

The assumption in this paper that consumers are overconfident (A*) implies that consumers underestimate demand conditional on it being high ($\theta > \theta^*$), but overestimate demand conditional on it being low ($\theta < \theta^*$). The underestimation of demand above θ^* drives marginal price above marginal cost at high quantities, as would naive beta-delta discounting. It is the overestimation of demand below θ^* that drives the region of zero marginal price at low quantities. Thus the balanced over and underestimation of demand captured by overconfidence is necessary for the result.

7.3 Price Discrimination with Common Priors

Under perfect competition, any rational model of cellular phone service pricing with common priors yields marginal cost pricing, which cannot explain observed tariffs. Determining whether price discrimination with common priors can explain observed tariffs under monopoly or imperfect competition is not a trivial problem, however. Considering the multi-tariff monopoly model discussed in Section 6.1 for the special case of common priors ($F^*(\theta|s) = F(\theta|s)$) provides useful insight (See Web Appendix E).

First, if the distribution of demand is increasing in a first order stochastic dominance (FOSD) sense, as a consumer's signal s increases, then marginal price should always be above marginal cost and consumption distorted downwards for all but those with the highest signal $\bar{s}$. Given such a type distribution, price discrimination with common priors would therefore not explain observed tariffs.

Second, given low marginal costs and free disposal, price discrimination with common priors could predict tariff menus qualitatively similar to those observed which couple increasing fixed fees with increasing numbers of included minutes and declining overage rates. However, to do so a rather implausible type distribution must be assumed. In particular, consumers' conditional priors over θ should satisfy equation (13) for some cutoff $\theta^*(s)$ increasing in s.

$$\frac{\partial}{\partial s}\left(1 - F(\theta|s)\right) \begin{cases} \leq 0 & \theta \leq \theta^*(s) \\ > 0 & \theta > \theta^*(s) \end{cases} \tag{13}$$

To understand why this type distribution generates such pricing, consider an example with two ex ante types. The high signal ($s = H$) type is an undergraduate whose valuation is high on average, but is also highly variable. The undergraduate is either on campus and has a high demand, or is away on break and has a low demand. The low signal ($s = L$) type is a graduate student who consistently has a moderate demand somewhere in between these two extremes. In this case, a monopolist will find it optimal to offer the undergraduate user unlimited usage at marginal cost for a high monthly fee. The graduate student will pay a low monthly fee for low marginal charges at low quantities followed by high marginal charges at high quantities. The high marginal charges at high quantities have little impact on either an undergraduate on break or a graduate student, but make the graduate student tariff much less attractive to an undergraduate on campus. The initial low marginal charges are attractive to the graduate student, and allow a higher monthly fee to be charged on the undergraduate tariff. This trade-off is a wash for an undergraduate on campus, but is unattractive to an undergraduate on break. Together, both distortions of the graduate student tariff away from marginal cost pricing increase the surplus that can be extracted from an undergraduate ex ante.

For two tariffs with $Q_1 < Q_2$ included minutes, marginal prices are zero on both tariffs for $q \in (0, Q_1)$. Thus assumptions about the distribution of demand for consumers on each plan map directly onto conclusions about distributions of consumption up to Q_1. A type distribution described by equation (13) therefore requires[24] that consumers selecting a tariff with $Q_2 > Q_1$ included minutes would be more likely to consume strictly less than Q_1 minutes than would consumers who actually selected the tariff with Q_1 included minutes. More specifically, it requires that the cumulative usage distribution of consumers choosing plan 1 be below that of consumers choosing plan 2, for all $q < Q_1$: $H(q|s_1) \leq H(q|s_2)$. This is implausible, and as shown in the following section, is not consistent with observed consumer behavior. As a result, the common-prior model does not appear to explain observed cellular phone service tariff menus.

8 Empirical Analysis

I have obtained billing data for 2,332 student accounts managed by a major US university for a national US cellular phone service provider.[25] The data span 40 of the 41 months February 2002 through June 2005 (December 2002 is missing), and include 32,852 individual bills. Within the data set there are several different menus of tariffs. For example, at any given time there are national calling plans, local calling plans, and a two-part tariff offered. Moreover, the menus offered differ over time. As a result, customers within my sample are on more than 50 distinct plans from more than 10 menus.

For my primary analysis, I focus on two similar menus with the most usage data. These are the

[24] Consumers who realized $\theta \leq \theta^*(s)$ would consume weakly below their included limit $Q = q^*(s, \theta^*(s))$, and consumers who realized $\theta > \theta^*(s)$ might make overages.

[25] Students received an itemized phone bill, mailed by default to their campus residence, which was separate from their university tuition bill. It is true that the sample of students is undoubtedly different than the entire cellular-phone-service customer-base. However, a pricing manager from one of the top US cellular phone service providers who kindly read through an earlier draft of this paper made the unsolicited comment that the empirical findings were highly consistent with their own internal analysis of much larger and representative customer samples.

set of local plans offered to students in the fall of 2002 and the fall of 2003 (Figure 7).[26] Within these menus I look at the four most popular plans. These include three-part tariffs with the smallest, second smallest, and third smallest monthly fixed-fees and included minutes, which I will refer to as plans 1, 2, and 3 respectively. While nearly 60% of students who signed up for a new tariff in the fall of 2002 or 2003 chose either plan 1, 2, or 3, an important alternative was a two-part tariff, which I call plan 0.[27] Plan 0 has a small monthly fixed-fee and a constant per-minute charge below the overage rates of plans 1-3.

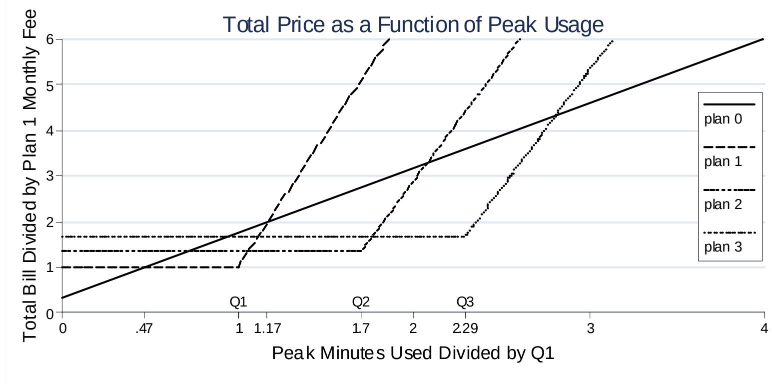

Figure 7: Total price as a function of peak usage for plans 0, 1, 2, and 3 (fall 2002 & 2003 menus). Usage is normalized by Q1 so that usage level 2 corresponds to twice as many minutes as are included in plan 1. Similarly, price is normalized so that bill level 2 corresponds to double the plan 1 monthly fee.

The overconfidence and price discrimination explanations of three-part tariff pricing can be distinguished by comparing customer usage patterns across different three-part tariffs on the same menu. Between 2002 and 2003, plans 1, 2, and 3 change only slightly, so I pool the usage data from the two menus.[28] Figure 8 plots the cumulative usage distributions $\hat{H}(q|plan)$ and their 95% confidence intervals[29] for customers on plans 1, 2, and 3. Bills for incomplete months of service in

[26] Most students sign up for cellular phone service at the beginning of the academic year, which is why the fall menus are most relevant. The fall 2002 menu was offered September 2002 through March 2003. Almost the same menu was offered again September 2003 through December 2003.

[27] Plan 0 was not offered to the general public, but only to the students who received service through the university. In the US it is typical for such two-part tariffs to be included in corporate rate packages, but not be offered to the general public. Students received additional benefits including up to 15% additional included minutes on plans, and a required service commitment of only 3 months rather than 12 months.

[28] Between 2002 and 2003 fall menus, minute allowances increased by 1-2%, and overage rates increased by 0-14%. Analyzing usage patterns separately for 2002 and 2003 yields similar results.

[29] If $\hat{H}(q)$ denotes the sample cumulative density function (CDF) for N observations, a 95% confidence interval is

which the monthly access fee and included minute limit were prorated are excluded, as are bills with missing usage information. In total the distribution plotted for plan 1 is based on 5,008 bills of 498 customers, while plan 2 is based on 2190 bills of 210 customers, and plan 3 is based on 283 bills of 31 customers.

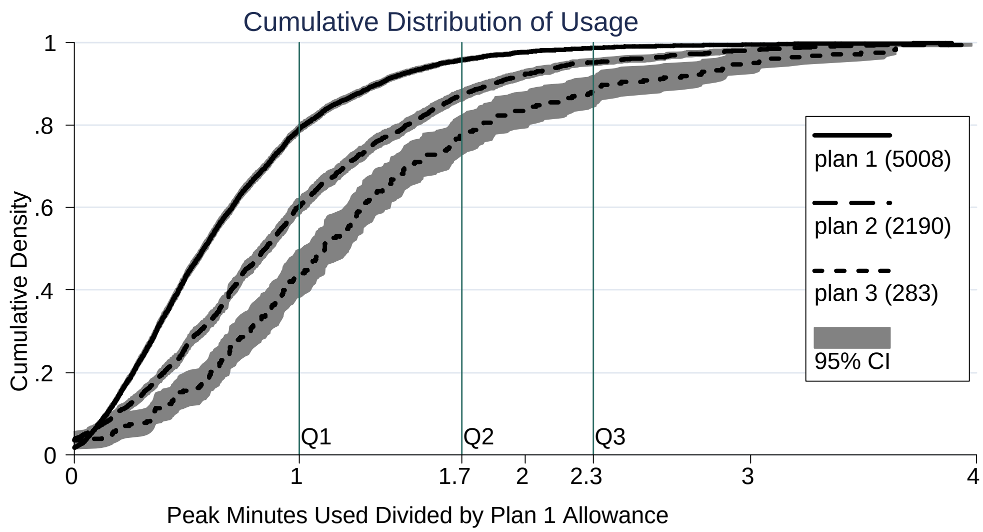

Figure 8: Cumulative usage distributions $\hat{H}(q|plan)$ and their 95% confidence intervals for customers on Plans 1, 2, and 3. Usage is normalized by the plan 1 allowance so that usage level 2 corresponds to twice the number of minutes that are included with plan 1. Vertical lines are drawn at the plan allowances. (Fall 2002 and 2003 menus.)

Figure 8 shows that the three usage distributions are statistically indistinguishable at the very bottom, and the very top, but everywhere else the distributions are consistent with strict a FOSD ordering. Formal pair-wise tests of first order stochastic dominance between the three distributions provide limited additional insight.[30] It is clear from the figure, however, that usage patterns are inconsistent with the assumption driving the common-prior alternative.

It is not the case that $\hat{H}(q|plan1) \leq \hat{H}(q|plan2)$ for $q \leq Q_1$. Customers choosing plan 2 are not

calculated point-wise as $\hat{H}(q) \pm 1.96\sqrt{\frac{\left(1-\hat{H}(q)\right)\hat{H}(q)}{N}}$. This is because for large N, $\hat{H}(q)$ is approximately normal with mean of the true CDF $H(q)$ and variance $\frac{(1-H(q))H(q)}{N}$.

[30]Barrett and Donald's (2003) test fails to reject the null hypothesis of FOSD for each pair at any reasonable significance level. Yet, because the distributions are statistically indistinguishable at the top and bottom, the KRS test Tse and Zhang (2004) describe, which is based on Kaur, Rao and Singh (1994), fails to reject the complementary null hypothesis for each pair at a 10% significance level. The DD test Tse and Zhang (2004) describe, which is based on Davidson and Duclos (2000), rejects the null hypothesis of distribution equality at a 1% significance level and accepts the first alternative hypothesis that the distributions have a FOSD ordering. (This test was based on 20 points equally spaced in the range of the plan 1 usage distribution using a critical value from Stoline and Ury (1979).)

"undergraduate" types who actually consume less than Q_1 minutes more frequently than plan 1 customers. Rather, plan 2 customers consume less than Q_1 minutes only 57% of the time, whereas customers choosing plan 1 consume less than Q_1 minutes 78% of the time. Similar comparisons with usage by plan 3 customers all fall out the same way. Therefore, in contrast to the overconfidence explanation, the alternative price discrimination model cannot simultaneously explain both observed pricing and observed usage patterns.

One might be concerned that the model of overconfidence is off the mark if one believes that customers only rarely exceed their included minutes. It is reasonable to hypothesize that observed tariffs are actually designed with the expectation that the included minutes serve as rather strict limits on usage, and that the typical overage rates of 35 to 45 cents are designed to be prohibitive outside of emergency situations. The model of overconfidence presented in this paper, however, explicitly incorporates the idea that many consumers will be surprised by higher demand than expected and use more than the included number of minutes. (Figure 10 in Web Appendix C illustrates a usage distribution predicted by the model for the example discussed in Section 5.5).

	Observations		(Usage / Allowance)	
	n	n/N	mean	std. dev.
Under Allowance	6176	83%	0.47	0.27
Over Allowance	1305	17%	1.45	0.48
Total	7481	100%	0.64	0.49

Table 1: Average usage as a fraction of included allowance across plans 1, 2, and 3 (fall 2002 & 2003 menus).

The data clearly show that overages are an important feature of customer behavior. This is apparent in Figure 8 and made explicit in Table 1. While 83% of the time customers on plans 1-3 do not exceed their allowance, using only half of included minutes on average, the other 17% of the time they exceed their allowance, by an average of nearly 50%. Moreover, overages are an important source of firm revenue. Within the entire data set, there are 18,064 individual bills from 1,484 unique customers who are on a tariff with a strictly positive number of included minutes. Within this sample, 19% of bills contain overages. Moreover, the average overage charge is 44% of the average monthly fixed-fee (229% conditional on an overage occurring), and represents 23% of average revenues (excluding taxes). In this regard, the model presented in this paper is consistent with customer behavior.

Consumer tariff choices indirectly reveal something about consumers' initial expectations for future usage. Comparing initial tariff choice with subsequent usage is therefore informative about consumer overconfidence. Plans 0-3 on the fall 2002 menu are identical in all dimensions other than peak usage pricing described by Figure 7, and in particular all offered free night and weekend calling. In fall 2003, plan 0 no longer offered free night and weekend calling, and is therefore less comparable to the three-part tariffs. As a result, I focus on the fall 2002 menu for the following analysis.

Plans 1 and 2 are cheaper than plan 0 only for a relatively narrow range of consumption: between 47% and 117% of Q1 for plan 1 and between 41% and 122% of Q2 for plan 2 (Figure 7). The fact that consumers signed up for plans 1 and 2 initially, implies that they believed their consumption would likely fall within these bounds. In fact, bills of plan 1 and 2 customers fall outside these bounds, both above and below, roughly half of the time. As a result, a large fraction of consumers make ex post "mistakes", in the sense that cumulatively over the duration of these customers' tenure in the data with a chosen plan, an alternative plan would have been lower cost for the same usage.

Table 2 shows that at least 65% of plan 1 customers would have saved money by initially choosing plan 0, and would have saved an average of 68% of the plan 1 monthly fixed-fee.[31] In fact, had all customers who chose plan 1 chosen plan 0 instead, consumers would have saved an average of 19% of the plan 1 monthly fixed-fee. Figures are similar although smaller for plan 2 customers. On the other hand, only 5% of plan 0 customers would have saved money by choosing plans 1, 2, or 3.

	Plan 0 Customers	Plan 1 Customers	Plan 2 Customers
Customers	393 (59%)	92 (15%)	124 (21%)
Alternative Considered	Plan 1, 2, or 3	Plan 0	Plan 0
Alternative Lower Cost Ex Post	5%	65%	50%
Average Saving*	25%	68%	53%

*Per month, conditional on occurrence as a percentage of Plan 1 monthly fixed-fee.

Table 2: Frequency and size of ex post "mistakes" (fall 2002 menu).

The prevalence and size of ex post mistakes show that consumers are uncertain about their future demand when making tariff choices, and that modeling this uncertainty is critical for understanding the market. Moreover, the direction of the mistakes described above provide additional evidence of consumer overconfidence. Finally, customers who quit or switch plans in less than 6 months

[31] Pro-rated months are excluded from the calculation. The frequency and size of mistakes are both underestimated. First, plan 0 includes unlimited free in-network calling, which plans 1-3 do not. I am able to correctly calculate the counter-factual cost of plans 1-3 to a plan 0 customer, but I overestimate the counter-factual cost of plan 0 to a plan 1 or 2 customer because I cannot distinguish in-network from out-of-network calls made by plan 1-3 users. Second, I do not account for the fact that customers could alter usage if enrolled in an alternative plan, making any potential switch more attractive. Moreover, if the entire choice set of plans are considered as possible alternatives, the frequency and size of ex post mistakes is substantially higher.

make more and larger mistakes than those who stay with their chosen plan longer.[32] This implies that customers are learning about their demand, and therefore, uncertainty will be greatest for new customers.

9 Extensions

9.1 Heterogeneous Overconfidence

Consider the case where fraction α of consumers share the firm's prior $F(\theta)$, while fraction $(1-\alpha)$ are overconfident and have an alternate prior $F^*(\theta)$. Under perfect competition, firms offer correct-prior types and overconfident types separate tariffs identical to those offered when consumers are homogeneous. Types with correct priors are offered price equal to cost $\left\{q^{FB}(\theta), C\left(q^{FB}(\theta)\right)\right\}$, just as if all consumers had correct beliefs. Overconfident types are offered the equilibrium tariff $\left\{q^*(\theta), P^C(\theta)\right\}$ characterized in Section 5, just as if all consumers were overconfident.

Note that the firm earns the same profit on a given tariff from a participating consumer with correct prior and a participating overconfident consumer. This is because both have the same true type distribution $F(\theta)$ from the firm's perspective, and therefore the same allocation $q(\theta)$ and payment $P(\theta)$ distributions. Thus the constraint set of zero-profit tariffs is the same regardless of whether a firm is serving consumers with correct priors or overconfident consumers.

Because the allocations q^{FB} and q^* each maximize the perceived expected utilities of their intended consumers over the same constraint set, consumers cannot believe themselves to be strictly better off under the other types' tariff. Consumers with correct priors prefer the standard tariff since $E\left[S^{FB}\right] \geq E\left[S^*\right]$, and overconfident types prefer the overconfident tariff since $E\left[\Psi^*\right] \geq E\left[\Psi^{FB}\right]$.

Under monopoly, there is one simple benchmark case where incentive compatibility of the homogenous tariffs is similarly guaranteed. Whenever overconfident types correctly estimate the surplus generated by the first best allocation:[33] $E^*\left[S^{FB}\right] = E\left[S^{FB}\right]$, in equilibrium correct-prior and overconfident types will be offered the same monopoly tariffs as if they were each the only type in the market: the standard tariff $\left\{q^{FB}(\theta), C\left(q^{FB}(\theta)\right) + E\left[S^{FB}\right]\right\}$ and the overconfident tariff $\left\{q^*(\theta), P^M(\theta)\right\}$ respectively.

Table 3 gives each party's perceived monopoly payoff under the standard and overconfident

[32] For instance, 75% of plan 1 customers who switch or quit in less than 6 months would have saved money on plan 0, but only 58% of longer term plan 1 customers would have saved money on plan 0. Moreover, the average potential savings for the two groups are respectively 102% and 35% of the plan 1 fixed monthly fee.

[33] When marginal costs are zero, the benchmark $E^*\left[S^{FB}\right] = E\left[S^{FB}\right]$ implies consumers correctly estimate their expected value of consuming up to their satiation points.

	$E[U]$	$E^*[U]$	$E[\Pi]$
Standard Tariff $\left\{q^{FB}, C+E\left[S^{FB}\right]\right\}$	0	$E\left[\Psi^{FB}\right]-E\left[S^{FB}\right]$	$E\left[S^{FB}\right]$
Overconfident Tariff $\left\{q^*, P^M\right\}$	$E[S^*]-E[\Psi^*]$	0	$E[\Psi^*]$

Table 3: Monopoly Payoffs

tariffs. The assumption $E^*\left[S^{FB}\right] = E\left[S^{FB}\right]$ ensures that overconfident types are indifferent between the two tariffs, since $E^*\left[S^{FB}\right] = E\left[\Psi^{FB}\right]$.[34] Further, by Lemma 3, it guarantees that consumers with correct priors will weakly prefer the first best tariff.

In general, however, first period incentive constraints may bind. In example 1, for instance, the overconfident type's true expected monopoly payoff is strictly positive. Consumers with correct priors would then prefer the overconfident tariff over the first best allocation with all surplus extracted. In equilibrium it should be expected that the overconfident tariff would be distorted towards marginal cost pricing in order to improve surplus extraction from types with correct priors, but that both types would earn strictly positive payoffs.

9.2 Underconfidence

Psychology literature documents the hard-easy effect:[35] While individuals are overconfident when making difficult predictions, they are actually underconfident when making simple predictions (Lichtenstein et al. 1982). Given the analysis in Section 5, pricing implications for underconfident consumers come at no extra cost. Simply reverse A^* by assuming that $F(\theta)$ crosses $F^*(\theta)$ once from below at θ^*, and denote this assumption A'.

Corollary 4 *Given A', for quantities at which there is no pooling: (1) If marginal cost is zero for all q then:*

$$P_q^*(q) > 0 \quad , \; q \in \left(\underline{q}, Q\right)$$
$$P_q^*(q) = 0 \quad , \; q \in (Q, \bar{q}) \cup \left\{\underline{q}, Q, \bar{q}\right\}$$

[34] See proof of Lemma 3 in Appendix A.

[35] It should be noted that the hard-easy effect has been documented for binary predictions such as, "Is London or Sydney more populous?" rather than continuous predictions such as "How far is it between London and Sydney?" which are relevant here. Moreover, some authors have called into question the validity of results documenting the hard-easy effect for such binary predictions (Juslin, Winman and Olsson 2000).

(2) If marginal cost is strictly positive for all q then:

$$\begin{aligned} P_q^*(q) &= C_q(q) > 0 \quad , \ q \in \{\underline{q}, Q, \bar{q}\} \\ P_q^*(q) &> C_q(q) > 0 \quad , \ q \in (\underline{q}, Q) \\ C_q(q) &> P_q^*(q) \geq 0 \quad , \ q \in (Q, \bar{q}) \end{aligned}$$

Proof. Follows directly from Proposition 2, assumption A', and $q^*(\theta)$ non-decreasing. ■

Corollary 4 shows that when marginal costs are zero, marginal price will initially be positive but will fall to zero above some threshold Q. When marginal costs are strictly positive, marginal price will fall below marginal cost above some threshold Q, and may fall all they way to zero at higher quantities, but will be positive at the top quantity $\bar{q}$.

This may explain the existence of loyalty programs such as that examined by Hartmann and Viard (2006) in which those who join a golf-club loyalty program receive a free or discounted game after playing ten. This implements a quantity discount without asking golfers to commit to play 11 games in advance. Relative to a fixed membership with games priced at marginal cost, this pricing scheme looks attractive to a golfer who believes she will probably either find a nicer course soon, and not play many games, or love the course and play a lot. The golf club will be happy to offer the scheme if the golfer ends up playing some intermediate number of games. In this case the golfer overvalues avoiding a high membership fee because she overestimates the probability of only playing a few games. The golfer also overestimates the value of the loyalty program, because she overestimates the probability of playing enough to earn the reward.

Of course there are already several good explanations of loyalty programs in the literature (see Hartmann and Viard (2006) for an overview). Frequent flyer programs are likely designed as kickbacks to exploit principal agent problems. Loyalty programs may be used to create artificial switching costs (Klemperer 1995). Hartmann and Viard (2006) also discuss how loyalty programs may be used for price-discrimination, and improve upon standard quantity discounts when there are reasons for consumers not to commit to future consumption in advance.

10 Conclusion

This paper has shown that given overconfident consumers, low marginal costs, and free disposal, optimal pricing involves included minutes at zero marginal price. This provides a promising explanation for the three-part tariff menus observed in the cellular phone services market. Empirical evidence shows that consumer usage patterns are consistent with the explanation, and in particular, that ex post "mistakes" by consumers are consistent with the underlying assumption of overconfidence. Although an alternative common-prior explanation exists, it appears to be inconsistent with consumer usage patterns. The model presented here is broadly applicable beyond cellular services, and is potentially relevant in any market in which consumers commit to a contract while they are uncertain about their eventual demand. In particular, the model can explain the use of three-part tariffs for other communication services such as internet access, a range of rental services, consumer credit card debt, and an increasing number of other services. Finally, if consumers are underconfident rather than overconfident, the model can explain the implementation of quantity discounts through loyalty programs.

11 Appendices

A Proofs

A.1 Proof of Lemma 1

Proof. (1). Monopoly: Consider any optimal tariff $\{q(\theta), P(\theta)\}$. If there is a type $\hat{\theta}$ who is offered $q(\hat{\theta}) > q^S(\hat{\theta})$, a monopolist would be weakly better off offering $q(\hat{\theta}) = q^S(\hat{\theta})$ instead. Consumer $\hat{\theta}$'s equilibrium consumption, and hence incentive and participation constraints would be unaffected. Any other consumer θ now finds it weakly less attractive to deviate by claiming type $\hat{\theta}$, so consumer θ's choice remains incentive compatible. Production costs would be weakly reduced however. This argument can be repeated for any $\hat{\theta}$, hence the tariff $\left\{\min\left\{q(\theta), q^S(\theta)\right\}, P(\theta)\right\}$ must be weakly more profitable, and therefore still optimal. Moreover, if costs are strictly increasing, lowering $q(\hat{\theta})$ to $q^S(\hat{\theta})$ for a positive measure of types $\hat{\theta}$ would strictly increase profits. (2) Perfect Competition: The argument is similar. Now however, reducing production costs relaxes the firm participation constraint, allowing the firm to reduce $P(\theta)$ by a fixed amount and improve consumers' perceived expected utility $E^*[U(\theta)]$ without affecting incentive constraints. ■

A.2 Proof of Proposition 1

Proposition 1 may be restated as follows: (1) If quantity payment pair $\left\{q^*(\theta), P^C(\theta)\right\}$ solves the perfect competition problem defined in Section 5.1, then the pair $\left\{q^*(\theta), P^C(\theta) + E[\Psi(q^*(\theta), \theta)]\right\}$ solves the monopoly problem defined in Section 5.1. Conversely, if the pair $\left\{q^*(\theta), P^M(\theta)\right\}$ solves the monopoly problem, then $\left\{q^*(\theta), P^M(\theta) - E[\Psi(q^*(\theta), \theta)]\right\}$ solves the perfect competition problem.

(2) If a quantity payment pair $\{q(\theta), P(\theta)\}$ solves either the monopoly or perfect competition problems stated in Section 5.1, then the pair $\{q^*(\theta), P(\theta)\}$, where $q^*(\theta) = \min\left\{q(\theta), q^S(\theta)\right\}$, solves the same problem. Moreover, $q^*(\theta)$ maximizes expected virtual surplus $E[\Psi(q(\theta), \theta)]$ subject to monotonicity, non-negativity, and satiation constraints, where virtual surplus is defined by equation (3), and $P(\theta)$ is given by equation (4) for perfect competition or equation (5) for monopoly. Conversely, if $q^*(\theta)$ maximizes expected virtual surplus subject to monotonicity, non-negativity, and satiation constraints, then there exists a unique pair of payments $\left\{P^C(\theta), P^M(\theta)\right\}$ such that $\left\{q^*(\theta), P^C(\theta)\right\}$ solves the perfect competition problem and $\left\{q^*(\theta), P^M(\theta)\right\}$ solves the monopoly problem. These payments are given by equations (4) and (5) respectively.

(3) If $q^*(\theta)$ maximizes expected virtual surplus subject to monotonicity, non-negativity, and satiation constraints, and the pair $\{q^*(\theta), P(\theta)\}$ is incentive compatible, then marginal price is

given by equation (6).

Note that the proof differs slightly from the outline in the text, in that I apply Lemma 1 at the end, rather than the beginning.

Proof. Let consumption of type θ who claims to be type $\hat{\theta}$ be $q^c(\theta,\hat{\theta}) \equiv \min\{q(\hat{\theta}), q^S(\theta)\}$ and consumption of an honest type θ be $q^c(\theta) \equiv q^c(\theta,\theta)$. Let $\Pi(\theta) \equiv P(\theta) - C(q(\theta))$ denote the firm's profit from serving a consumer who reports type θ.

1. By the standard approach, global incentive compatibility can be replaced with a local incentive constraint and a monotonicity condition. Note that $\frac{\partial}{\partial\theta}U(\theta,\hat{\theta}) = V_\theta\left(q^c(\theta,\hat{\theta}),\theta\right)$, because $V_q\left(q^S(\theta),\theta\right) = 0$ by the definition of $q^S(\theta)$. As a result, global incentive compatibility and application of an envelope theorem (e.g. Milgrom and Segal (2002) Theorem 2) implies that $U'(\theta) = V_\theta(q^c(\theta),\theta)$ almost everywhere and $U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} V_\theta(q^c(z),z)\,dz$. Further, global incentive compatibility and increasing differences $V_{q\theta} > 0$ implies that consumption $q^c(\theta)$ will be non-decreasing in θ. These two conditions are also sufficient for global incentive compatibility for the standard reason.

2. Applying the local incentive compatibility condition and integrating by parts implies that the true expected utility from the firm's perspective and the consumers' perceived expected utility may be expressed as given by equations (14) and (15) respectively.

$$E[U(\theta)] = U(\underline{\theta}) + E\left[V_\theta(q^c(\theta),\theta)\frac{1-F(\theta)}{f(\theta)}\right] \tag{14}$$

$$E^*[U(\theta)] = U(\underline{\theta}) + E\left[V_\theta(q^c(\theta),\theta)\frac{1-F^*(\theta)}{f(\theta)}\right] \tag{15}$$

3. The participation constraints must bind in both problems. For any allocation $q(\theta)$ and implied consumption quantity $q^c(\theta)$, there is a unique payment function $P(\theta)$ which satisfies both local incentive compatibility and the relevant participation constraint with equality. This payment function can be found first by expressing payments in terms of consumer utility: $P(\theta) = U(\theta) - V(q^c(\theta),\theta)$. Next, by applying local incentive compatibility $U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} V_\theta(q^c(z),z)\,dz$ which pins down consumer utility, and therefore payments, up to a constant $U(\underline{\theta})$:

$$P(\theta) = V(q^c(\theta),\theta) - \int_{\underline{\theta}}^{\theta} V_\theta(q^c(z),z)\,dz - U(\underline{\theta}) \tag{16}$$

   Finally, binding participation constraints determine the constant $U(\underline{\theta})$. Under monopoly $E^*[U(\theta)] = 0$, so by equation (15) $U(\underline{\theta}) = -E\left[V_\theta(q(\theta),\theta)\frac{1-F^*(\theta)}{f(\theta)}\right]$. Note that expected firm profits are always equal to the difference between expected surplus and the consumers' true expected utility: $E[\Pi(\theta)] = E[S(q(\theta),\theta)] - E[U(\theta)]$. Hence under perfect competition $E[\Pi(\theta)] = 0$ implies $E[U(\theta)] = E[S(q(\theta),\theta)]$, and so by equation (14) $U(\underline{\theta}) = E\left[V(q^c(\theta),\theta) - C(q(\theta)) - V_\theta(q^c(\theta),\theta)\frac{1-F(\theta)}{f(\theta)}\right]$. As a result monopoly and perfect competition payments are given by equations (17) and (18):

$$P^M(\theta) = V(q^c(\theta),\theta) - \int_{\underline{\theta}}^{\theta} V_\theta(q^c(z),z)\,dz + E\left[V_\theta(q^c(\theta),\theta)\frac{1-F^*(\theta)}{f(\theta)}\right] \tag{17}$$

$$\begin{aligned} P^C(\theta) \quad = \quad & V(q^c(\theta),\theta) - \int_{\underline{\theta}}^{\theta} V_\theta(q^c(z),z)\,dz \\ & -E\left[V(q^c(\theta),\theta) - C(q(\theta)) - V_\theta(q^c(\theta),\theta)\frac{1-F(\theta)}{f(\theta)}\right] \end{aligned} \tag{18}$$

4. By substituting the unique payment function $P^M(\theta)$ or $P^C(\theta)$ implied by the allocation $q(\theta)$, local incentive compatibility, and the relevant (binding) participation constraint in place of $P(\theta)$ in each problem, the problems reduce to maximizations over allocation $q(\theta)$ subject to non-negativity and monotonicity. Note that for all possible allocations, and not just the optimal allocation, the implicit payment functions guarantee local incentive compatibility and binding participation. Hence, in the reduced monopoly problem, consumers' perceived expected utility $E^*[U(\theta)]$ is a constant equal to zero for all allocations, and can be added to the objective function without altering the solution. In this case, the monopoly objective function becomes: $E[\Pi(\theta)] + E^*[U(\theta)]$. Similarly, in the reduced perfect-competition problem, $E[\Pi(\theta)]$ is a constant equal to zero for all allocations. Hence it can be added to the objective function without altering the solution. In this case, the perfect-competition objective function becomes: $E[\Pi(\theta)] + E^*[U(\theta)]$. This shows that the objective functions are the same in both reduced problems. Since expected firm profits are always equal to the difference between expected surplus and the consumers' true expected utility, the objective functions can be written as:

$$E[V(q^c(\theta)) - C(q(\theta))] + E^*[U(\theta)] - E[U(\theta)] \tag{19}$$

Equations (14) and (15) imply that the fictional surplus is:

$$E^*[U(\theta)] - E[U(\theta)] = E\left[V_\theta(q^c(\theta),\theta)\frac{F(\theta) - F^*(\theta)}{f(\theta)}\right] \tag{20}$$

5. By Lemma 1, and the refinement $q(\theta) \leq q^S(\theta)$, I can replace $q^c(\theta)$ with $q(\theta)$ in the objective

function defined by equations (19) and (20) as well as in the expected utility and payment expressions in equations (14-18) by making the same substitution in the monotonicity constraint and adding a satiation constraint $q(\theta) \leq q^S(\theta)$ to the simplified maximization problems. This substitution completes the proof of Proposition 1 parts 1 and 2. Following the substitution, differentiating equation (16) for $P(\theta)$ and making a change of variables gives the final result:

$$\frac{d}{dq}P(q) = V_q(q, \theta(q))$$

This is valid for quantities at which there is no pooling, where $\theta(q)$ is a well defined function.

■

A.3 Small Lemma 4

The satiation quantity $q^S(\theta)$ must be characterized prior to solving for the equilibrium allocation $q^*(\theta)$ as it is the upper bound of the constraint set. Moreover, both satiation $q^S(\theta)$ and first best $q^{FB}(\theta)$ quantities are important benchmarks to which $q^*(\theta)$ is compared in the paper. Lemma 4 therefore gives relevant properties of each.

Lemma 4 *1. Satiation $q^S(\theta)$ and first best $q^{FB}(\theta)$ quantities are continuously differentiable, strictly positive, and strictly increasing.*

2. Satiation quantity is higher than first best quantity, and strictly so when marginal costs are strictly positive at q^{FB}.

$$q^S(\theta)\begin{cases} = q^{FB}(\theta) & C_q\left(q^{FB}(\theta)\right) = 0 \\ > q^{FB}(\theta) & C_q\left(q^{FB}(\theta)\right) > 0 \end{cases}$$

Proof. Given maintained assumptions, $q^S(\theta)$ and $q^{FB}(\theta)$ exist, and are continuous functions. Moreover they are everywhere characterized by the first order conditions $V_q\left(q^S(\theta), \theta\right) = 0$ and $S_q\left(q^{FB}(\theta), \theta\right) = 0$ respectively. The implicit function theorem implies $q^S(\theta)$ and $q^{FB}(\theta)$ are both continuously differentiable with derivatives:

$$\frac{d}{d\theta}q^S = -\frac{V_{q\theta}}{V_{qq}}$$

$$\frac{d}{d\theta}q^{FB} = -\frac{V_{q\theta}}{V_{qq} - C_{qq}}$$

It then follows that both are strictly increasing. Further, when $C_q\left(q^{FB}(\theta)\right)=0$, the first order condition $S_q\left(q^{FB}(\theta),\theta\right)=0$ implies that $V_q\left(q^{FB}(\theta),\theta\right)=0$. This is simply the first order condition for $q^S(\theta)$ which must be satisfied, so $q^S(\theta)=q^{FB}(\theta)$. When $C_q\left(q^{FB}(\theta)\right)>0$, the first order condition $S_q\left(q^{FB}(\theta),\theta\right)=0$ implies that $V_q\left(q^{FB}(\theta),\theta\right)>0$. Since V is concave in q, this implies that $q^S(\theta)>q^{FB}(\theta)$. ■

A.4 Proof of Lemma 2

Proof. Under maintained assumptions, the constraint set $D(\theta)=\left[0,q^S(\theta)\right]$ is convex and compact valued, continuous, and non-empty. Further, virtual surplus $\Psi(q,\theta)$ is continuous and strictly concave in q:

$$\Psi_{qq}(q,\theta)=\underbrace{V_{qq}(q,\theta)}_{<0}-\underbrace{C_{qq}(q)}_{\geq 0}+\underbrace{V_{qq\theta}(q,\theta)}_{=0}\frac{F(\theta)-F^*(\theta)}{f(\theta)}<0$$

Therefore q^R is a continuous function. (Note that this is where the stricter assumption $V_{qq\theta}=0$ is used rather than the standard assumption $V_{qq\theta}\leq 0$.)

Since $\Psi(q,\theta)$ is twice continuously differentiable and strictly concave in q for all θ, $q^R(\theta)$ is characterized by the first order condition $\Psi_q(q,\theta)=0$ unless either the non-negativity or free disposal constraints are binding.

Over the interior of any interval for which q^R is characterized the FOC $\Psi_q\left(q^R,\theta\right)=0$ and F^* is continuously differentiable, then by the implicit function theorem q^R is continuously differentiable with derivative $\frac{d}{d\theta}q^R=-\frac{\Psi_{q\theta}}{\Psi_{qq}}$. Since $q^S(\theta)$ is continuously differentiable (Lemma 4) this implies that $q^R(\theta)$ is piecewise smooth. Kinks may occur when q^R hits the constraints 0 or q^S, or a kink in F^* when the constraints are not binding. ■

A.5 Proof of Proposition 2

Proof. (1) Following Proposition 1, at non-pooling quantities marginal price is equal to $V_q(q,\theta(q))$. Thus at pooling quantity q, marginal price must increase discontinuously since above and below q marginal price is given by $V_q(q,\inf\{\theta(q)\})$ and $V_q(q,\sup\{\theta(q)\})$ respectively and by assumption $V_{q\theta}>0$.

(2) Further, at non-pooling quantities the non-negativity and monotonicity constraints are not binding and $q^*(\theta)=q^R(\theta)$ by Lemma 2 part 2. Thus by Lemma 2 part 1, either (i) the equilibrium allocation is characterized by the first order condition $\Psi_q(q,\theta)=0$, or (ii) the satiation constraint is binding and marginal price is zero since the satiation quantity is characterized by $V_q\left(q^S(\theta),\theta\right)=0$.

In the former case, the first order condition can be solved for $V_q(q,\theta)$, and therefore marginal price:

$$V_q(q,\theta) = C_q(q) + V_{q\theta}(q,\theta)\frac{F^*(\theta) - F(\theta)}{f(\theta)}$$

When $\left[C_q(q) + V_{q\theta}(q,\theta(q))\frac{F^*(\theta(q))-F(\theta(q))}{f(\theta(q))}\right]$ is negative, the first order condition $\Psi_q = 0$ implies $V_q(q,\theta)$ is negative and therefore $q \geq q^S(\theta)$. This is precisely when the satiation constraint binds, ensuring marginal price to be weakly positive. Thus marginal price is equal to $\left[C_q(q) + V_{q\theta}(q,\theta(q))\frac{F^*(\theta(q))-F(\theta(q))}{f(\theta(q))}\right]$ whenever that quantity is positive, and zero otherwise.

(3) Since $\theta(q)$ is a continuous function at non-pooling quantities, marginal price is as well. Thus payment $P^*(\theta(q))$ is continuously differentiable at non-pooling quantities. Moreover, incentive compatibility requires that types who pool at the same quantity pay the same price. Thus $P^*(\theta(q))$ is well defined and continuous at pooling quantities. ■

A.6 Proof of Lemma 3

Proof. Given the characterization of q^* in Proposition 1, expected virtual surplus must be weakly higher under the equilibrium allocation than under the first best allocation: $E[\Psi^*] \geq E\left[\Psi^{FB}\right]$. Moreover, under marginal cost pricing expected profits are equal to the fixed fee regardless of the prior over θ. Thus under the first best allocation, the expected virtual surplus is equal to the perceived expected surplus: $E\left[\Psi^{FB}\right] = E^*\left[U^{FB}\right] + E\left[\Pi^{FB}\right] = E^*\left[U^{FB}\right] + E^*\left[\Pi^{FB}\right] = E^*\left[S^{FB}\right]$. Together this implies that $E[\Psi^*] \geq E^*\left[S^{FB}\right]$. The assumption $E^*\left[S^{FB}\right] \geq E\left[S^{FB}\right]$ therefore implies that the firm is better off: $E[\Psi^*] \geq E\left[S^{FB}\right]$. This in turn implies that consumers are worse off since total welfare is lower. ■

References

Arkes, Hal R., "Overconfidence in judgemental forecasting," in J. S. Armstrong, ed., *Principles of Forecasting*, Boston: Kluwer Academic Publishers, 2001, pp. 495–515.

Baron, D. P. and D. Besanko, "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1984, *1* (3).

Barrett, G.F. and S.G. Donald, "Consistent tests for stochastic dominance," *Econometrica*, 2003, *71* (1), 71–104.

Bolger, Fergus and Dilek Onkal-Atay, "The effects of feedback on judgmental interval predictions," *International Journal of Forecasting*, 2004, *20* (1), 29–39.

Courty, Pascal and Hao Li, "Sequential screening," *The Review of Economic Studies*, 2000, *67* (4), 697–717.

Davidson, R. and J.Y. Duclos, "Statistical inference for stochastic dominance and for the measurement of poverty and inequality," *Econometrica*, 2000, *68* (6), 1435–1464.

DellaVigna, Stefano and Ulrike Malmendier, "Contract design and self-control: Theory and evidence," *The Quarterly Journal of Economics*, 2004, *119* (2), 353–402.

Edlin, A.S. and C. Shannon, "Strict monotonicity in comparative statics," *Journal of Economic Theory*, 1998, *81* (1), 201–219.

Fudenberg, Drew and Jean Tirole, *Game theory*, Cambridge, Mass.: MIT Press, 1991.

Gabaix, Xavier and David Laibson, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

Griffin, Dale and Amos Tversky, "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology*, 1992, *24*, 411–435.

Hartmann, Wesley R. and Brian V. Viard, "Do Frequency Reward Programs Create Switching Costs?," Working Paper August 22 2006.

Juslin, P., A. Winman, and H. Olsson, "Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect," *Psychological Review*, 2000, *107* (2), 384–396.

Kaur, A., B.L.S.P. Rao, and H. Singh, "Testing for Second-Order Stochastic Dominance of two Distributions," *Econometric Theory*, 1994, *10* (5), 849–866.

Klemperer, P., "Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade," *The Review of Economic Studies*, 1995, *62* (4), 515–539.

Lambrecht, Anja and Bernd Skiera, "Paying Too Much and Being Happy About It: Existence, Causes, and Consequences of Tariff-Choice Biases," *Journal of Marketing Research*, 2006, *43* (2), 212–223.

——, **Katja Seim, and Bernd Skiera**, "Does Uncertainty Matter? Consumer Behavior under Three-Part Tariffs," Working Paper 2005.

Leonard, Daniel and Ngo van Long, *Optimal control theory and static optimization in economics*, Cambridge ; New York: Cambridge University Press, 1992.

Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence D. Phillips, "Calibration of probabilities: The state of the art to 1980," in Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under uncertainty : heuristics and biases*, Cambridge ; New York: Cambridge University Press, 1982, pp. 306–334.

Loewenstein, G., T. O'Donoghue, and M. Rabin, "Projection bias in predicting future utility," *The Quarterly Journal of Economics*, 2003, *118* (4), 1209–1248.

Milgrom, Paul R. and Ilya Segal, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 2002, *70* (2), 583–601.

Miravete, Eugenio J., "Screening consumers through alternative pricing mechanisms," *Journal of Regulatory Economics*, 1996, *9* (2), 111–132.

——, "Choosing the wrong calling plan? Ignorance and learning," *American Economic Review*, 2003, *93* (1), 297–310.

——, "The Welfare Performance of Sequential Pricing Mechanisms," *International Economic Review*, 2005, *46* (4), 1321–1360.

Mirrlees, J. A., "An Exploration in Theory of Optimum Income Taxation," *The Review of Economic Studies*, 1971, *38* (114), 175–208.

Mussa, Michael and Sherwin Rosen, "Monopoly and Product Quality," *Journal of Economic Theory*, 1978, *18* (2), 301–317.

Rochet, Jean-Charles and Lars Stole, "The Economics of Multidimensional Screening," in M. Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in economics and econometrics: theory and applications - eighth world Congress*, Vol. 36 of Econometric Society Monographs, New York: Cambridge University Press, 2003.

Rothschild, Michael and Joseph E. Stiglitz, "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 1970, *2* (3), 225–243.

—— **and** ——, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, 1976, *90* (4), 629–649.

Seierstad, Atle and Knut Sydsæter, *Optimal Control Theory with Economic Applications* Advanced Textbooks in Economics, New York: North-Holland, 1987.

Stole, Lars A., "Nonlinear Pricing and Oligopoly," *Journal of Economics and Management Strategy*, 1995, *4* (4), 529–62.

Stoline, M.R. and H.K. Ury, "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons Among Means," *Technometrics*, 1979, *21* (1), 87–93.

Topkis, Donald M., "Minimizing a Submodular Function on a Lattice," *Operations Research*, 1978, *26*, 305–321.

Train, Kenneth E., *Optimal Regulation: The Economic Theory of Natural Monopoly*, Cambridge, MA: MIT Press, 1991.

Tse, Y.K. and X.B. Zhang, "A Monte Carlo investigation of some tests for stochastic dominance," *Journal of Statistical Computation and Simulation*, 2004, *74* (5), 361–378.

Wilson, Robert B., *Nonlinear Pricing*, New York, NY: Oxford University Press, 1993.