Selling to Overconfident Consumers

Citation: Grubb, Michael D. 2009. "Selling to Overconfident Consumers." American Economic Review, 99(5): 1770–1807. DOI:10.1257/aer.99.5.1770

As Published: http://dx.doi.org/10.1257/aer.99.5.1770

Publisher: American Economic Association

Persistent URL: http://hdl.handle.net/1721.1/52652

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Selling to Overconfident Consumers*

Michael D. Grubb
MIT Sloan School of Management
Cambridge, MA 02142
mgrubb@mit.edu
www.mit.edu/~mgrubb

May 2, 2008

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, Simon Board, Carlos Corona, Liran Einav, Erik Eyster, Bob Gibbons, Richard Holden, Peter Lorentzen, Greg Rosston, Brian Viard, Bob Wilson, and seminar participants at Stanford, Berkeley, Northwestern, MIT, Princeton, Yale, Columbia, UCLA, Caltech, Harvard, and Penn. 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 plans 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 often include a mileage allowance and charge per mile thereafter.[1] In a variety of rental markets, contracts charge a flat rate for a specified period followed by steep late fees. Finally, introductory credit card offers may also be 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 nonlinear pricing does not provide a compelling explanation for such pricing patterns. For perfect competition one expects prices to be driven down to cost, while standard nonlinear pricing models suggest the highest demand consumer will pay 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

[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."

firms 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 units offered at zero marginal price, followed by positive marginal prices for additional units. This result holds not only under monopoly, but also under perfect competition. Furthermore, while overconfidence always reduces total surplus, it may increase consumer welfare and reduce monopoly profits. Predicted prices may be fully nonlinear, however a three-part tariff is always a good approximation when overconfident consumers are primarily uncertain about the volume of desirable units, relative to a fairly consistent value for units that are desirable.

I extend the model to allow for ex ante heterogeneity and screening at the contracting stage via a menu of multiple tariffs. I characterize a monopolist's optimal two-tariff menu given two ex ante types. The qualitative pricing results of the single tariff model are robust as long as overconfidence is sufficiently high relative to ex ante heterogeneity in average demand. 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 three 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 consider 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. As 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. Section 4 presents and analyzes the single-tariff model, and Section 5 extends the analysis to multiple-tariff menus (focusing on a menu of two tariffs). Potential alternative explanations are described in Section 6, and then tested empirically in Section 7. Finally, Section 8 concludes. (Supplementary material referred to in the paper is contained in a Web Appendix that is available for download from my website: www.mit.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 steep marginal charges. Standard nonlinear pricing models (Mussa and Rosen 1978, Maskin and Riley 1984, Wilson 1993) predict marginal cost pricing for the last unit, and higher marginal prices for all lower quantities. They 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. Several papers explicitly model two-stage screening in which agents choose a contract using a signal about their preferences, but later learn their true preferences before making a quantity choice (Baron and Besanko 1984, Riordan and Sappington 1987, Miravete 1996, Courty and Li 2000, Miravete 2005). Although none of this research specifically addresses three-part tariff pricing, this branch of the nonlinear pricing literature is a natural reference point.

In particular, 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

[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 does not explain 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 7). This undermines the interpretation that a menu of three-part tariffs is an implementation of Mussa and Rosen's (1978) optimal fully-nonlinear tariff, and other explanations based on weak optimality (e.g. Oi (1971) and Jensen (2006)), since these all predict that every customer consumes at or beyond his or her included allowance.

predict pricing qualitatively similar to a single three-part tariff. Moreover, both the multi-tariff monopoly model with overconfidence presented in Section 5 and the price discrimination model with common priors discussed in Section 6 are closely related to Courty and Li (2000). These two models differ from Courty and Li (2000) by incorporating continuous demand with declining marginal valuations and free disposal, and either by adding overconfidence (Section 5) or by considering alternative type distributions (Section 6).

I incorporate overconfidence by relaxing the common prior assumption. There are a few related papers which also investigate firm pricing when consumers have biased priors about their future demand. Eliaz and Spiegler (2008) consider a model where consumers have biased priors but only two types of ex post demand: high or low, in contrast to the continuous demand case studied here. The difference is important because although Eliaz and Spiegler (2008) can capture consumers who over or under estimate average demand, they cannot capture overconfident consumers who underestimate the likelihood of both upper and lower tails of demand by overweighting the center of the distribution. Allowing for more than two ex post types is also crucial (along with free disposal and consumer satiation) to generate three-part tariff pricing.[4] (Eliaz and Spiegler's 2006 and 2008 papers are similar, but the earlier analysis examines dynamically inconsistent consumers.)

Uthemann (2005) considers a model that like mine has continuous ex post demand, but unlike mine (and similar to Eliaz and Spiegler (2008)), does not allow for free disposal or satiation. As a result, the optimal tariffs under both monopoly and competition are qualitatively different than those in this paper, and in particular have strictly positive marginal prices everywhere.[5] Finally, Sandroni and Squintani (2004) characterize equilibrium pricing in a competitive Rothschild and Stiglitz (1976) insurance market where a fraction of customers underestimate their risk of suffering a loss. A similarity between our models is that, unlike Eliaz and Spiegler (2008) or Uthemann (2005), we allow firms' beliefs about the "true" distribution of demand to vary across consumers as well as for heterogeneity in consumers' own beliefs. However, like Eliaz and Spiegler (2008), Sandroni and Squintani (2004) assume that there are only two consumer types ex post - in this case those who suffer a loss and those who do not.

There are a number of other related papers which consider the effect of alternate consumer biases or non-standard preferences on optimal nonlinear pricing (DellaVigna and Malmendier 2004, Oster

[4] Technically, one needs at least three types to separate overconfidence from over-optimism. Similarly, one needs many ex-post equilibrium quantities to pin down uniquely the shape of the optimal contract.

[5] On the technical side, an important difference is that Uthemann (2005) only characterizes the solution to a relaxed problem without providing sufficient conditions on model primitives to guarantee that it coincides with the solution to the full problem. As I show in Web Appendix C, the two solutions do not coincide when consumer overconfidence is high.

and Scott Morton 2005, Esteban and Miyagawa 2005, Gabaix and Laibson 2006, Esteban, Miyagawa and Shum 2007). I discuss several of these as potential explanations of observed pricing patterns in Section 6.

2.2 Overconfidence

Loewenstein et al. (2003) present a variety of laboratory and survey evidence demonstrating the prevalence of projection bias, and Conlin, O'Donoghue and Vogelsang (2007) document projection bias in the field using catalog sales and returns data. 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 experimental evidence shows that individuals are overconfident about the precision of their own predictions when making difficult[6] forecasts (e.g. Lichtenstein, Fischhoff and Phillips (1982)). In other words, individuals tend to set overly narrow confidence intervals relative to their own confidence levels. A typical psychology study 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. A typical finding is that the true answer lies outside a subject's 98% confidence interval about 30% to 40% of the time. Ex post tariff-choice "mistakes" made by cellular phone customers are consistent with such overconfidence, as documented in Section 7.

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.[7] 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 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.

[6] 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.

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

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.[8] 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 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,

[8] 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.

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 4.8).

4 Single-Tariff Analysis

4.1 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 $q(\theta')$ at $t = 2$.

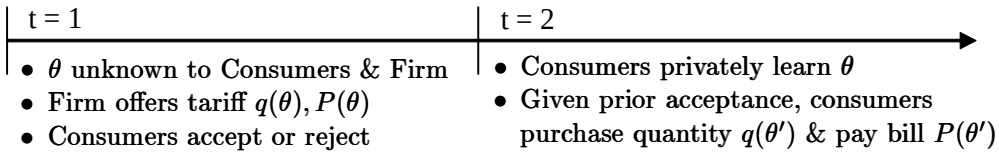

Figure 1: Time Line

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: $V(0, \theta) = 0$.

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 consumer type θ who purchases $q(\theta')$ units will only consume the minimum of $q(\theta')$ and $q^S(\theta)$, and will receive consumption value $V(\min\{q^S(\theta), q(\theta')\}, \theta)$.[9]

[9] The satiation point $q^S(\theta)$ is the point at which type θ's (strictly decreasing) marginal value of consumption becomes negative. Finite satiation implies that consumers need not purchase all units with zero marginal price. Rather than explicitly allowing for free disposal, it would have been equivalent to assume directly that consumers

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[10] 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*:[11] *$F^*(\theta)$ 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.

Within the context of this model, the equilibrium tariff, or allocation and payment pair $\{\hat{q}(\theta), \hat{P}(\theta)\}$, 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 $\left[\underline{\theta}, \overline{\theta}\right]$, a range which includes the support of consumers' prior $F^*(\theta)$.

---

have value function $\widetilde{V}(q,\theta) = V\left(\min\left\{q, q^S(\theta)\right\}, \theta\right)$ for which the marginal value of consumption is zero beyond the finite point $q^S(\theta)$.

[10] 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*).

[11] 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. Conversely, 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.

Finally assumption A* is closely related to second order stochastic dominance (SOSD). Let μ^* and μ be the mean values of distributions F^* and F respectively. Given A*, F^* SOSD F if and only if $\mu^* \geq \mu$ (Hanoch and Levy 1969, Theorem 3).

Finally, I restrict the partial derivative $V_{qq\theta}$ by equation (1), where $q^{FB}(\theta) \equiv \arg\max_q [V(q,\theta) - C(q)]$ denotes the first best allocation. This ensures that the virtual surplus function described in Proposition 1 is strictly quasi-concave. Note that when marginal costs are zero (the focus of the paper), consumer satiation coincides with the first best allocation and equation (1) reduces to the standard assumption: $V_{qq\theta} \geq 0$ (e.g. see Fudenberg and Tirole (1991) Chapter 7).

$$V_{qq\theta}(q,\theta) \begin{cases} \geq 0, & 0 < q < q^{FB}(\theta) \\ \leq 0, & q^{FB}(\theta) < q < q^{S}(\theta) \end{cases} \tag{1}$$

Parallel demand curves ($V_{qq\theta} = 0$) are a special case of equation (1) for which it is 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 $V(q^c,\theta) = v(q^c) + q^c\theta$. Moreover, consumers who correctly predict the mean of θ also correctly predict their mean value of each unit. Another special case of equation (1) are demand curves that are translated horizontally by θ and are concave in quantity if marginal costs are zero, or are concave below the first best allocation and convex above if constant marginal costs are strictly positive. Unlike parallel demand curves, these can capture preferences close to $V(q^c,\theta) = p\min\{q^c,\theta\}$ which describe consumers who have a relatively consistent and certain value for desirable units, but are uncertain about the quantity desired.

4.2 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:

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

such that:

1. Global IC $\quad U(\theta,\theta) \geq U(\theta,\theta') \quad \forall \theta,\theta' \in \left[\underline{\theta},\bar{\theta}\right]$
2. Consumer Participation[12] $\quad E^*[U(\theta)] \geq 0$

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

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

constraint applies: the utility $U(\theta,\theta') \equiv V(\min\{q^S(\theta), q(\theta')\},\theta) - P(\theta')$ of a consumer of type θ who reports θ' 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 $\min\{q^S(\theta), q(\theta')\}$ rather than the purchased quantity $q(\theta')$. Second, contracting occurs ex ante at which time consumers' beliefs differ from those of the firm. 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 $\quad U(\theta,\theta) \geq U(\theta,\theta') \;\; \forall \theta,\theta' \in \left[\underline{\theta},\bar{\theta}\right]$
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, as otherwise there would be an opportunity for profitable entry. In contrast, under monopoly firm payoff is maximized subject to consumer participation.

4.3 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, as 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, incentives constraints of other consumers would have been weakly relaxed, and the firm could have reduced production costs. This result is stated formally in Lemma 1.

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

Proof. Follows from free disposal. See Appendix A. ■

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 $\min\{q(\theta), q^S(\theta)\}$, 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.[13] 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. (This is true because consumers are homogeneous at the time of contracting.) 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 2). Moreover, having substituted local incentive compatibility and participation constraints in place of payments,

[13] This payment function can be found first by expressing payments in terms of consumer utility: $P(\theta) = V\left(\min\{q(\theta), q^S(\theta)\}, \theta\right) - U(\theta)$. 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 $U(\underline{\theta})$.

fictional surplus is given by equation (3).

$$E\left[\Psi\left(q\left(\theta\right),\theta\right)\right]=E\left[S\left(q\left(\theta\right),\theta\right)\right]+E^{*}\left[U\left(\theta\right)\right]-E\left[U\left(\theta\right)\right] \tag{2}$$

$$E^{*}\left[U\left(\theta\right)\right]-E\left[U\left(\theta\right)\right]=E\left[V_{\theta}\left(q\left(\theta\right),\theta\right)\frac{F\left(\theta\right)-F^{*}\left(\theta\right)}{f\left(\theta\right)}\right] \tag{3}$$

When consumers and the firm share the same prior ($F^{*}\left(\theta\right)=F\left(\theta\right)$) fictional surplus is zero, so the equilibrium tariff maximizes expected surplus $E\left[S\left(q\left(\theta\right),\theta\right)\right]$. This implies first best allocation and marginal payment equal to marginal cost. 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 $\hat{q}\left(\theta\right)=q^{M}\left(\theta\right)=q^{C}\left(\theta\right)$, 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.[14] Proposition 1 summarizes these results precisely.

Proposition 1 *Under both monopoly and perfect competition:*

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

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

$$\Psi\left(q,\theta\right)\equiv V\left(q,\theta\right)-C\left(q\right)+V_{\theta}\left(q,\theta\right)\frac{F\left(\theta\right)-F^{*}\left(\theta\right)}{f\left(\theta\right)} \tag{4}$$

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

$$P^{M}\left(\theta\right)=V\left(\hat{q}\left(\theta\right),\theta\right)-\int_{\underline{\theta}}^{\theta}V_{\theta}\left(\hat{q}\left(z\right),z\right)dz+E\left[V_{\theta}(\hat{q}\left(\theta\right),\theta)\frac{1-F^{*}\left(\theta\right)}{f\left(\theta\right)}\right] \tag{5}$$

$$P^{C}\left(\theta\right)=P^{M}\left(\theta\right)-E\left[\Psi\left(\hat{q}\left(\theta\right),\theta\right)\right] \tag{6}$$

[14] 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 $V\left(q,\theta,d\right)=V\left(q,\theta\right)-d$). 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).) The extension is straight forward as, apart from their brand preferences, consumers are homogeneous at the time of contracting.

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

4.4 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 part 2 of Proposition 1.

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

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

Proof. Part 1 is true because equation (1) ensures Ψ is strictly quasi-concave in q. The proof of part 2 is omitted as it closely follows ironing results for the standard screening model. (See Appendix A.) ■

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

When overconfidence is limited, and consumers' beliefs are close to those of the firm, the relaxed solution will be strictly increasing. However, when consumers are extremely overconfident, the relaxed solution will violate the monotonicity constraint (Web Appendix C, Proposition 6). This is because the relaxed solution is distorted upwards (weakly above first best) below θ^*, but downwards (below first best) above θ^* (Web Appendix C, Proposition 5). When overconfidence is sufficiently high, the distortions in opposing directions are large enough either side of θ^* that the relaxed solution must be strictly decreasing at θ^*. Thus to avoid excluding interesting cases, Web Appendix C completes the characterization of ironed allocations (Lemma 4).

4.5 Pricing Implications

Together, Proposition 1 and Lemma 2 characterize the equilibrium allocation $\hat{q}(\theta)$ consumed and payment $\hat{P}(\theta)$ paid by each type θ. Rather than using a direct revelation mechanism, however,

in practice this tariff will be implemented by setting price as a function of quantity. Let $\hat{\theta}(q) \equiv \inf\{\theta : \hat{q}(\theta) = q\}$. Then (with a slight abuse of notation) the tariff will be implemented by: $\hat{P}(q) = \hat{P}(\hat{\theta}(q))$. It is now possible to draw implications about pricing using Proposition 1.

Proposition 2 *The equilibrium payment $\hat{P}(q)$ 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, because the monotonicity constraint does not bind, marginal price is given by:*

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

Proof. See Appendix A. ■

As 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 $[\hat{P}_q(q) - C_q(q)]$ is equal to the sign of $[F^*(\theta) - F(\theta)]$ except when $F^*(\theta) < F(\theta)$ and marginal cost is zero, as then marginal price is also zero. This is informative about equilibrium pricing, as 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:

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

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:*

$$\begin{aligned}
&\hat{P}_q(q) = 0 \quad , \; q \in \left(\underline{q}, Q\right) \cup \left\{\underline{q}, Q, \bar{q}\right\} \\
&\hat{P}_q(q) > 0 \quad , \; q \in (Q, \bar{q})
\end{aligned}$$

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

$$\begin{aligned}
&\hat{P}_q(q) = C_q(q) > 0 \quad , \; q \in \left\{\underline{q}, Q, \bar{q}\right\} \\
&C_q(q) > \hat{P}_q(q) \geq 0 \quad , \; q \in \left(\underline{q}, Q\right) \\
&\hat{P}_q(q) > C_q(q) > 0 \quad , \; q \in (Q, \bar{q})
\end{aligned}$$

Proof. Follows directly from Proposition 2, assumption A*, and $\hat{q}(\theta)$ 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. (In contrast to observed tariffs, the model predicts that marginal price is always equal to marginal cost for the last unit sold. However, this result is not robust to ex ante heterogeneity. See Section 5 and Figure 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 overestimate the probability of consuming initial units. Thus a firm can overcharge consumers for the first Q minutes 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. Overconfident consumers also underestimate the probability of extremely high usage. Thus a firm cannot extract consumers' true value for 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 usage above Q. The drawback of such marginal fees is that they inefficiently distort consumption below first best, which limits their optimal size.

Absent free disposal, the firm could do better by setting negative marginal prices on initial units in order to charge consumers for purchasing less than anticipated. However free disposal ensures that such negative marginal pricing is futile at best, and strictly suboptimal when marginal costs are positive. Consumers can costlessly avoid underusage penalties by purchasing negative marginal price units and disposing of them. It is free disposal that makes zero a sharp lower bound for marginal price in equation (8). (Web Appendix B provides additional intuition based on option pricing.)

4.6 Example

The implications of Proposition 2 that are summarized in Corollary 1 are best illustrated with figures from specific examples.

Example 1 *Firms have a fixed cost of* \$25 *and a constant marginal cost of* $c \geq 0$ *per unit. Consumers' inverse demand function is linear in* q *and* θ *(Figure 2):* $V(q,\theta) = \frac{3}{2}q - \frac{3}{2000}q^2 + \frac{3}{2}q\theta$. *The firm and consumers' priors are uniform, centered on* 0 *(Figure 2):* F *is* $U\left[-\frac{1}{2},\frac{1}{2}\right]$ *and* F^* *is*

$U\left[-\frac{1-\Delta}{2}, \frac{1-\Delta}{2}\right]$. *Consumers and the firm both agree that the mean of* θ *is equal to* 0. *The parameter* $\Delta \in [0,1]$ *measures the percent by which consumers underestimate the standard deviation of* θ. *For* $\Delta = 0$, *consumers are not overconfident at all, and share the firm's prior. For* $\Delta = 1$, *consumers are extremely overconfident and believe* $\theta = 0$ *with probability one.*

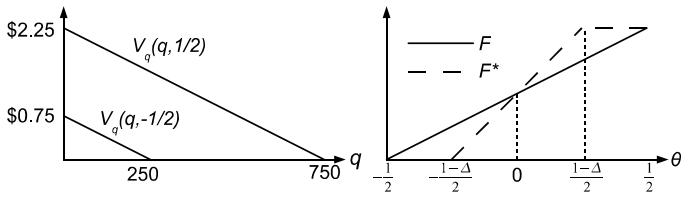

Figure 2: Inverse demand curves and priors in Example 1.

Figure 3 illustrates Corollary 1 given zero marginal costs, using the example described above. In the top row, plots A and B show total prices and total costs under perfect competition. There are three curves: optimal total price (solid line), total cost (dashed line), and the best three-part tariff approximation to optimal pricing (dot-dashed line). In the bottom row, plots C and D show marginal prices and marginal costs, under either perfect competition or monopoly. Both the optimal marginal price and its best three-part tariff approximation are shown. In the left hand column, plots A and C assume low overconfidence, $\Delta = 0.25$, for which there is no pooling. In the right hand column, plots B and D assume high overconfidence, $\Delta = 0.75$, 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. In short, optimal pricing shares important qualitative features of observed three-part tariffs. However the fit is not perfect, because optimal pricing is fully nonlinear. 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 $\bar{q}$.

For comparison, the dot-dashed lines show the best three-part tariff approximations to optimal pricing. In fact, the plots also show the best two-part tariff approximations to optimal pricing since these coincide with the dashed cost curves.[15] Optimal pricing yields a \$0.98 improvement over the

[15] Given linear demand curves and correct estimation of the mean ($E^*[\theta] = E[\theta]$) the optimal two-part tariff marginal price is simply marginal cost. To characterize the best two-part tariff approximation in general, let $q^T(\theta, p) = \{q : V_q(q, \theta) = p\}$ be the quantity chosen by type θ given a two-part tariff marginal price p. Then the first order condition for the optimal p is: $\int_{\underline{\theta}}^{\bar{\theta}} \frac{\Psi_q\left(q^T(\theta,p),\theta\right)}{V_{qq}\left(q^T(\theta,p),\theta\right)} f(\theta)\, d\theta = 0$. To characterize the best three-part tariff

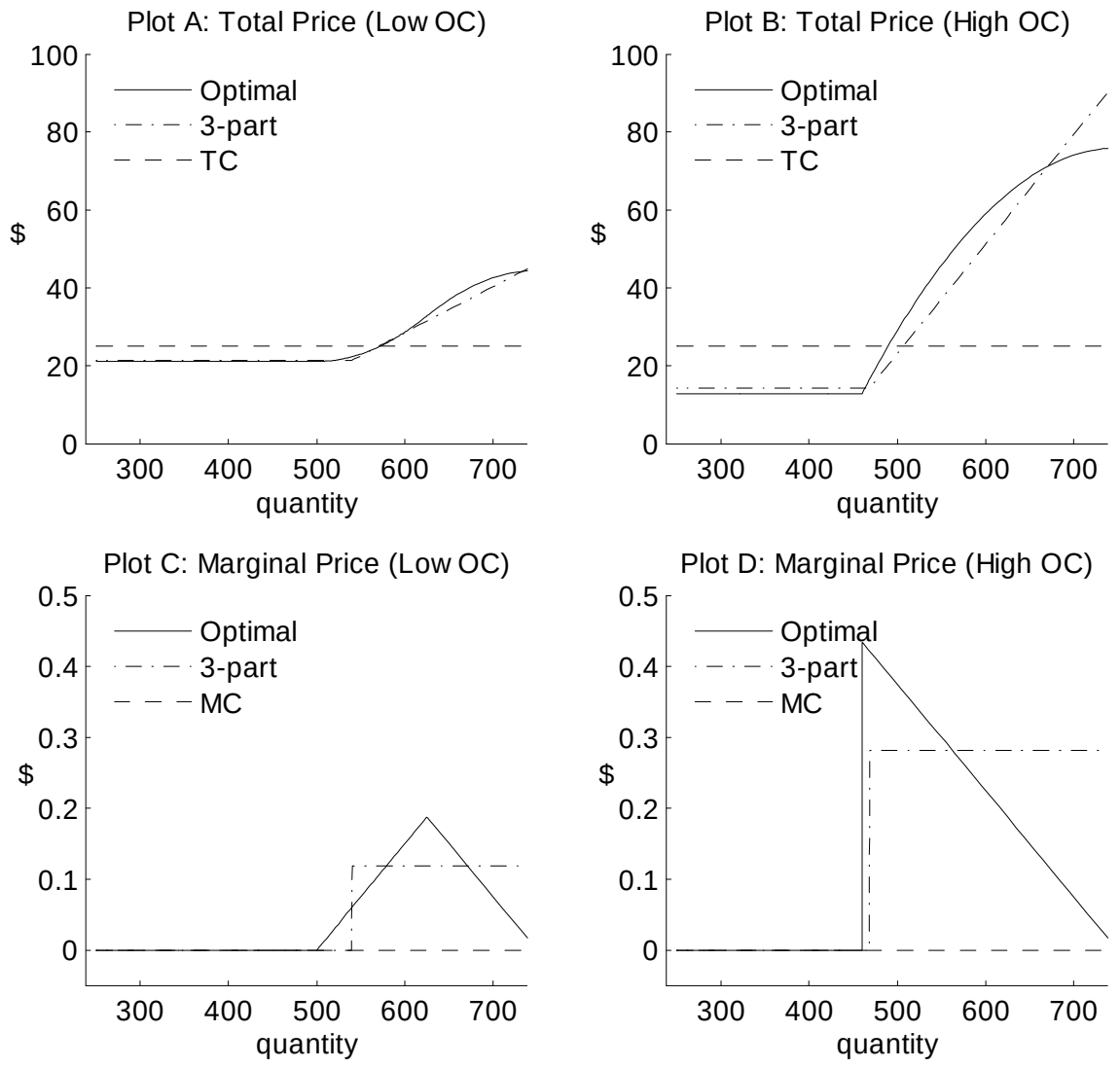

Figure 3: The fully nonlinear equilibrium price and its closest 3-part tariff approximation are depicted under perfect competition and zero marginal cost. Prices are shown for low overconfidence ($\Delta = 0.25$) in the left hand column and for high overconfidence ($\Delta = 0.75$) in the right hand column. Total prices and costs are shown in the top row, and marginal prices and costs are shown in the bottom row.

best two-part tariff given low overconfidence, and a \$7.89 improvement given high-overconfidence. In both cases, the best three-part tariff captures more than 80% of this gain, leaving only a \$0.13 or \$1.30 approximation loss for low-overconfidence and high-overconfidence cases respectively (Table 1). The fact that deviating from marginal cost pricing becomes increasingly valuable with overconfidence is expected, since absent overconfidence marginal cost pricing is optimal.

marginal cost	\$0	\$0	\$0.035	\$0.035
overconfidence	low	high	low	high
2-part loss	\$0.98	\$7.89	\$1.43	\$9.38
3-part loss	\$0.13	\$1.30	\$0.16	\$1.26
$1-\frac{\text{3-part loss}}{\text{2-part loss}}$	87%	84%	89%	87%

Table 1: Two and three-part tariff approximation losses.

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, as marginal cost is strictly positive, marginal price is strictly positive near $\underline{q}$. The best three-part tariff approximations offer similarly small approximation losses relative to the best two-part tariffs (Table 1). 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 might never bind, and marginal price could be strictly positive at all quantities.

4.7 Pricing in Practice

As Figures 3-4 illustrate, the optimal price characterized by Proposition 2 is fully nonlinear: marginal price may vary for each increment sold beyond the included allowance. In contrast, overage rates are constant for observed tariffs. The fact that observed tariffs are simpler than those predicted by a theory which ignores costs of complexity is not surprising. As Wilson (1993) points out, "In practice,... a different price for each increment is differentiation too fine to justify the transaction costs incurred by customers and the firm."[16]

---

approximation, let $\theta(Q,p) = \{\theta : V_q(Q,\theta) = p\}$ be the type who purchases exactly Q units given a marginal price p. The first order conditions for optimal included allowance Q and overage rate p are: $\int_{\theta(Q,0)}^{\theta(Q,p)} \Psi_q(Q,\theta) f(\theta)\, d\theta = \int_{\theta(Q,p)}^{\bar{\theta}} \frac{\Psi_q\left(q^T(\theta,p),\theta\right)}{V_{qq}\left(q^T(\theta,p),\theta\right)} f(\theta) = 0$. (The optimal initial marginal price is zero in these examples.)

[16] Costs arise for at least two reasons: First, a tariff with many marginal prices would be costly to communicate to consumers, costly for consumers to evaluate, and costly to implement in firm billing systems. Second, the cost of acquiring enough information about demand to distinguish optimal prices for one unit from the next could be substantial. Most firms that charge nonlinear prices specify n-part tariffs, which consist of a fixed fee and $(n-1)$ intervals of constant marginal price. (The three-part tariffs I focus on are a special case in which the first interval of

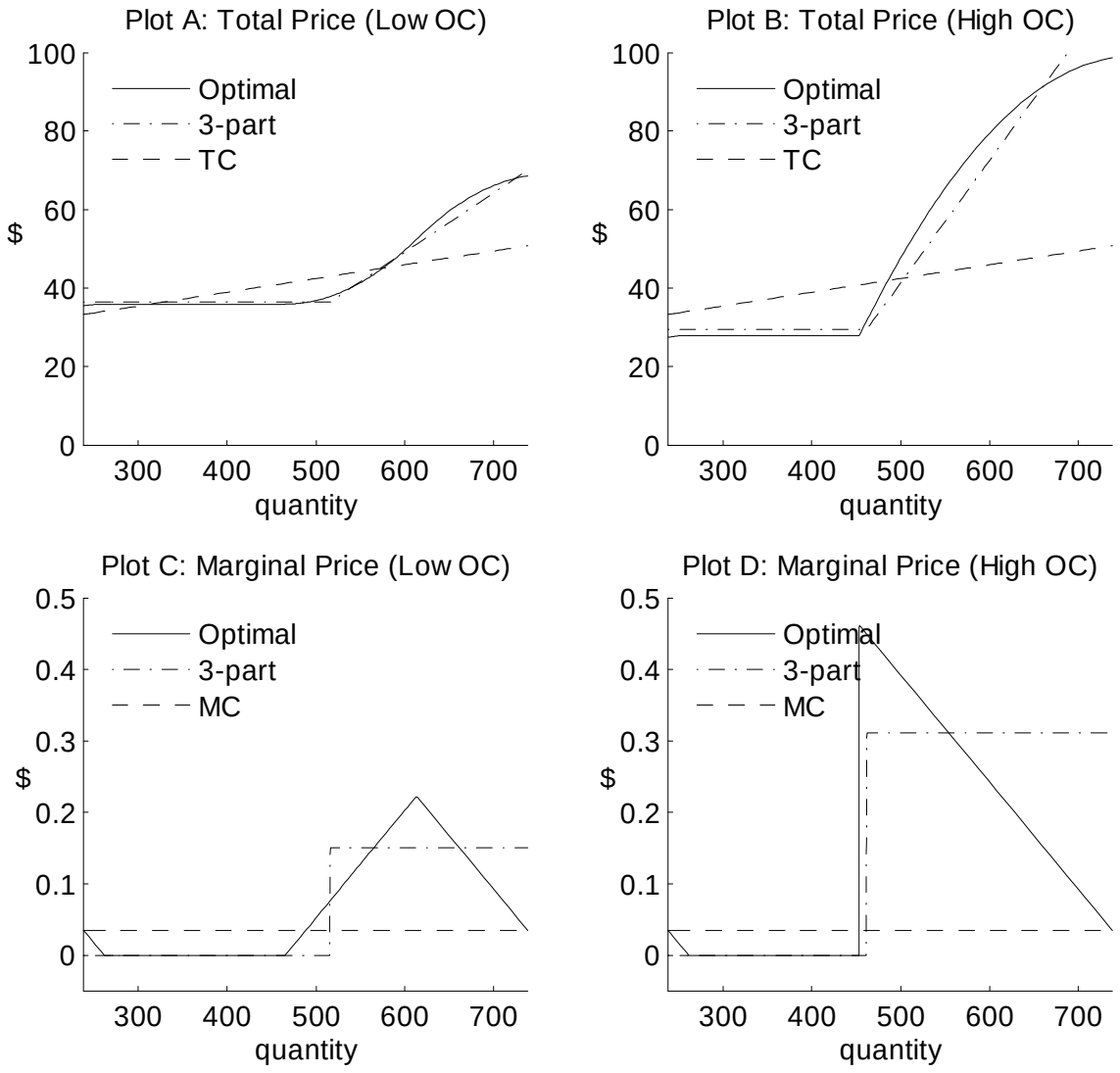

Figure 4: The fully nonlinear equilibrium price and its closest 3-part tariff approximation are depicted under perfect competition and positive marginal cost: $c = \$0.035$. Prices are shown for low overconfidence ($\Delta = 0.25$) in the left hand column and for high overconfidence ($\Delta = 0.75$) in the right hand column. Total prices and costs are shown in the top row, and marginal prices and costs are shown in the bottom row.

Whether or not a three-part tariff is an optimal approximation depends on complexity costs, consumers preferences, and beliefs. However, I now show that a three-part tariff is an excellent approximation whenever overconfident consumers are primarily uncertain about the quantity of desirable units, and tend to have a consistent and certain value for units that are desirable. Moreover, a generalization of Example 1 in Section 4.6 suggests that three-part tariffs may provide large improvements over two-part tariffs relative to the remaining approximation loss for a broader range of preferences.

Consider a generalization of the illustrative example in Section 3. If overconfident consumers value the first θ units at p each and are satiated thereafter ($V(q,\theta) = p \cdot \min\{q,\theta\}$), then the optimal contract is a three-part tariff that includes θ^* units at zero marginal price followed by a constant overage rate p. A constant overage rate is optimal regardless of firm and consumers' specific beliefs because although consumers are uncertain ex ante about how much they will want to consume, they have a uniform and certain valuation p for any units they do eventually find desirable. Proposition 3 shows that this is not a knife-edge result: If overconfident consumers have preferences close to $V(q,\theta) = p \cdot \min\{q,\theta\}$ (so that they are primarily uncertain about their purchase volume, rather than their value for future purchases) then three-part tariffs offer large improvements over two-part tariffs, and closely approximate optimal pricing.

Proposition 3 *Assume that marginal cost is zero, and that consumers' preference $V(q,\theta)$ is close to $p\min\{q,\theta\}$ in the sense that $V_q(q,\theta) \in \left[p \cdot 1_{q\leq\theta}, (p+\delta_p) \cdot 1_{q\leq\theta+\delta_q}\right]$ for some $\delta_q, \delta_p > 0$. For any $\varepsilon > 0$, there exist δ_p, $\delta_q > 0$ such that: (1) The best three-part tariff yields profits within ε of optimal profits. (2) If beliefs are symmetric around θ^*, then the best three-part tariff (with included units at zero marginal price) yields profits at least $\frac{p}{2}\left(E^*\left[\theta|\,\theta \leq \theta^*\right] - E\left[\theta|\,\theta \leq \theta^*\right]\right) - \varepsilon$ higher than the best two-part tariff.*

Proof. The inequalities are satisfied if $2\delta_q p + \delta_p E[\theta] + \delta_p\delta_q \leq \varepsilon$. See Appendix A. ■

For alternative preferences, a three-part tariff is optimal only for specific consumer and firm beliefs. For example, if $V(q,\theta) = v(q) + q\theta$, marginal costs are zero, consumers are sufficiently overconfident, and the true distribution of θ is exponential, then a three-part tariff is optimal. In otherwise the same context, however, the optimal overage rate would vary if the true distribution of θ had a varying hazard rate. (If $V_{q\theta}$ is constant and consumers believe $\theta = \theta^*$ with probability one, Proposition 2 shows that the optimal overage rate is proportional to the inverse hazard rate

---

constant marginal price sets marginal price equal to zero.) Wilson (1993) shows that "The profit and total surplus foregone by using an optimal n-part tariff rather than an optimal tariff with continuously varying prices is of order $1/n^2$ for large values of n; thus, cost considerations need not be large for a tariff with only a few parts to be optimal." Given his assumption that monotonicity is not binding, Wilson's (1993) asymptotic result extends to this setting.

$\frac{1-F(\theta)}{f(\theta)}$.) Nevertheless, three-part tariffs can still be a good approximation in this context even for firm beliefs with a widely varying hazard rate. If demand is linear, beliefs are uniform, consumers estimate the mean of θ correctly, and marginal cost is zero, (as in Example 1 Figure 3), it can be shown that three-part tariffs capture 82% or more of the virtual surplus (true surplus + fictional surplus) lost by using a two-part tariff instead of the optimal fully nonlinear tariff.[17] This suggests that three-part tariffs are likely to be a good approximation for a variety of preferences beyond those close to $V(q,\theta) = p \cdot \min\{q,\theta\}$.

4.8 Welfare

To evaluate welfare I assume that the firm's prior $F(\theta)$ is correct.[18] 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, as firm profits are always zero.

Under monopoly, total welfare is also lower when consumers are overconfident, but in general it is ambiguous as to whether consumers or the firm are better or worse off.

Lemma 3 *Under monopoly, if overconfident consumers overestimate the surplus created by the first-best allocation*[19] *($E^*\left[S^{FB}\right] \geq E\left[S^{FB}\right]$) then the firm is better off and consumers are worse off due to their overconfidence. If overconfident consumers do not overestimate the mean of θ ($E^*[\theta] \leq E[\theta]$) and $V_{\theta\theta} \geq 0$, then the firm is worse off and consumers are better off due to their overconfidence.*

Proof. See Appendix A. ■

If overconfident consumers overestimate expected first-best surplus, then they are willing to pay a larger fixed fee for marginal cost pricing than a monopolist could extract from consumers with

[17] Example 1 generalizes to $V_q(q,\theta) = \alpha - \beta q + \gamma\theta$ and $\theta \sim U\left[\underline{\theta}, \overline{\theta}\right]$, with consumers who underestimate the standard deviation of θ by $100\Delta\%$. If $q(\underline{\theta}) > 0$, then the ratio $1 - \frac{\text{3-part loss}}{\text{2-part loss}}$ (reported in the final row of Table 1 for Example 1) is independent of α, β, γ, or the support of θ. For zero marginal cost, the ratio can be calculated analytically as a function solely of the overconfidence level $\Delta > 0$. Its minimum is approximately 82%.

[18] See Footnote 10.

[19] *Note that given zero marginal cost, 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.*

correct-priors. The tables are turned if consumption value V is convex in θ and overconfident consumers do not overestimate average θ (and hence underestimate their expected utility),[20] because this creates bargaining power. The firm cannot extract all surplus ex ante, and to extract it ex post the firm must give away information rents as the customer is privately informed about θ in period two. This is the case in the examples discussed in Section 4.6 as consumers are assumed to estimate the mean of θ correctly. These consumers are strictly better off when overconfident while the monopolist is worse off and would prefer customers to have correct priors.

5 Multi-Tariff Menus

The model presented in Section 4 assumes that consumers have homogeneous priors ex ante, and therefore firms offer only a single tariff. In this section, I examine the case where consumer beliefs are initially heterogeneous and firms offer a menu of contracts.

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 $\theta \sim F_s^*(\theta)$. 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). Preferences are identical to those in the single-tariff model, except that in the monopoly case I make the stronger restriction $V_{qq\theta} = 0$ in place of equation (1).

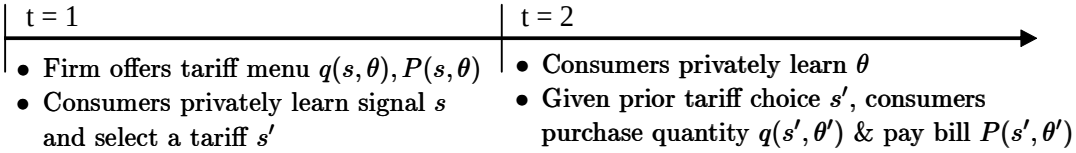

Figure 5: Time line for multi-tariff model.

There are two natural dimensions of heterogeneity to explore. First, consumers could all be similarly overconfident, but vary in their average demand. Second, consumers could all have the same true distribution of demand, but vary in their degree of overconfidence. (The later scenario could arise because experienced consumers overcome overconfidence via learning but new customers

[20] Given $V_{\theta\theta} \geq 0$, consumers' second period utility $U(\theta)$ is convex in θ for any incentive compatible allocation $q(\theta)$. Assumption A^* and $E^*[\theta] \leq E[\theta]$ imply that $F(\theta)$ RSOSD $F^*(\theta)$. (RSOSD is defined in Section 5.1.) Convexity and RSOSD imply $E^*[U(\theta)] \leq E[U(\theta)]$.

continually enter the market.) Extending the model in this way requires separate analysis for monopoly and perfect competition market conditions.

5.1 Monopolist's Multi-Tariff Menu

The monopolist's multi-tariff 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. For this analysis, I replace equation (1) by the stricter assumption $V_{qq\theta} = 0$. I consider signal spaces which order consumer beliefs $F_s^*(\theta)$ either by FOSD, or a more general reverse second order stochastic dominance (RSOSD), defined below.[21] This is sufficiently general for me to capture heterogeneity either in average demand or in degree of overconfidence.

Definition 1 *S is ordered by RSOSD*[22] *if $\int_\theta^{\bar{\theta}} F_H^*(x)\,dx \leq \int_\theta^{\bar{\theta}} F_L^*(x)\,dx$, $\forall\theta$ and $\forall H \geq L$.*

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 $\{q_H(\theta), P_H(\theta)\}$. Beyond the familiar non-negativity, satiation, and monotonicity constraints, substituting incentive and participation constraints in place of 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 using a sufficient (but not necessary) monotonicity condition. Within the current model, if $F_H^*(\theta)$ FOSD $F_L^*(\theta)$, then the analogous monotonicity condition that is sufficient (but not necessary) for upward incentive compatibility is: $q_L(\theta) \leq q_H(\theta)$ for all θ. As this sufficient condition can easily fail for overconfident consumers,[23] a more productive approach in this context is to directly incorporate the upward incentive constraint using optimal control techniques. Doing so leads to the following characterization of marginal prices.

---

[21] 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.

[22] *This is reverse SOSD rather than SOSD because I integrate from the top down rather than the bottom up. Thus higher s corresponds to higher spread and higher mean rather than lower spread and higher mean. Increasing s could thus be thought of as the combination of a mean preserving spread and a first order stochastic dominant shift. Given F_L crosses F_H once from below, F_H RSOSD F_L if and only if $\mu_H \geq \mu_L$.*

[23] Given $F_H^*(\theta)$ FOSD $F_L^*(\theta)$, the binding downward incentive constraint is sufficient for upward incentive compatibility if $q_L(\theta) \leq q_H(\theta)$. Overconfidence can generate greater downward distortions on the H contract than on the L contract. This is because (excluding screening effects) extreme overconfidence drives overage rates upwards towards ex post monopoly prices (Web Appendix B). These are higher (which implies greater downward quantity distortion) for higher average demand. Ex ante screening will ameliorate this effect by distorting quantity downwards on the L contract, but not necessarily completely. The upward incentive constraint is likely to bind when the monotonicity condition is severely violated by the relaxed ($\gamma = 0$) solution, and must bind when it is violated at all θ.

Proposition 4 *Given (i) $F_H^*(\theta)$ FOSD $F_L^*(\theta)$, or (ii) $V_{\theta\theta} \geq 0$ and $F_H^*(\theta)$ RSOSD $F_L^*(\theta)$: The monopoly payment functions $\hat{P}_L(q)$ and $\hat{P}_H(q)$ 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. Let $\hat{\theta}(q,s)$ be the inverse of the monopoly allocation $\hat{q}_s(\theta)$ when the latter is invertible. For quantities at which there is no pooling, marginal prices are given by equations (9-10) where $\alpha = \Pr(H)$ and $\gamma \geq 0$ is the shadow price of the upward incentive constraint.*

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

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

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

5.1.1 Heterogeneity in average demand

To capture heterogeneity in average demand among similarly overconfident consumers, assume that consumers' conditional priors $F_s^*(\theta)$ cross firms' conditional priors $F_s(\theta)$ once from below at $\theta^*(s)$, and that signals are ordered by FOSD so that $F_H^*(\theta) \leq F_L^*(\theta)$. Proposition 4 applies and equations (9-10) show that when the upward incentive constraint is not binding ($\gamma = 0$), marginal prices for the high-demand user are identical to those in the single-tariff case. On the other hand, marginal prices for the low-demand user 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 while 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 effect 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.

Example 2 *I extend Example 1 to the case of two ex ante signals. Half of consumers, who receive signal L ex ante, have the same true and perceived distributions of demand as in the low-overconfidence case ($\Delta = 0.25$) of Example 1: F_L is $U[-.5, .5]$ and F_L^* is $U[-.375, .375]$. The other half of consumers, who receive signal H ex ante, have higher demand in a FOSD sense: F_H is $U[-.5, .7]$ and F_H^* is $U[-.35, .55]$. Both low and high average demand consumers estimate the mean of θ correctly but underestimate the standard deviation of θ by 25%. Marginal cost is \$0.035.*

Figure 6 depicts a monopolist's optimal two contract menu for Example 2. Plot A shows total prices for contracts L and H, while Plot B shows marginal cost and marginal prices for contracts L and H. In this example, upward incentive constraints do not bind ($\gamma = 0$), so screening does not distort contract H away from the single-contract benchmark.[24] Ex ante screening does distort contract L. Plot B shows as a dotted line marginal prices that would be charged on contract L if there were no distortions due to screening, which coincides with the single-contract benchmark shown in Plot C of Figure 4. By comparison, the solid line in Plot B shows that optimal marginal prices for contract L are distorted upwards. Due to the distortion, contract L provides roughly 85 fewer 'free' units, and a maximum marginal price of \$0.56 rather than only \$0.22.

Despite the screening distortion, contracts L and H display key qualitative features of observed three-part tariff menus: increasing fixed fees for increasing quantities of 'free' units followed by steep marginal charges. Moreover, unlike the single-contract benchmark, marginal price is strictly positive for the last unit sold on contract L. (Note that no consumption occurs on the dotted portion of contract L in equilibrium, and the model only places a lower bound on prices in this region. The dotted extension represents the minimum constant marginal price extension of the contract. Marginal price doesn't fall to marginal cost because the support of the type distributions for low and high average demand groups do not coincide.)

[24] Upward incentive compatibility is not binding despite the fact that optimal marginal prices of the two contracts briefly cross, and hence allocations are not monotonically increasing from contract L to H.

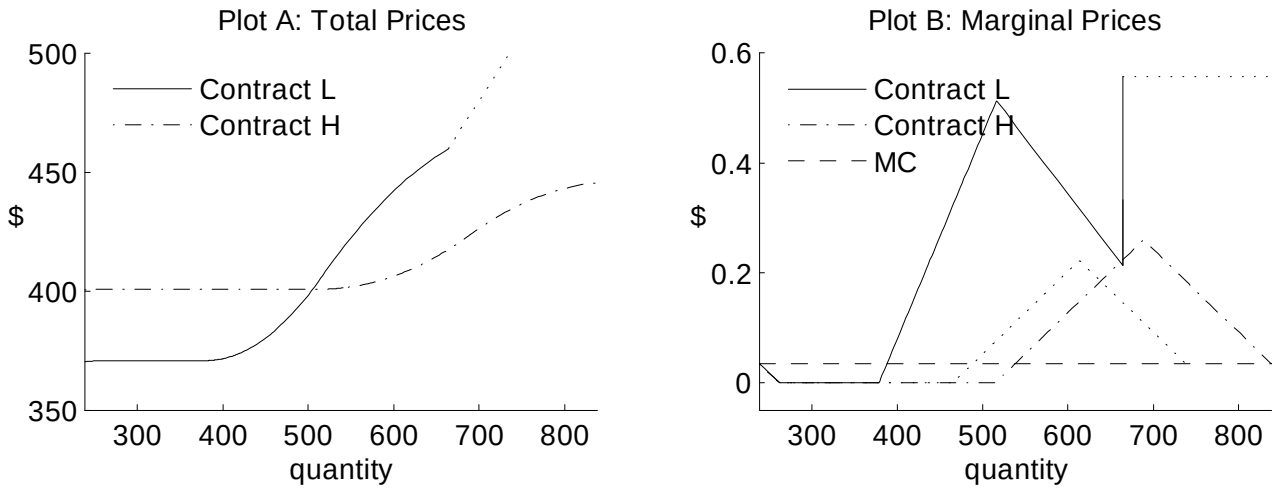

Figure 6: A monopolist's two contract menu (Example 2)

5.1.2 Heterogeneity in degree of overconfidence

To capture heterogeneity in degree of overconfidence (rather than average demand), assume that the firm's conditional priors are independent of s ($F_s(\theta) = F(\theta)\ \forall s$), high-types have correct beliefs ($F_H^*(\theta) = F(\theta)$), and low-types are overconfident ($F_L^*(\theta) = F^*(\theta)$ crosses $F(\theta)$ once from below). If overconfident consumers do not overestimate the mean of θ, so that $F_H^*(\theta)$ RSOSD $F_L^*(\theta)$, and in addition $V_{\theta\theta} \geq 0$, then Proposition 4 applies.[25] Equations (9-10) reduce to equations (11-12) and show that screening between overconfident and correct-prior consumers exacerbates the distortion in marginal prices for overconfident consumers. When the upward incentive constraint is not binding ($\gamma = 0$), marginal price equals marginal cost for the correct-prior consumer. On the other hand, if the upward incentive-constraint binds ($\gamma > 0$), then correct-prior consumers will pay above marginal cost for initial quantities and below marginal cost for higher quantities (as if they

[25] If overconfident consumers overestimate the mean of θ, then Proposition 4 does not apply. In this case solving the monopolist's problem is more difficult since it is not clear which first-period incentive and participation constraints will bind. However, there is one simple benchmark case where neither upward nor downward incentive constraints bind. If overconfident types correctly estimate the surplus generated by the first best allocation ($E^*\left[S^{FB}\right] = E\left[S^{FB}\right]$), then 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 $\left\{\hat{q}(\theta), P^M(\theta)\right\}$ respectively. (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.) For a proof see Appendix A.

were underconfident).

$$\frac{d\hat{P}_L(q)}{dq} = \max\left\{0, C_q(q) + \frac{1+\gamma}{1-\alpha}V_{q\theta}(q,\theta)\frac{F^*(\theta)-F(\theta)}{f(\theta)}\right\} \quad (11)$$

$$\frac{d\hat{P}_H(q)}{dq} = \max\left\{0, C_q(q) - \frac{\gamma}{\alpha}V_{q\theta}(q,\theta)\frac{F^*(\theta)-F(\theta)}{f(\theta)}\right\} \quad (12)$$

Note that Lemma 3 extends to this setting, and overconfident consumers are better off than they would be if all consumers had correct beliefs. At the same time, the fraction of consumers with correct priors must be weakly better off than their overconfident counterparts because correct beliefs lead to better decisions. Thus the presence of overconfident types in the marketplace improves the outcome for both types.

5.2 Perfectly Competitive Multi-Tariff Menu

If firms believe that all consumers share the same true distribution of θ, while consumers have heterogeneous beliefs, for instance because they differ in severity of overconfidence, then screening between consumers with different beliefs does not distort prices. Consumers who receive signal s will be offered and choose the tariff described by the single-tariff model as if they were the only type in the market.[26]

A general characterization of the equilibrium tariff menu under perfect competition when consumers differ both in their beliefs and in their true distributions of demand is a difficult problem left for future research. Unlike the monopoly case, 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 some cases, ex ante screening is not distortionary and each contract on the equilibrium menu matches that in the single-tariff model. One such example is illustrated by Figure 7. This is a variation of the example shown in column 2 of Figure 4, in which marginal cost is \$0.035 and consumers are highly overconfident ($\Delta = 0.75$). The variation is that consumers receive one of three signals ex ante, low, medium, or high, which correspond to future θ being distributed

[26] Note that a firm expects the same profit on a given tariff from all participating consumers who the firm believes have the same type distribution $F(\theta)$. Thus the set of zero-profit tariffs is the same regardless of whether a firm is serving consumers with one prior or another. The competitive single-tariff solution is the tariff in this set that maximizes the perceived expected utility of its target customer. Hence, consumers cannot believe themselves to be strictly better off under another offered tariff.

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 7.

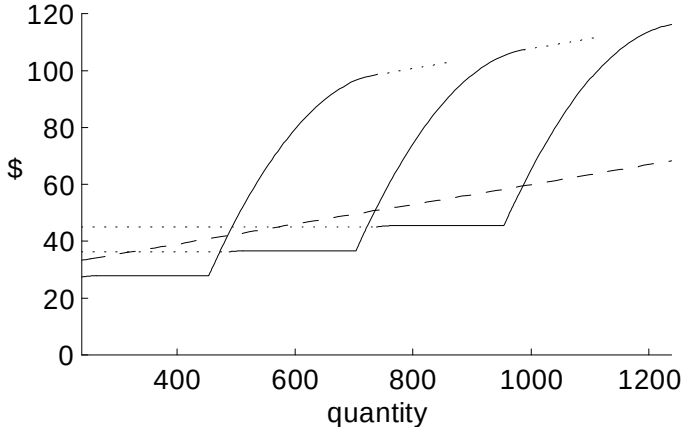

Figure 7: Total pricing for a 3-tariff menu under perfect competition. Solid portions of the tariffs are uniquely optimal. Dashed portions of the tariffs are illustrative extensions where no consumption takes place. The straight line shows total costs.

6 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.

6.1 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. (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.)

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 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 7).

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. ■

While flat-rate bias alone does not appear to explain three-part tariffs, the taxi-meter effect could be complementary to overconfidence. Overconfident consumers underestimate the likelihood of making overages, and in the extreme could believe that they will always stay within the flat portion of their chosen tariff. Given such beliefs, the taxi-meter effect would reinforce the attractiveness of three-part tariffs.

6.2 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. Bar-Gill (2006) draws similar conclusions when consumers underestimate their future usage.

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.

6.3 Price Discrimination with Common Priors

Under perfect competition, common priors yield marginal cost pricing,[27] 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 5 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, then marginal price should always be above marginal cost and consumption distorted downwards for all but those with the highest average demand. Given such a type distribution, price discrimination with common priors does 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

[27] This is true not only in the current model, but in more general models that maintain key features of the framework including the contracting time-line, quasi-linear utility, and deterministic costs.

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, with a continuum of ex ante types, consumers' conditional priors over θ should satisfy equation (13) for some cutoff $\theta^*(s)$ increasing in s. Alternatively, with two ex ante types, $F_L(\theta)$ must cross $F_H(\theta)$ once from below (See Proposition 4).

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

To understand why this type distribution generates such pricing, consider an example with two ex ante types. The high type ($s = H$) 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 type ($s = L$) 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 small or zero marginal charges below marginal cost at low quantities followed by high marginal charges above marginal cost 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 small or zero marginal charges are attractive to the graduate student, and allow a higher monthly fee to be charged on the graduate 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 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.[28] More specifically, it requires that the cumulative usage distribution of plan 1 customers be below that of plan 2 customers, for all $q < Q_1$: $H\left(q|s_1\right) \leq H\left(q|s_2\right)$. As shown in the following section, this is not consistent with observed consumer behavior. As a result, the common-prior model does not appear to explain observed cellular

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

phone service tariff menus.

7 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.[29] 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 sample there more than 50 distinct tariffs from more than 10 menus. These include national calling plans, local calling plans, and a two-part tariff, which all vary over time.

For my primary analysis, I focus on two similar menus with the most usage data. These are the set of local plans offered to students in the fall of 2002 and the fall of 2003 (Figure 8).[30] 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 allowances, 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.[31] Plan 0 has a small monthly fixed-fee and a constant per-minute charge below the overage rates of plans 1-3.

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.[32] Figure 9 plots the cumulative usage distributions $\hat{H}(q|plan)$ and their 95% confidence intervals[33] for customers on plans 1, 2, and 3. Prorated bills for incomplete months of

[29] Students received an itemized phone bill, mailed by default to their campus residence, which was separate from their university tuition bill. 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.

[30] 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.

[31] 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.

[32] 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.

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

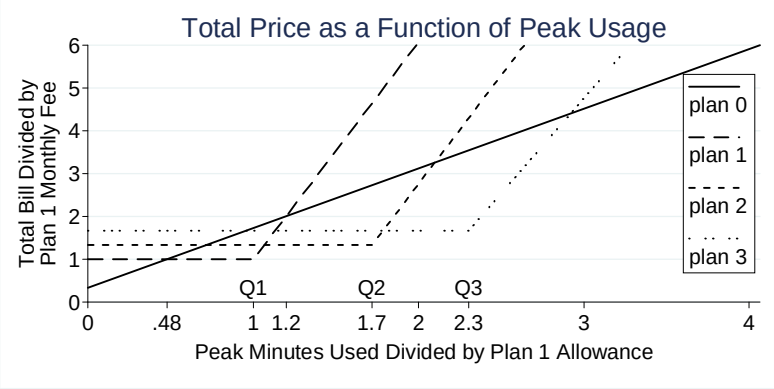

Figure 8: Total price as a function of peak usage for plans 0, 1, 2, and 3 (fall 2002 & 2003 menus). Usage and price are measured as a fraction of the plan 1 included allowance (Q_1) and monthly fee respectively.

service 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 9 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 a strict FOSD ordering. Formal pair-wise tests of first order stochastic dominance between the three distributions provide limited additional insight.[34] It is clear from the figure, however, that usage patterns are inconsistent with the assumption driving the common-prior alternative.

It is not true that $\hat{H}(q|plan1) \leq \hat{H}(q|plan2)$ for $q < Q_1$. Plan 2 customers are not "undergraduate" types who actually consume less than Q_1 minutes more frequently than plan 1 customers. Rather, plan 2 customers use less than Q_1 minutes only 60% of the time, while plan 1 customers use less than Q_1 minutes 79% of the time. Comparisons with usage by plan 3 customers are similar. Therefore, in contrast to the overconfidence explanation, the alternative price discrimination model cannot simultaneously explain both observed pricing and observed usage patterns.

---

mean of the true CDF $H(q)$ and variance $\frac{(1-H(q))H(q)}{N}$.

[34] 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).)

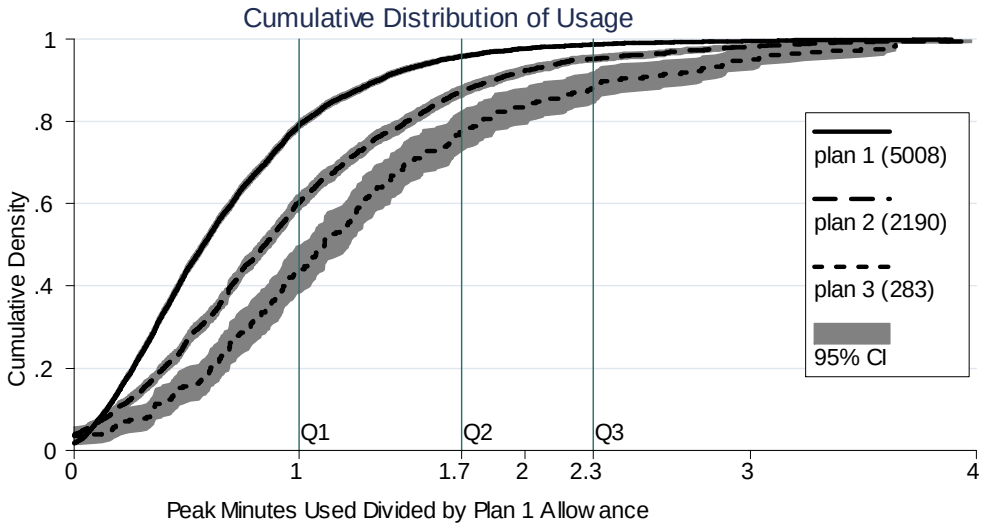

Figure 9: Cumulative usage distributions $\hat{H}(q|plan)$ and their 95% confidence intervals for customers on Plans 1, 2, and 3. Usage is measured as a fraction of Q_1, the plan 1 allowance. Vertical lines are drawn at all three plan allowances. (Fall 2002 and 2003 menus.)

One might be concerned that the model of overconfidence is off the mark if one believes that customers only rarely exceed their included allowances. It is reasonable to hypothesize that 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.

	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 2: Mean usage as a fraction of the plan allowance across plans 1-3 (fall 2002 & 2003 menus).

Figure 9 and Table 2 clearly show that overages are an important feature of customer behavior. 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 18,116 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 (230% 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 8, and in particular all offered free night and weekend calling. In fall 2003, plan 0 no longer offered free off-peak 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 48% and 120% of Q_1 for plan 1 and between 42% and 122% of Q_2 for plan 2 (Figure 8). Consumers' choice of plans 1 or 2 implies an initial belief that 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 3 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 42% of the plan 1 monthly fixed-fee.[35] In fact, had all plan 1 customers chosen plan 0 instead, they would have saved an average of 19% of the plan 1 monthly fixed-fee. Similarly, 50% of plan 2 customers would have saved an average of 36% of the plan 1 monthly fixed-fee by choosing plan 0, although had all plan 2 customers chosen plan 0 average savings would not be significantly different from zero. In contrast, only 5% of plan 0 customers would have saved money by choosing plans 1, 2, or 3. In all cases, customers who quit or switch plans in less than 6 months make more and larger mistakes than those who stay with their chosen plan longer. This is consistent with learning, but nevertheless mistakes are still prevalent among experienced customers.[36]

---

[35] The frequency and size of mistakes are underestimated. First, plan 0 includes 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.

[36] 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 99% and 33% of the plan 1 fixed monthly fee.

To avoid overstating the size of mistakes, Tables 3 and 4, and the text report bill-weighted average mistake sizes.

	Plan 0 Customers	Plan 1 Customers	Plan 2 Customers
Customers	393 (62%)	92 (15%)	124 (20%)
Bills	5,551	899	1,193
Alternative Considered	Plan 1, 2, or 3	Plan 0	Plan 0
Alternative Lower Cost Ex Post	5%	65%	50%
Conditional Avg. Saving†	11% *	42% **	36% **
Unconditional Avg. Saving†	NA	19% **	2%

†Average per month, as a percentage of Plan 1 monthly fixed-fee.
* 90% confidence. ** 99% confidence.

Table 3: Frequency and size of ex post "mistakes" (fall 2002 menu).

For the plan 1 and 2 customers who could have saved money on plan 0, Table 4 shows the rate of underusage and overusage, as well as their contributions to the potential savings on plan 0. Underusage (overusage) occurs in any specific month when plan 0 would have been cheaper due to low (high) usage. On a monthly basis, the most common ex post mistake is underusage, which does not justify payment of a high fixed fee. However, the less frequent ex post error of overusage is typically a much more expensive mistake due to high overage charges. Hence when dollar weighted, both underusage and overusage are important sources of potential savings. For the 65% of plan 1 customers who could have saved money on plan 0, potential savings from underusage occur on 56% of bills and average 20% of the plan 1 fixed fee across all bills. Potential savings from overusage occur on only 16% of bills, but average 30% of the plan 1 fixed fee across all bills. These potential savings are partially offset by 28% of bills in which there is neither underusage nor overusage. Figures for plan 2 customers are similar, although dollar weighted underusage contributes slightly more to mistakes than overusage, rather than the reverse.

	Plan 1 Customers (60)		Plan 2 Customers (62)	
	Bills (527)	Potential Saving†	Bills (612)	Potential Saving†
Underusage	56%	20%	58%	29%
Intermediate	28%	(8%)	34%	(17%)
Overusage	16%	30%	9%	24%
Total	100%	42%	100%	36%

†Average per month as a percentage of Plan 1 monthly fixed-fee.

Table 4: Underusage versus overusage for customers who could have saved on plan 0.

Since those customers who make the largest mistakes tend to switch or quit earlier than others, and therefore have fewer bills, customer-weighted average mistake sizes are larger. For instance, customer-weighted average potential savings conditional on a mistake are 25%, 68%, and 53% of the plan 1 fixed-fee for plan 0, 1, and 2 customers respectively. In comparison the corresponding bill-weighted averages reported in Table 3 are 11%, 42%, and 36%. Moreover, overusage is relatively more important for customer-weighted average mistakes, due to the presence of customers who switch plans or quit after a handful of extremely large overages.

The prevalence and size of ex post mistakes show that customers are uncertain about their future demand when making tariff choices, and that modeling this uncertainty is critical for understanding the market. Moreover, the consistent direction of the mistakes, and in particular the fact that on average plan 1 customers could have saved money had they all chosen plan 0, provide evidence that consumers have biased beliefs ex ante. Were mistakes due primarily to underusage, the consumers' bias might be systematic overestimation of demand and could cause flat rate bias (Lambrecht and Skiera 2006). Were mistakes due primarily to overusage, the consumers' bias might be systematic underestimation of demand, consistent with naive quasi-hyperbolic discounting (DellaVigna and Malmendier 2004). Instead, the important contribution of both underusage and overusage to mistakes is consistent with consumer overconfidence.

8 Conclusion

This paper has shown that given overconfident consumers, low marginal costs, and free disposal, optimal pricing involves included units at zero marginal price followed by high marginal charges. This provides a promising explanation for the three-part tariff menus observed in the cellular phone services market. The theory ignores tariff complexity costs, and hence does not necessarily predict overage rates to be constant as observed. When consumers are primarily uncertain about the volume of desirable units, relative to a fairly consistent value for units that are desirable, three-part tariffs will be a good approximation for optimal pricing in the sense that they provide a large improvement over two-part tariffs relative to the remaining approximation losses. An example with linear demand and uniform beliefs suggests that this is true for other reasonable cases as well. Empirical evidence shows that consumer usage patterns are consistent with the overconfidence 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.

I have not modeled the possibility that over time consumer beliefs become calibrated correctly. Even if this happens in the long run, the psychology literature suggests that learning is slow (Alpert and Raiffa 1982, Plous 1995, Bolger and Önkal Atay 2004). As a result, my characterization of a one-month contract may be a good approximation for an optimal multi-month contract. Moreover, the multi-tariff model of Section 5 can capture the heterogeneity in overconfidence that arises when experience eliminates overconfidence for some fraction of the population but new consumers who are overconfident continue to enter the market. In this case firms would be expected to offer three-part tariffs to overconfident consumers, but less convex tariffs to sophisticated consumers. It may

be that Cingular's "Rollover" contracts, which carry forward unused minutes from the included allowance from one month to the next, are specifically targeting such sophisticated consumers who realize that their demand varies month to month.

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, car leasing, a range of rental services, consumer credit card debt, and an increasing number of other services.

9 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 θ' who is offered $q(\theta') > q^S(\theta')$, a monopolist would be weakly better off offering $q(\theta') = q^S(\theta')$ instead. A type θ' consumer'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 θ', so type θ consumer's choice remains incentive compatible. Production costs would be weakly reduced however. This argument can be repeated for any θ', 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 $q(\theta')$ to $q^S(\theta')$ for a positive measure of types θ' 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 Restatement of Proposition 1

Proposition 1 may be restated as follows: (1) If quantity payment pair $\left\{\hat{q}(\theta), P^C(\theta)\right\}$ solves the perfect competition problem defined in Section 4.2, then the pair $\left\{\hat{q}(\theta), P^C(\theta) + E[\Psi(\hat{q}(\theta), \theta)]\right\}$ solves the monopoly problem defined in Section 4.2. Conversely, if the pair $\left\{\hat{q}(\theta), P^M(\theta)\right\}$ solves the monopoly problem, then $\left\{\hat{q}(\theta), P^M(\theta) - E[\Psi(\hat{q}(\theta), \theta)]\right\}$ 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 4.2, then the pair $\{\hat{q}(\theta), P(\theta)\}$, where $\hat{q}(\theta) = \min\left\{q(\theta), q^S(\theta)\right\}$, solves the same problem. Moreover, $\hat{q}(\theta)$ maximizes expected virtual surplus $E[\Psi(q(\theta), \theta)]$ subject to monotonicity, non-negativity, and satiation constraints, where virtual surplus is defined by equation (4), and $P(\theta)$ is given by equation (5) for monopoly or equation (6) for perfect competition. Conversely, if $\hat{q}(\theta)$ maximizes expected virtual surplus subject to monotonicity, non-negativity, and satiation constraints, then there exists a unique pair of payments $\left\{P^M(\theta), P^C(\theta)\right\}$ such that $\left\{\hat{q}(\theta), P^M(\theta)\right\}$ solves the monopoly problem and $\left\{\hat{q}(\theta), P^C(\theta)\right\}$ solves the perfect competition problem. These payments are given by equations (5) and (6) respectively.

A.3 Proof of Proposition 1

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 θ' be $q^c(\theta,\theta') \equiv \min\{q(\theta'), q^S(\theta)\}$ 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 $\frac{\partial}{\partial\theta}U(\theta,\theta') = V_\theta\left(q^c(\theta,\theta'),\theta\right)$, 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):

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

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

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. As 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.

■

A.4 Proof of Lemma 2

Proof. Part 1: Under maintained assumptions, the constraint set $D(\theta) = \left[0, q^S(\theta)\right]$ is convex and compact valued, continuous, and non-empty. Further, as shown below, virtual surplus $\Psi(q, \theta)$ is continuous and strictly quasi-concave in q. (Note that this is where I use the restriction on $V_{qq\theta}$ described by equation (1).) Therefore $q^R(\theta)$ is a continuous function. Moreover, as $\Psi(q, \theta)$ is twice continuously differentiable 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.

Virtual surplus $\Psi(q, \theta)$ is strictly quasi-concave in q:

Case (a) $\frac{F(\theta)-F^*(\theta)}{f(\theta)} > 0$: If $q \leq q^{FB}(\theta)$ then $V_q - C_q \geq 0$ and so Ψ is strictly increasing in q:

$$\Psi_q(q, \theta) = \underbrace{V_q(q, \theta) - C_q(q)}_{\geq 0 \ (q \leq q^{FB}(\theta))} + \underbrace{V_{q\theta}(q, \theta)}_{>0} \underbrace{\frac{F(\theta) - F^*(\theta)}{f(\theta)}}_{>0 \text{ (case a)}} > 0$$

If $q \in (q^{FB}(\theta), q^S(\theta))$, then by equation (1) $V_{qq\theta} \leq 0$, and Ψ is strictly concave.

$$\Psi_{qq}(q, \theta) = \underbrace{V_{qq}(q, \theta) - C_{qq}(q)}_{<0} + \underbrace{V_{qq\theta}(q, \theta)}_{\leq 0 \ (q > q^{FB})} \underbrace{\frac{F(\theta) - F^*(\theta)}{f(\theta)}}_{>0 \text{ (case a)}} < 0$$

Since Ψ is continuous, this is sufficient for strict quasi-concavity on $D(\theta) = \left[0, q^S(\theta)\right]$.

Case (b) $\frac{F(\theta)-F^*(\theta)}{f(\theta)} < 0$: Similarly, if $q \geq q^{FB}$ then $V_q - C_q \leq 0$ and so Ψ is strictly decreasing in q. If $0 < q < q^{FB}(\theta)$, then by equation (1) $V_{qq\theta} \geq 0$, and Ψ is strictly concave. Since Ψ is continuous, this is sufficient for strict-quasi concavity on $D(\theta) = \left[0, q^S(\theta)\right]$.

Case (c) $\frac{F(\theta)-F^*(\theta)}{f(\theta)} = 0$: Virtual surplus equals true surplus, which is strictly concave and hence also strictly quasi-concave.

Part 2: It follows from the application of standard results in optimal control theory (Seierstad and Sydsæter 1987) and the Kuhn-Tucker theorem. Note that because the virtual surplus function is strictly quasi-concave, but not necessarily strictly concave, optimal control results give necessary rather than sufficient conditions for the optimal allocation. (See Web Appendix C for additional

discussion.) ■

A.5 Proof of Proposition 2

Proof. (1) Differentiating equation (6) for $P(\theta)$ and making a change of variables yields an expression for marginal price that is valid at non-pooling quantities where $\hat{q}(\theta)$ is invertible: $\frac{d}{dq}P(q) = V_q(q, \hat{\theta}(q))$.

(2) At any pooling quantity q, by assumption $V_{q\theta} > 0$, marginal price must increase discontinuously as above and below q marginal price is given by $V_q(q, \inf\{\theta : \hat{q}(\theta) = q\})$ and $V_q(q, \sup\{\theta : \hat{q}(\theta) = q\})$ respectively.

(3) At non-pooling quantities Lemma 2 implies that either (i) $\hat{q}(\theta)$ satisfies the first order condition $\Psi_q(\hat{q}(\theta), \theta) = 0$, or (ii) satiation binds and marginal price is zero since $V_q\left(q^S(\theta), \theta\right) = 0$. In the former case, the first order condition implies marginal price is:

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

When $\left[C_q(q) + V_{q\theta}(q, \hat{\theta}(q))\frac{F^*(\hat{\theta}(q)) - F(\hat{\theta}(q))}{f(\hat{\theta}(q))}\right]$ 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 $\left[C_q(q) + V_{q\theta}(q, \hat{\theta}(q))\frac{F^*(\hat{\theta}(q)) - F(\hat{\theta}(q))}{f(\hat{\theta}(q))}\right]$ whenever that quantity is positive, and zero otherwise.

(4) As $\hat{\theta}(q)$ is a continuous function at non-pooling quantities, marginal price is as well. Thus payment $\hat{P}(\hat{\theta}(q))$ is continuously differentiable at non-pooling quantities. Moreover, $\hat{P}(\hat{\theta}(q))$ is continuous because incentive compatibility requires that types who pool at the same quantity pay the same price. ■

A.6 Proof of Lemma 3

Proof. 1. Given the characterization of $\hat{q}$ in Proposition 1, expected virtual surplus must be weakly higher under the equilibrium allocation than under the first best allocation: $E[\hat{\Psi}] \geq E\left[\Psi^{FB}\right]$. 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 $E[\hat{\Psi}] \geq E^*\left[S^{FB}\right]$. The assumption $E^*\left[S^{FB}\right] \geq E\left[S^{FB}\right]$ therefore implies that the firm is better off: $E[\hat{\Psi}] \geq E\left[S^{FB}\right]$. This in turn implies that consumers are worse off as total welfare is lower.

2. Differentiating consumer utility shows that $U(\theta)$ is convex in θ for any incentive compatible

allocation if $V_{\theta\theta} \geq 0$. This follows from local incentive compatibility, increasing differences $V_{q\theta} > 0$, and monotonicity $q_\theta(\theta) \geq 0$:

$$U_{\theta\theta}(\theta) = \underbrace{V_{q\theta}(q(\theta),\theta)}_{(+)} \underbrace{q_\theta(\theta)}_{(+) \text{ by IC}} + \underbrace{V_{\theta\theta}(q(\theta),\theta)}_{\geq 0 \text{ (assumption)}} \geq 0$$

Convexity implies that $E[U(\theta)] \geq E^*[U(\theta)]$ if F RSOSD F^*. (For RSOSD see Definition 1. The result is analogous to the standard result that if X second order stochastically dominates (SOSD) Y then $E[h(X)] \geq E[h(Y)]$ for any concave utility h (Rothschild and Stiglitz 1970, Hadar and Russell 1969, Hanoch and Levy 1969). The proof is similar and hence omitted.) Assumption A^* and $E^*[\theta] \leq E[\theta]$ jointly imply that F RSOSD F^*. This fact is analogous to Hanoch and Levy's (1969) Theorem 3. Coupled with the participation constraint $E^*[U(\theta)] \geq 0$, $E[U(\theta)] \geq E^*[U(\theta)]$ implies that overconfident consumers are weakly better off for being overconfident, and hence the monopolist is weakly worse off. (Note that for the first best allocation, $E[U(\theta)] \geq E^*[U(\theta)]$ implies that $E\left[S^{FB}\right] \geq E^*\left[S^{FB}\right]$ as marginal cost pricing ensures that expected profits are independent of the perceived distribution of quantity sold. Hence parts 1 and 2 of the lemma are consistent.) ■

A.7 Proof of Proposition 3

Proof. Part (1): The first step is to derive an upper bound for optimal profits. Let $Q \equiv \hat{q}(\theta^*)$. Then Q units are included at zero marginal price on the optimal tariff (Corollary 1). The optimal tariff earns $\int_{\underline{\theta}}^{\bar{\theta}} V\left(\min\left\{Q, q^S(\theta)\right\}, \theta\right) f^*(\theta)\, d\theta$ via the fixed fee for the perceived expected value of Q included units. Since $\hat{q}(\theta)$ is non-decreasing, profits from units above Q can be written as:

$$\int_{\theta^*}^{\bar{\theta}} ((V(\hat{q}(\theta),\theta) - V(Q,\theta)) f^*(\theta) + (P(\hat{q}(\theta)) - P(Q))(f(\theta) - f^*(\theta)))d\theta$$

Now, integrating by parts, and applying local incentive compatibility ($\frac{d}{d\theta}U(\theta) = V_\theta(q(\theta),\theta)$) implies:

$$\begin{aligned} & \int_{\theta^*}^{\bar{\theta}} (V(\hat{q}(\theta),\theta) - V(Q,\theta) - P(\hat{q}(\theta)) + P(Q))(f(\theta) - f^*(\theta))\, d\theta \\ = & \int_{\theta^*}^{\bar{\theta}} (V_\theta(\hat{q}(\theta),\theta) - V_\theta(Q,\theta))(F^*(\theta) - F(\theta))\, d\theta \end{aligned}$$

Since $F(\theta) \geq F^*(\theta)$ for all $\theta \geq \theta^*$, and $V_{q\theta} > 0$ this must be positive. Rearranging terms gives

$$\int_{\theta^*}^{\bar{\theta}} ((V(\hat{q}(\theta),\theta) - V(Q,\theta)) f^*(\theta) + (P(\hat{q}(\theta)) - P(Q))(f(\theta) - f^*(\theta)))d\theta \leq \int_{\theta^*}^{\bar{\theta}} (V(\hat{q}(\theta),\theta) - V(Q,\theta)) f(\theta)\, d\theta$$

Given the satiation constraint and fixed cost FC, this implies that an upper bound on profits is

$$\Pi^{\text{optimal}} \leq \int_{\underline{\theta}}^{\bar{\theta}} V\left(\min\left\{Q, q^S(\theta)\right\}, \theta\right) f^*(\theta)\, d\theta + \int_{\theta^*}^{\bar{\theta}} \left(V\left(q^S(\theta), \theta\right) - V(Q, \theta)\right) f(\theta)\, d\theta - FC$$

Moreover, since preferences are close to $p \min\{q, \theta\}$:

$$\Pi^{\text{optimal}} \leq \int_{\underline{\theta}}^{\bar{\theta}} V\left(\min\left\{Q, q^S(\theta)\right\}, \theta\right) f^*(\theta)\, d\theta + (p + \delta_p) \int_{Q-\delta_q}^{\bar{\theta}} (\theta + \delta_q - Q) f(\theta)\, d\theta - FC$$

The second step is to compare the preceding upper bound on optimal tariff profits, to minimum profits on a (possibly sub-optimal) 3-part tariff which also includes Q units at zero marginal price, and charges an overage rate p. Like the optimal tariff, this 3-part tariff earns $\int_{\underline{\theta}}^{\bar{\theta}} V\left(\min\left\{Q, q^S(\theta)\right\}, \theta\right) f^*(\theta)\, d\theta$ for the first Q units. In addition, it earns at least $\int_{Q}^{\bar{\theta}} p(\theta - Q) f(\theta)\, d\theta$ from units above Q. Therefore, lost profits due to using a 3-part approximation are at most $\delta_q p + \delta_p E[\theta] + \delta_q \delta_p$.

Part (2): Let p^* be the optimal 2-part tariff marginal charge. I will derive a lower bound for the profit difference between the optimal 2-part tariff, and a (possibly suboptimal) 3-part tariff with $Q = \theta^*$ included units and overage rate equal to $\max\{p, p^*\}$.

Starting from the optimal 2-part tariff, I analyze the shift to the 3-part tariff in 3 stages. First, if $p^* > p$, reduce marginal price everywhere to $\min\{p, p^*\}$. At worst, this reduces marginal revenues and profits by $\delta_p E[\theta] + \delta_p \delta_q$. Second, increase marginal price from $\min\{p, p^*\}$ to p for units above $Q = \theta^*$. Since all customers still buy at least θ units, types above θ^* generate additional marginal fees, partially offset by reduced fixed fees, on units $q \in [\theta^*, \theta]$ that increase profits by:

$$(p - \min\{p, p^*\}) \int_{\theta^*}^{\bar{\theta}} (\theta - \theta^*) \left(f(\theta) - f^*(\theta)\right) d\theta$$

Third, reduce marginal price from $\min\{p, p^*\}$ to 0 for units below $Q = \theta^*$. Both before and after the marginal price reduction, any type θ bought all units $q \leq \min\{\theta, \theta^*\}$. So all types generate additional fixed fees, partially offset by reduced marginal charges, on units $q \in [0, \min\{\theta, \theta^*\}]$ that increase profits by:

$$\min\{p, p^*\} \int_{\underline{\theta}}^{\bar{\theta}} \min\{\theta, \theta^*\} \left(f^*(\theta) - f(\theta)\right) d\theta$$

Now both the second and third price changes additionally affect profits extracted from all types for units $q \in [\theta, \theta + \delta_q]$. Even if the 2-part tariff with marginal price $\min\{p, p^*\}$ captured all the surplus on these units, first through the fixed fee ex ante, and then again in marginal fees ex post, while the 3-part tariff captured none, the loss would be at most $2\delta_q p$ (extraction of value above p is unaffected by the second and third price changes). Putting everything together, profits increase

by at least:

$$\begin{aligned}\Delta\Pi \quad \geq \quad & (p - \min\{p, p^*\})(1 - F(\theta^*))(E[\theta|\,\theta \geq \theta^*] - E^*[\theta|\,\theta \geq \theta^*]) \\ & + \min\{p, p^*\} F(\theta^*)(E^*[\theta|\,\theta \leq \theta^*] - E[\theta|\,\theta \leq \theta^*]) - 2\delta_q p - \delta_p E[\theta] - \delta_p \delta_q\end{aligned}$$

For symmetric distributions, this implies a profit increase of at least:

$$\Delta\Pi \geq \frac{p}{2}(E^*[\theta|\,\theta \leq \theta^*] - E[\theta|\,\theta \leq \theta^*]) - 2\delta_q p - \delta_p E[\theta] - \delta_p \delta_q$$

■

A.8 Proof of Claim in Footnote 25

Proof. Table 5 gives each party's perceived monopoly payoff under the standard and overconfident tariffs. The assumption $E^*\left[S^{FB}\right] = E\left[S^{FB}\right]$ ensures that overconfident types are indifferent between the two tariffs, because $E^*\left[S^{FB}\right] = E\left[\Psi^{FB}\right]$ (See proof of Lemma 3). Further, by Lemma 3, it guarantees that consumers with correct priors will weakly prefer the first best tariff.

	$E[U]$	$E^*[U]$	$E[\Pi]$
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 $\left\{\hat{q}, P^M\right\}$	$E[S^*] - E\left[\hat{\Psi}\right]$	0	$E\left[\hat{\Psi}\right]$

Table 5: Monopoly Payoffs

■

References

Alpert, M. and H. Raiffa, "A Progress Report on the Training of Probability Assessors," in Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press, 1982, pp. 294–305.

Bar-Gill, Oren, "Bundling and Consumer Misperception," *The University of Chicago Law Review*, 2006, *73* (1), 33–61.

Baron, D. P. and D. Besanko, "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1984, *1* (3), 267–302.

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

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

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang, "Projection Bias in Catalog Orders," *American Economic Review*, 2007, *97* (4), 1217–1249.

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.

Eliaz, Kfir and Ran Spiegler, "Contracting with Diversely Naive Agents," *The Review of Economic Studies*, 2006, *73* (3), 689–714.

___ and ___ , "Consumer Optimism and Price Discrimination," Working Paper February 2008.

Esteban, Susanna and Eiichi Miyagawa, "Optimal Menu of Menus with Self-Control Preferences," Working Paper 2005.

___ , ___ , and Matthew Shum, "Nonlinear Pricing with Self-Control Preferences," *Journal of Economic Theory*, 2007, *135* (1), 306–338.

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.

Hadar, Josef and William R. Russell, "Rules for Ordering Uncertain Prospects," *American Economic Review*, 1969, *59* (1), 25.

Hanoch, Giora and Haim Levy, "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies*, 1969, *36* (107), 335.

Jensen, Sissel, "Implementation of Competitive Nonlinear Pricing: Tariffs with Inclusive Consumption," *Review of Economic Design*, 2006, *10* (1), 9–29.

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.

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.

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.

Maskin, Eric and John Riley, "Monopoly with Incomplete Information," *RAND Journal of Economics*, 1984, *15* (2), 171–196.

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.

___ , "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.

Oi, Walter Y., "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," *Quarterly Journal of Economics*, 1971, *85* (1), 77–96.

Oster, Sharon M. and Fiona M. Scott Morton, "Behavioral Biases Meet the Market: The Case of Magazine Subscription Prices," *Advances in Economic Analysis & Policy*, 2005, *5* (1). Available at: `http://www.bepress.com/bejeap/advances/vol5/iss1/art1`.

Plous, S., "A Comparison of Strategies for Reducing Interval Overconfidence in Group Judgments," *Journal of Applied Psychology*, 1995, *80* (4), 443–454.

Riordan, Michael H. and David E. M. Sappington, "Awarding Monopoly Franchises," *American Economic Review*, 1987, *77* (3), 375–387.

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.

Sandroni, Alvaro and Francesco Squintani, "The Overconfidence Problem in Insurance Markets," Working Paper, SSRN 2004.

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.

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.

Uthemann, A., "Competitive Screening of Customers with Non-Common Priors," Working Paper November 9th 2005.

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