

Dynamic Nonlinear Pricing: biased expectations, inattention, and bill shock

Michael D. Grubb*

Abstract

Recent research highlights the importance of biased expectations and inattention for nonlinear pricing in dynamic environments. Findings are: (1) Three-part tariffs, such as cellular service contracts, exploit consumer overconfidence. (2) Surprise penalty fees may be used to further exploit biased beliefs or alternatively to price discriminate more efficiently whenever consumers are inattentive. (3) Implementing the recent bill-shock agreement between cellular carriers and the FCC is predicted to harm rather than help consumers when endogenous price changes are taken into account.

Keywords: nonlinear pricing, dynamic, inattention, overconfidence, bill shock, cellular

1. Introduction

Nonlinear pricing schemes, such as volume discounts, are important tools for converting consumer surplus into profits. In the large literature on nonlinear pricing, models are predominantly static. Nevertheless, nonlinear pricing often arises in dynamic environments where consumer purchase decisions are

*Massachusetts institute of Technology, Sloan School of Management.
mgrubb@gmail.com. www.mit.edu/~mgrubb

spread out over time. For instance, a cellular phone customer in the US must first choose a calling plan, which typically involves trading off higher monthly fees with larger buckets of *included* minutes. Only later does the consumer choose how many calls to make, which is done on a call-by-call basis throughout the course of a month. In such a dynamic environment, consumer behavior depends importantly both on expectations about future consumption patterns (such as when choosing a calling plan) and on consumer recall about past consumption (such as when trying to recall if one has already exhausted an allowance of included minutes).

Recent research provides compelling evidence from consumer behavior and firm pricing that our standard assumptions of rational expectations and perfect recall are both unrealistic and unsatisfactory for understanding nonlinear pricing in such dynamic environments. I review three recent papers that highlight the importance of modeling and measuring both bias and inattention to understand firms' pricing decisions and evaluate public policies such as the Federal Communication Commission's (FCC) recent bill-shock agreement with cellular carriers (Grubb, 2009, 2011; Grubb and Osborne, 2011).¹

In Grubb (2009) I show that consumer overconfidence can explain why cellular phone service plans include 'free' minutes followed by steep *over-*

¹Each of the papers reviewed here discusses related work on dynamic nonlinear pricing. Moreover, Rochet and Stole (2003, Section 8) survey dynamic nonlinear pricing models with standard consumers, such as Courty and Li (2000). Spiegel (2011) surveys models with nonstandard consumers, such as DellaVigna and Malmendier (2004) and Eliaz and Spiegel (2008).

age fees. The paper highlights the importance of consumers' expectations about future consumption at the time of contracting but still assumes that there is single consumption choice after a contract is signed. Two more recent papers relax this assumption by recognizing that consumption choices themselves are spread out over time. Consumer recall of past consumption becomes important in this environment, which allows me to examine the role of consumer inattention and its proposed remedy: bill-shock alerts that notify consumers when they reach consumption thresholds. In Grubb (2011) I show that bill-shock regulation will harm some consumers and lower total welfare in fairly competitive markets if consumers are inattentive but unbiased, but that such regulation can be beneficial if consumers underestimate their own future demand. In Grubb and Osborne (2011) we structurally estimate consumer biases from cellular-phone billing-data and counterfactually simulate the effect of implementing the recent bill-shock agreement between cellular carriers and the FCC. Initial results suggest that the agreement will be bad for consumers on average, assuming risk neutrality, although it will lower bill volatility.

2. Selling to Overconfident Consumers

Firms commonly offer three-part tariffs, or a menu 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 cellular phone service pricing in the US. The standard literature on nonlinear pricing does not provide a compelling explanation for such pricing patterns. Under

perfect competition one expects price to be driven down to cost, while standard nonlinear pricing models suggest the highest demand consumer will pay the lowest marginal price.

In Grubb (2009), I develop two theoretical explanations for observed three-part tariffs. The first explanation maintains the standard common-prior assumption and is based on price discrimination, while the second relies on consumer overconfidence. Using cellular phone customer billing data, I rule out the first explanation and conclude that overconfidence is a good explanation for three-part tariff pricing in the cellular context.²

At the heart of both explanations developed in Grubb (2009) is the recognition of one dynamic element of pricing: consumers must choose a calling plan when they are still uncertain about their future calling needs. The first potential explanation for three-part tariffs has the flavor of standard price discrimination, but consumers are separated based on the level of their uncertainty about future consumption rather than purely on the level of their demand. The second explanation is that consumers tend to underestimate the variance of their future demand when choosing a tariff.

Two important biases lead to the tendency of consumers to underestimate the variance of their future demand: forecasting overconfidence (sometimes called overprecision), which has been well documented in the psychology lit-

²Eliaz and Spiegel (2006) show that menus of three-part tariffs may be used to screen consumers with varying degrees of awareness about their own time-inconsistency. While I assume consumers buy all units from a single seller, Chao (2011) assumes that consumers buy from multiple sellers and argues that a dominant firm may use a three-part tariff to accommodate a competitor.

erature, and projection bias, which is described by Loewenstein et al. (2003). To understand the distinction, assume that in each period t consumer i realizes $\theta_{it} = \mu_i + \varepsilon_{it}$ potential calling opportunities (measured in minutes) where μ_i is consumer i 's average number of calling opportunities and ε_{it} captures volatility month-to-month. Moreover, assume that consumers do not know their own average tastes μ_i but form beliefs after receiving a noisy signal. Forecasting overconfidence implies a consumer will underestimate his own uncertainty about his average tastes μ_i . Projection bias implies that a consumer will underestimate the volatility of her tastes over time (the variance of ε_{it}). The two biases are isomorphic in the context of Grubb (2009) because the paper models a single consumption period and hence only beliefs about θ_{i1} matter. (The two biases are distinguished in Grubb and Osborne (2011) which models consumer learning about average tastes μ_i using panel data.)

Intuitively, underestimating variance of future demand can lead to three-part tariff pricing 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 own consumption. (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.” (Grubb, 2009)) Thus consumers expect to pay a low average price per unit, but sellers profit ex post when consumers make large revisions in either direction.

Grubb (2009) introduces a panel of billing records spanning February

2002 through June 2005 for approximately 2300 student accounts managed by a major US university for a national US cellular phone service provider. I find that customer tariff choices and subsequent usage decisions are 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 price discrimination model predicts that firms offer the ‘free’ minutes inherent in three-part tariffs only when cumulative usage distributions cross, which is ruled out by FOSD.)

A simple example from Grubb (2009) illustrates the role of overconfidence. Assume that firm marginal costs are 5 cents per minute and fixed costs are \$50 per customer. Consumers value each minute of calling at 45 cents up to some satiation point, beyond which they have no additional value for calling. When consumers sign contracts at time one, they are homogeneously uncertain about their satiation points. Then, at time two, consumers learn their satiation points and make their consumption choices. In particular, one third of consumers learn that they will be satiated after 100 minutes, one third after 400 minutes, and the remaining third after 700 minutes.

If consumers and the firm 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.³ A monopolist 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.

A monopolist 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.

³In a richer setting marginal cost pricing would be uniquely optimal.

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, low demand consumers use only 100 minutes but don't receive a refund for their unused allowance, while high demand consumers buy minutes 400-700 at the monopoly price of 45 cents per minute.

3. Bill Shock: Inattention and Price-Posting Regulation

In late 2011, US President Barack Obama said that,

Far too many Americans know what it's like to open up their cell-phone bill and be shocked by hundreds or even thousands of dollars in unexpected fees and charges. But we can put an end to that with a simple step: an alert warning consumers that they're about to hit their limit before fees and charges add up.

Obama's statement was made at the announcement of a voluntary agreement between cellular carriers and the FCC to begin providing such usage alerts by April 2013 (CTIA - The Wireless Association, 2011).

The situation Obama refers to, known as bill shock, can arise even when consumers are fully informed about contract terms. For example, a cellular-phone customer may know that the first 400 minutes are included while later minutes are charged at an overage rate of forty-five cents per minute. Nevertheless, he may be uncertain whether the next call will cost zero cents or forty-five cents because he does not know how many minutes he has already used. Such marginal-price uncertainty at the point of sale occurs whenever marginal prices vary with the level of consumption and, due to inattention, consumers are unaware of their past consumption when making additional consumption choices. Thus similar bill-shock regulation might be relevant to a wide variety of products and services including electricity, health insurance, and debit and credit-card transactions. For instance US checking account holders are often uncertain at the point of sale whether or not a \$35 overdraft fee will apply if they have not kept track of their bank balance and this could be addressed via balance alert regulation.⁴

Obama describes the new bill-shock agreement as part of his Administration's "ongoing efforts to protect American consumers by making sure financial transactions are fair, honest and transparent" (CTIA - The Wireless Association, 2011). Holding pricing fixed, usage alerts should at least weakly benefit consumers as Obama assumes by giving them more information to make better choices. However, firms will change prices in response to new disclosure requirements and it is therefore worth asking whether the bill-shock agreement or similar bill-shock regulation in other contexts will help

⁴Since 2010, American consumers must have opted-in to overdraft protection to be in this situation.

consumers or raise total welfare. This is the goal of two recent papers: Grubb (2011) models the effects of consumer inattention and bill-shock regulation theoretically and Grubb and Osborne (2011) evaluates bill-shock regulation in the specific context of cellular phones via structural estimation and counterfactual simulation.

To understand the effect of bill-shock regulation one must understand the answer to a related question: Why do firms both charge *penalty fees* (so that high usage triggers high marginal charges) and make them a *surprise* by not alerting consumers when they cross the relevant threshold?⁵

In Grubb (2011) I model consumer consumption choices and firm pricing responses when consumers are inattentive to past consumption and are aware of their own inattention when forecasting their future consumption choices.⁶ I show that there are at least two reasons for firms to charge inattentive consumers surprise penalty fees for excessive usage: First, surprise penalty fees can be useful for price discrimination. Second, surprise penalty fees can be useful for exploiting biases in consumer beliefs. Thus either heterogeneity or bias in consumer expectations at the time of contracting can explain both firms' use of surprise penalty fees and their resistance to regulation.

⁵Prior to its agreement with the FCC, the wireless industry trade group C.T.I.A. - The Wireless Association opposed proposed bill-shock regulation on the grounds that it "violates carriers' First Amendment protections. . . . against government compelled speech" (Altschul et al., 2011).

⁶Liebman and Zeckhauser (2004) analyze optimal pricing for consumers who confuse average price with marginal price (*ironing*) and for consumers who myopically consider only the marginal price of the current unit (*spotlighting*). In contrast, I assume consumers make choices optimally conditional on their limited memory.

Whether firms' use of surprise penalty fees is motivated by price discrimination concerns or their value in exploiting consumer bias matters for the consequences of regulation. I first consider the case in which consumers are unbiased ex ante but may have either a low or a high expectation about their future demand. A surprising result is that the combination of surprise penalty fees and consumer inattention can be socially valuable (as well as privately valuable to firms) and benefit some consumers by reducing allocative distortions imposed by price discriminating firms. Moreover this is always the case in fairly competitive markets. In the cellular context, firms clearly use packages of included minutes followed by overage charges to price discriminate across low and high usage segments. The implication is therefore that implementing the recent bill-shock agreement will hurt some consumers and lower total welfare if the cellular services market is fairly competitive and consumers are unbiased.

The intuition for the result that bill-shock regulation can harm some consumers and lower total welfare follows in two parts. Part one of the intuition is the standard logic of price discrimination with standard consumers: Suppose that if a firm could practice third-degree price discrimination it would offer low-usage consumers a discounted markup via a discounted fixed fee. When consumers self select contracts the firm must ensure that high-usage consumers are not tempted to choose a contract with a discounted markup intended for low-usage consumers. Thus, to offer low-usage consumers a discounted markup and discounted fixed fee, some marginal prices must be raised above marginal cost on the discounted contract to encourage high-usage consumers to pay a higher markup and fixed fee. When consumers are

attentive, this necessarily means that allocations will be distorted downwards inefficiently.

Part two of the intuition points out an important difference when consumers are inattentive: Price changes that keep expected marginal price equal to marginal cost remain efficient. For instance, assume marginal cost is five cents per minute and consider a contract that offers included minutes at zero marginal price followed by a surprise overage rate of fifty cents per minute. If consumers anticipate overages one in five months, then the expected marginal price is five cents and allocations are efficient. Nevertheless, this price structure can still encourage a higher-volume consumer (who would make an overage more than one in five months on the contract and hence face a higher expected marginal price) to pay a higher fixed fee and markup for a contract charging five cents per minute. Thus the combination of inattention and surprise penalty fees can allow firms to charge different markups without distorting allocations. Bill-shock regulation removes this option and forces firms to impose standard allocative distortions to price discriminate, thereby reducing total welfare. Because of this inefficiency, price discrimination is less profitable and firms will charge different customer groups more similar markups. This means raising markups for low-volume customers who are unambiguously worse off but lowering markups for high-volume consumers - the only group to benefit from the regulation. Nevertheless, consumer advocacy groups who do not take into account firms' price response to the regulation would be expected to advocate for it strongly - since holding prices fixed all consumers would be made better off.

Because Grubb (2009) finds evidence that cellular consumers are biased, I also examine the role of bill-shock regulation when consumers are homogeneous but biased. In this case firms use surprise penalty fees to exploit consumers' bias. Bill-shock regulation limits revenues from penalty fees because it helps consumers avoid accidental overages. Under monopoly this means that bill-shock regulation serves its intended role of consumer protection by limiting monopoly rents. In an oligopoly, however, if firms compete on fixed fees then bill-shock regulation will not effect equilibrium markups and consumers are residual claimants of social surplus. (Revenues from penalty fees are always rebated through lower fixed fees.) Moreover, whether the regulation will increase or decrease total welfare (and hence help or hurt consumers) will depend finely on the level of marginal costs and the nature of consumer bias. As a result, the total effect of bill-shock regulation in the cellular market is an empirical question, explored further in Grubb and Osborne (2011). However, Grubb (2011) shows that if fixed fees have already been competed down to zero, such as in the case of the US checking account market where *free checking* prevails, then bill-shock regulation can be expected to reduce equilibrium markups and benefit consumers if account holders underestimate the frequency of paying penalties (overdrafting). This result suggests that US checking account holders could benefit substantially if regulators moved beyond the recent adoption of opt-in regulation for overdraft fees to requiring zero-balance alerts.

4. Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock

Whether or not the recent bill-shock agreement will benefit consumers or be welfare improving depends on a number of factors including biases in consumer beliefs. It is therefore an empirical question whether the regulation will be good or bad. In Grubb and Osborne (2011), we develop and estimate a dynamic model of calling plan choice and calling usage using the panel of individuals' phone bills introduced in Grubb (2009). We identify the distribution of consumers' true demand for cellular calls from their actual usage. In addition, we identify consumers' beliefs about their future demand from their calling plan choices. The joint distribution of beliefs and realized demand determine whether beliefs are biased in the population. Moreover, we estimate the rate at which consumers learn about their own average demand overtime and switch plans to lower their bills. Given our estimates, we are able to make counterfactual simulations of bill-shock regulation and consumer de-biasing.⁷

We show that consumers are responsive to marginal prices (calling jumps at 9pm when off-peak hours begin) but do not adjust calling behavior throughout the billing cycle. Our explanation is that consumers are inattentive to their own past usage. (An attentive consumer would cut back calling following a period of high usage, and would do so more strongly if the period

⁷Most closely related are Goettler and Clay (2010), which measures consumer bias in the online grocery-delivery market, and Jiang (2011), which analyzes bill-shock regulation but assumes consumers are unbiased.

of high usage arose within the current billing cycle. We find no evidence of such behavior.) We therefore assume that consumers use a constant-threshold strategy, making all calls valued above a chosen level v^* .⁸ This threshold is constant throughout the month because consumers do not keep track of past usage and hence cannot condition calling behavior on past usage. By following this approach, our's is the first empirical model which can endogenously incorporate the information arrival due to usage alerts under a bill-shock regulation regime.

Our estimates show that consumers underestimate their own uncertainty about future calling demand. In Grubb (2009), I note that this may arise from forecasting overconfidence or projection bias but do not distinguish the two causes, often referring to both together as overconfidence. However, using consumers' switching behavior we are able to separately identify forecasting overconfidence and projection bias as two distinct biases and reserve the term overconfidence only for the first. Our preliminary estimates are that consumers underestimate their own uncertainty about their average demand by 82% (overconfidence) and that consumers underestimate the monthly volatility of their demand by 54% (projection bias). The relative magnitudes of the two biases are identified by consumers' low rate of switching and shed light, for instance, on the success of AT&T's *roll-over* minutes pricing which exploits overconfidence but not projection bias.

Holding prices fixed, our counterfactual simulations show that bill-shock regulation increases consumer welfare about \$21 per customer per year. No-

⁸Grubb (2011) shows that this is the optimal strategy for an inattentive consumer.

tably, we find that the large effect is due to the presence of bias. When all biases are removed, the effect drops to less than \$2 per customer per year. Of course holding prices fixed is unrealistic. In our endogenous price counterfactual simulations, preliminary results show that consumer welfare falls by about \$25 per person. Consumers reduce calling after receiving alerts that they have exceeded their allowance of minutes. As a result, overage revenues fall. Equilibrium markups are not affected, however, as firms make up the difference through higher fixed fees. Thus consumers make fewer phone calls but end up paying the same amount, so are worse off. This suggests that the long run effect of implementing the recent bill-shock agreement could be negative for consumers.

Altschul, Michael F., Christopher Guttman-McCabe, and Brian M. Josef, “Comments of CTIA - The Wireless Association,” January 10th 2011. <http://fjallfoss.fcc.gov/ecfs/document/view?id=7021025497>.

Chao, Yong, “Strategic Effects of Three-Part Tariffs,” 2011.

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

CTIA - The Wireless Association, “CTIA-The Wireless Association, Federal Communications Commission and Consumers Union Announce Free Alerts to Help Consumers Avoid Unexpected Overage Charges,” Technical Report October 17 2011. <http://www.ctia.org/media/press/body.cfm/prid/2137>.

- DellaVigna, Stefano and Ulrike Malmendier**, “Contract Design and Self-Control: Theory and Evidence,” *The Quarterly Journal of Economics*, 2004, 119 (2), 353–402.
- Eliasz, 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,” *Theoretical Economics*, 2008, 3 (4), 459–497.
- Goettler, Ronald L. and Karen B. Clay**, “Tariff Choice with Consumer Learning and Switching Costs,” Working Paper 2010.
- Grubb, Michael D.**, “Selling to Overconfident Consumers,” *American Economic Review*, 2009, 99 (5), 1770–1807.
- , “Bill Shock: Inattention and Price-Posting Regulation,” Working Paper 2011.
- and **Matthew Osborne**, “Cellular service demand: Tariff choice, usage uncertainty, biased beliefs, and learning,” Working Paper 2011.
- Jiang, Lai**, “The Welfare Effects of “Bill Shock” Regulation in Mobile Telecommunication Markets,” Working Paper July 2011.
- Liebman, Jeffrey B. and Richard Zeckhauser**, “Scheduling,” Working Paper October 2004.
- Loewenstein, George, Ted O’Donoghue, and Matthew Rabin**, “Projection Bias in Predicting Future Utility,” *The Quarterly Journal of Economics*, 2003, 118 (4), 1209–1248.

Rochet, Jean-Charles and Lars Stole, *The Economics of Multidimensional Screening*, Vol. 36 of *Econometric Society Monographs*, New York: Cambridge University Press, 2003.

Spiegler, Ran, *Bounded Rationality and Industrial Organization*, Oxford University Press, 2011.