Operations Research

Optimal Three-Part Tariff Plans

Gadi Fibich, Roy Klein, Oded Koenigsberg, Eitan Muller

OPERATIONS RESEARCH
Vol. 65, No. 5, September–October 2017, pp. 1177–1189
http://pubsonline.informs.org/journal/opre/
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Optimal Three-Part Tariff Plans

Gadi Fibich,[a] Roy Klein,[a] Oded Koenigsberg,[b] Eitan Muller[c,d]

[a] School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel; [b] London Business School, London NW1 4SA, United Kingdom; [c] Interdisciplinary Center (IDC), Herzelia 4610101, Israel; [d] New York University, New York, New York 10012

Contact: fibich@tau.ac.il (GF); roykleín.g@gmail.com (RK); okoenigsberg@london.edu (OK); emuller@stern.nyu.edu (EM)

Received: April 6, 2015
Revised: December 15, 2015; July 25, 2016; December 5, 2016
Accepted: December 15, 2016
Published Online in Articles in Advance: June 14, 2017

Subject Classifications: marketing: pricing
Area of Review: Marketing Science

https://doi.org/10.1287/opre.2017.1609

Abstract. Service providers, such as cell phone carriers, often offer three-part tariff plans that consist of three levers: A fixed fee, an allowance of free units, and a price per unit above the allowance. In previous studies the optimal three-part tariff contract was characterized using the standard first-order conditions approach. Because this optimization problem is nonsmooth, however, it could only be solved in a few simple cases. In this study we employ a different methodology that is based on obtaining a global bound for the firm profit, and then showing that this bound is attained by the optimal plan. This approach allows us to explicitly calculate the optimal three-part tariff plan under quite general conditions, where consumers are rational, they have a general utility function, they experience psychological costs when they exceed the number of free units, they have deterministic or stochastic consumption rates, they are homogeneous or heterogeneous, and the firm costs are fixed or depend on the usage level.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2017.1609.

Keywords: nonlinear pricing • three-part tariff • nonsmooth optimization

1. Introduction

Three-part tariff plans consist of a fixed fee (*access price*), the number of free units (*usage allowance*), and the price per unit above the number of free units (*overage price*). These contracts are popular in service industries such as the telecommunication industry (charging for each minute above the monthly allowance), car rentals (charging for miles above a mileage allowance), flights (charging for additional services), and Internet data storage. In this study we explicitly compute the optimal three-part tariff plan when consumers act rationally. We extend on previous work by considering consumers with a general valuation function and with a deterministic or random consumption rate. The consumers may be homogeneous or heterogeneous, and the firm cost may or may not depend on the usage level. We also take into account that consumers may incur a psychological cost when they exceed their allowance. For ease of exposition, we refer to the cellular phone market and use of cellular calling minutes as our unit of analysis.

Calculating the optimal firm strategy in the presence of rational consumers involves two nested optimization problems. The "inner" optimization problem is the calculation of the optimal strategy for consumers for any given three-part plan. From this calculation one obtains the firm's revenue from rational consumers under any three-part plan. Then the "outer" optimization problem is the calculation of the optimal three-part plan that maximizes the firm's revenue. Unfortunately, both the utility of the consumer and the firm revenue are *nonsmooth* at the point where the number of minutes used is equal to the monthly allowance. Since this nested optimization problem is non-smooth, the standard optimization approach, which is based on first-order conditions, leads to extremely long calculations that can only be solved in a few simple cases. For that reason, there have been few analytical results in the literature on optimal three-part tariffs plans.

In this study we avoid the nonsmoothness obstacle by adopting a different methodology, whereby we obtain a global bound on the firm's revenue under any three-part plan, and then find a plan that attains that bound. Therefore, this plan has to be optimal. This approach allows us to handle problems that are intractable using first-order conditions. Moreover, any plan that attains this bound is a global maximum, in contrast with the first-order conditions approach, where even if a solution can be found, it is not always clear whether it corresponds to a local or global maximum or minimum.

As noted, we assume that consumers are rational decision makers who seek to maximize their utility, which is the difference between their service value (service utility) from the minutes that they use, and the sum of (i) the monetary price that they pay to the firm and (ii) the psychological cost that they incur when they exceed the free minutes allowance. We allow for the consumers' usage rate to be deterministic or stochastic. The latter case corresponds to situations where consumers either cannot expect or cannot control how many minutes they will use (as is the case in the U.S. mobile market where consumers pay for incoming calls).

We find that when the firm costs are independent of consumers' usage and consumers are homogeneous, the optimal strategy for the firm is to let consumers use as many minutes as they want, which effectively reduces the three-part tariff plan to a fixed-price contract. This result, as well as all subsequent results, hold regardless of whether the usage rate is deterministic or stochastic. Thus, the firm sets a sufficiently high allowance, guaranteeing that consumers never exceed it. Therefore, consumers attain their maximal service value. Then the firm sets the fixed fee to be equal to consumers' maximal service value, which effectively reduces the consumers' overall utility to zero. In this contract, the marginal price per minute is irrelevant. We also find that the firm's revenue decreases as the consumer consumption rate becomes more stochastic.

The above result may seem to suggest that in the case of homogeneous consumers, a three-part tariff plan is not needed. However, allowing consumers to use as many minutes as they want is not the optimal strategy when the firm incurs a cost for every minute that consumers talk. In such a case, the firm should set a usage allowance and prevent consumers from exceeding it by charging a sufficiently high per-minute overage price. The usage allowance threshold is the point at which the consumers' marginal service value from talking becomes equal to the firm's marginal cost. Therefore, even when consumers are homogeneous, a three-part plan is needed if the firm costs are taken into account.

To investigate the case in which consumers are heterogeneous, we divide them into two segments of heavy and light users. We analyze this problem under both deterministic and stochastic demand. A priori, when the firm offers one plan for all users, there are two potential optimal strategies. The first is to target the heavy consumers exclusively. In this case, the firm allows the heavy users to talk as much as they want and sets the fixed fee to be equal to their maximal valuation from talking. The light consumers do not join the plan, because the fixed fee is too high for them. The second strategy is to target both consumer segments. The intuitive contract in that case is to maximize the firm's profit from light consumers through the fixed fee by allowing them to talk as much as they want, and then maximize the extra profits from the heavy users with a proper choice of the per-minute overage charge. Interestingly, however, this contract is suboptimal. Rather, both the fixed fee and the usage allowance should be lower than those that extract the maximal profit from the light users. The firm can also choose to offer two three-part tariff plans: one that allows the light ones to talk as much as they want and a second plan that maximizes the revenues from the heavy users. Adding a second plan increases the firm profits, compared to a single plan. Even with two plans, however, allowing the light users to talk as much as they want is always suboptimal. Whether the firm should focus on the heavy users exclusively or on all users depends on the level of heterogeneity in the consumers' valuations and on the ratio of the number of heavy to light users.

1.1. Literature Review

Nonlinear pricing was studied in the economics, operations research, and marketing literature. Most of the literature on three-part tariff plans is empirical or numerical, and only a single paper calculated the optimal three-part tariff plan analytically. Lambrecht et al. (2007) considered a three-part tariff under uncertainty associated with Internet data packages. They set up a quadratic utility function and estimated the demand. They did not, however, determine the optimal packages. Rather they measured the consumers' preferences for flat-rate plans relative to pay-per-use plans and found it to be significant. Iyengar et al. (2008) considered three-part tariff plans for mobile phone services. They used conjoint data to estimate the model parameters and then used a grid search to compute the optimal plans numerically. Iyengar et al. (2007) analyzed data from a single wireless service provider. They developed a model for plan choice and consumption that incorporates consumers' usage uncertainty and consumers' learning for service quality and usage. Ascarza et al. (2012) considered the effect of the free allowance part on the consumer's choice in a three-part tariff pricing. The setting was that the firms add a three-part tariff plan to their existing menu that consisted exclusively of two-part tariff plans. Optimal packages, however, were not one of the objectives of these papers.

Iyengar et al. (2007) and Lambrecht et al. (2007) considered the randomness of the consumption rate. In those studies, the consumer chooses the optimal number of minutes assuming he has a deterministic consumption rate. Only then, the uncertainty in the consumption rate is taken into account by the consumer (who decided whether to join the plan) and by the firm (in determining its expected profits). In our model, the consumer chooses his desired consumption rate while taking into account the uncertainty in his or her consumption rate. This makes the consumer optimization problem more challenging to compute, but it makes the model more realistic.

Several studies on nonlinear pricing in service industries examined two-part tariff plans. Essegaier et al. (2002) computed the optimal two-part tariff plan under constraints on service capacity and heterogeneous consumer use. They assumed that usage rates of individual consumers vary and that the marginal cost of serving a customer is low and independent of the consumers usage rate. They showed that flat-fee pricing is the only sustainable pricing structure once the industry has developed sufficient excess capacity. Cachon and Feldman (2011) asked whether a firm should charge per use or sell subscriptions when congestion is unavoidable and found that subscription pricing is preferable, despite its limitations with respect to congestion. Desai et al. (2016) study the role of family plans in the telecommunication industry.

A few studies investigated some characteristics of three-part tariff pricing (see Huang 2008 and Kim et al. 2010 for a review of those studies). None of these studies, however, calculated the optimal three-part tariff plan. For example, Bagh and Bhargava (2013) analyzed the ability of alternative nonlinear pricing structures to price discriminate. They showed that three-part tariffs are more efficient than two-part tariffs as price-discriminating mechanisms for heterogeneous consumers.

We are only aware of a single paper that calculated optimal three-part tariff optimization problem analytically. Grubb (2009) computed the optimal three-part tariff plan when consumers are overconfident, by assuming that each consumer has an estimated demand and an actual demand and chose a plan based on the estimated demand. He showed that for consumers who are not overconfident, the firm's optimal strategy is to offer a plan that has a high fixed fee and thus takes all of the surplus of the consumers. Furthermore, the firm earns a greater profit when consumers are overconfident. In that model, the firm knows both the estimated and actual demand of the consumers, but consumers only know their estimated demand. We consider a different situation of symmetric information between the firm and the consumers. In addition, in Grubb's model, consumers have a predetermined number of minutes that they want to use. Therefore, they only have to decide whether to join the calling plan. In our model, the number of minutes consumers want to use depends on the calling plan parameters. Hence, our model leads to a nested optimization problem, whereas Grubb's model does not.

Our paper can also be linked to the rich literature on product lines that dates back to the seminal paper by Mussa and Rosen (1978) (see also Moorthy 1984, Johnson and Myatt 2003, and Villas-Boas 2004). In the models in those studies, consumers differed in how much they valued product quality. The firm knew the distribution of consumers' taste for quality but could not identify the tastes of individual consumers. The firm offered multiple products, and consumers self-selected the product that matched their tastes. In our work, consumers differ in preferred rates of consumption. The firm knows the distribution of consumers' taste for consumption but cannot identify the tastes of individual consumers. The firm offers three-part tariff contracts (more, obviously, when it offers multiple three-part contracts), and consumers self-select how many minutes to consume given their contract plan, which is a de facto differentiation of consumer segments based on their preferences. Their self-selection creates a product line in which the products differ according to the individuals' rates of consumption. A firm offering two three-part tariff contracts is equivalent to introduction of a regular product line if the overage price is decided by a regulator or any other external entity. The firm chooses the fixed fees and usage allowances, which correspond to the products' prices and levels of quality. Therefore, the time allowances act as the perceived quality of the plans, and customers self-select a package, which is equivalent to choosing different products (quality and price). The equivalence breaks down, however, when an overage price is added. In that case, the three-part tariff contracts are equivalent to consumers buying additional bits of quality for an additional price that is decided by the firm. Our paper also relates to studies of product lines that capture heterogeneity in consumers' consumption rates. In Koenigsberg et al. (2010), for example, the authors model a firm's decisions about quality, price, and package size when the consumption rate is exogenous. In our study, each consumer's consumption rate is a decision variable determined by the underlying distribution of the consumption rate, the consumer's degree of uncertainty, and the contract parameters.

The paper is organized as follows. In Section 2 we compute the optimal three-part tariff plan when consumers are homogeneous and have a deterministic demand, and the firm costs are independent on consumers' usage level. In Section 3 we allow the firm's costs to depend on consumers' usage. In Section 4 we analyze the case of heterogeneous consumers, and in Section 5 we show how the results can be extended to the case of consumers with a stochastic demand. Section 6 concludes with a discussion. To streamline the presentation, most proofs are relegated to the appendix.

2. Homogeneous Consumers with a Deterministic Demand

Consider a market with rational consumers whose valuation from talking $x \geqslant 0$ minutes is

$$V(x) = \int_0^x v(y)\,dy, \tag{1}$$

where $v(x)$ is the consumer surplus valuation for the x minute. We assume that $v(x)$ is continuous, $v(x) > 0$ for $0 \leqslant x < x_V^{\max}$ and $v(x) < 0$ for $x > x_V^{\max}$, where $0 < x_V^{\max} < \infty$. Therefore, $V(x)$ is continuously differentiable, its global maximum is positive, finite, and is attained at $x_V^{\max}$, i.e.,

$$x_V^{\max} := \arg\max_{x\geqslant 0} V(x), \quad V^{\max} := V(x_V^{\max}), \\ 0 < x_V^{\max} < \infty, \qquad 0 < V^{\max} < \infty. \tag{2}$$

Thus, when unrestricted, a rational consumer will talk exactly $x_V^{\max}$ minutes.

The assumption that the consumer maximal valuation is attained at a finite $x_V^{\max}$ is essential for the analysis. There are two possible approaches to justify this assumption:

1. Assumption (2) is satisfied by the quadratic valuation function $V(x) = \alpha_1 x - \alpha_2 x^2$ that is common in the empirical literature on two-part and three-part tariff pricing (see, for example, Iyengar et al. 2007, Lambrecht et al. 2007, Iyengar et al. 2008, Ascarza et al. 2012). Furthermore, the assumption that the surplus valuation becomes negative above a finite $x_V^{\max}$ is consistent with empirical evidence that consumers with unlimited plans speak well below 24 hours per day.

Nevertheless, this assumption on $V(x)$ seemingly violates the conditions of monotonicity and local non-satiation that are fundamental in microeconomic modeling of consumer preferences (see e.g., Mas-Colell et al. 1995). While this is true for a general valuation function, since the variable x is number of minutes per period, say a day, the valuation function V contains an implicit constraint: a limit X that the consumer has per period on the time available (e.g., 24 hours per day). Moreover, if the consumer does not use all the available time for one activity (talking over the phone), the consumer has other uses for it. Thus, we posit the second approach of achieving this condition:

2. Assume that the consumer has a finite budget constraint $x \leqslant X < \infty$, and that her valuation when talking x minutes is $V(x) = \int_0^x v_1(y)\,dy + \int_0^{X-x} v_2(y)\,dy$, where $v_1(y)$ and $v_2(y)$ are her surplus valuations from talking and from all the alternative usage of her time, respectively. We then have the following result:

Lemma 1. *Assume that $v_1(y)$ and $v_2(y)$ are positive and monotonically decreasing in y. If $v_1(X) < v_2(0)$ and $v_2(X) < v_1(0)$, Then $V(x)$ satisfies* (2).

Proof. We have that $V(x) = \int_0^x v_1(y)\,dy + \int_0^X v_2(y)\,dy - \int_{X-x}^X v_2(y)\,dy = C_2 + \int_0^x v(y)\,dy$, where $C_2 = \int_0^X v_2(y)\,dy$ is a constant and $v(y) = v_1(y) - v_2(X-y)$. Since $v(0) > 0$, $v(X) < 0$, and $v'(x) = v_1'(x) + v_2'(X-x) < 0$, there exists a unique $0 < x_V^{\max} < X$ such that $v(y)$ is positive for $y < x_V^{\max}$ and negative for $y > x_V^{\max}$. Consequently, $\max V(x)$ is finite, and is attained a finite x.

Note that Lemma 1 provides a theoretical foundation for satiated utility functions that are used in the empirical literature.

A monopolistic service provider (firm) offers a monthly plan (p, T, F), such that if a consumer signs up to the plan, she pays a fixed fee of F dollars ("*access fee*") and in return gets T minutes of free calls. For every minute in excess of T, the consumer pays an additional price of p dollars per minute. Thus, the firm's revenue from a consumer that talks x minutes is

$$\pi(x,p,T,F) = \begin{cases} F, & \text{if } x \leqslant T, \\ F + p(x-T), & \text{if } x > T. \end{cases} \tag{3}$$

We assume that when a consumer is charged $p(x-T)$ for exceeding his monthly allowance, he may experience a "psychological cost," which we denote by $S(x,p,T)$. Therefore,

$$\begin{cases} S(x,p,T) = 0, & \text{if } x \leqslant T, \\ S(x,p,T) \geqslant 0, & \text{if } x > T. \end{cases} \tag{4}$$

This effect was not considered in previous studies of three-part tariff plans but is consistent with prospect theory. The consumer's utility $U(x,p,T,F)$ is the difference between his valuation of the service and his monetary and psychological costs, i.e.,

$$U(x,p,T,F) = V(x) - \pi(x,p,T,F) - S(x,p,T). \tag{5}$$

Therefore,

$$U(x,p,T,F) = \begin{cases} V(x) - F, & \text{if } x \leqslant T, \\ V(x) - F - p(x-T) & \\ \quad - S(x,p,T), & \text{if } x > T. \end{cases} \tag{6}$$

For a given plan (p,T,F), the optimal number of minutes for a consumer is

$$x_U^{\text{opt}}(p,T,F) := \arg\max_{x \geqslant 0} U(x,p,T,F). \tag{7}$$

In this case, his utility is

$$\begin{aligned} U^{\text{opt}}(p,T,F) &:= \max_{x \geqslant 0} U(x,p,T,F) \\ &= U(x_U^{\text{opt}}(p,T,F), p, T, F). \end{aligned} \tag{8}$$

A rational consumer signs up to the plan (and talks x_U^{opt} minutes) if $U^{\text{opt}}(p,T,F) > 0$, but the consumer does not sign up to the plan if $U^{\text{opt}}(p,T,F) < 0$. When $U^{\text{opt}}(p,T,F) = 0$, the consumer is "indifferent" between signing or not signing. In practice, the firm can always set a slightly lower fixed fee, leading the consumer to sign up. Hence, from now on we assume that if $U^{\text{opt}}(p,T,F) = 0$, the consumer signs up to the plan.

When the firm offers a plan (p,T,F), its revenue per (rational) consumer is

$$\Pi(p,T,F) := \begin{cases} \pi(x_U^{\text{opt}}(p,T,F), p, T, F), & \text{if } U^{\text{opt}}(p,T,F) \geqslant 0, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The firm optimization problem is to find the plan $(p^{\text{opt}}, T^{\text{opt}}, F^{\text{opt}})$ that maximizes its profits:

$$(p^{\text{opt}}, T^{\text{opt}}, F^{\text{opt}}) = \arg\max_{p,T,F \geqslant 0} \Pi(p,T,F).$$

Note that to find the optimal firm plan, one first needs to calculate the optimal consumer response; see (7). This nested optimization problem is nonsmooth, because $U(x,p,T,F)$ is not smooth at $x=T$. Therefore, it cannot be solved using the first-order conditions, except in some very simple cases. This nonsmooth nested optimization problem can be solved explicitly using a different mathematical approach, leading to

Proposition 1. *The optimal firm plan is*

$$F^{\text{opt}} = V^{\max}, \quad T^{\text{opt}} \geqslant x_{\text{V}}^{\max}, \quad p^{\text{opt}} \geqslant 0, \tag{10}$$

where $V^{\max}$ and $x_{\text{V}}^{\max}$ are defined in (2). In addition,

1. *the consumer talks $x_{\text{V}}^{\max}$ minutes, i.e., as much as she would in an unlimited plan;*
2. *the consumer utility is* 0;
3. *the firm revenue is $V^{\max}$.*

Proof. This is a special case of Proposition 5.

Thus, the optimal firm strategy is to let consumers talk as much as they want, so that they would maximize their valuation. Therefore, it sets $T^{\text{opt}} \geqslant x_{\text{V}}^{\max}$. Then, it extracts all their utility through the fixed fee. Since the consumers do not exceed their allowance, the value of p^{opt} is insignificant.

For future reference, we note the following result:

Lemma 2. *There is no optimal strategy in which a portion of the firm revenues comes from overage usage, i.e., there is no optimal strategy with $F < V^{\max}$ and $T < x_{\text{V}}^{\max}$.*

Proof. Assume that there is an optimal strategy with $F < V^{\max}$. Then $x_{\text{V}}^{\max} > T$ and $p > 0$, since otherwise the firm revenue will be F, which is suboptimal. When a consumer exceeds T he incurs psychological costs that reduce his utility. Even if psychological costs are neglected, since a rational consumer stops talking once $V'(x) \leqslant p$, he talks less than $x_{\text{V}}^{\max}$ minutes. Therefore, his utility will be smaller than $V^{\max}$. Since the overall payment of the consumer cannot exceed his utility, the firm revenues will be smaller than $V^{\max}$.

3. Variable Firm Cost

In Proposition 1 we saw that the optimal firm strategy is to let consumers talk as much as they want and then extract all their utility using the fixed fee. This is no longer true, however, when the firm cost depends on the number of minutes that consumers use, since then above a certain usage level the consumer's marginal utility becomes smaller than the firm marginal cost.

To analyze this case, we denote by $C(x)$ the firm cost when a consumer talks x minutes. The firm revenue per consumer is the difference between its profits and costs, i.e.,

$$\pi_{\text{c}}(x,p,T,F) = \pi(x,p,T,F) - C(x).$$

Thus,

$$\pi_{\text{c}}(x,p,T,F) = \begin{cases} F - C(x), & \text{if } x \leqslant T, \\ F - C(x) + p(x-T), & \text{if } x > T. \end{cases}$$

Consequently, the firm optimization problem reads

$$(p^{\text{opt}}, T^{\text{opt}}, F^{\text{opt}}) = \underset{p,T,F \geqslant 0}{\arg\max}\, \Pi_{\text{c}}(p,T,F),$$

where

$$\Pi_{\text{c}}(p,T,F) := \begin{cases} \pi_{\text{c}}(x_{\text{U}}^{\text{opt}}(p,T,F), p, T, F), & \text{if } U^{\text{opt}}(p,T,F) > 0, \\ 0, & \text{otherwise}, \end{cases}$$

and $x_{\text{U}}^{\text{opt}}$ and U^{opt} are given by Equations (7) and (8), respectively.

Proposition 2. *Suppose that $V(x)$ is concave, $C(x)$ is monotonically increasing, and $V(x) - C(x)$ has a unique global maximum at*

$$x_{\text{V},c}^{\max} := \underset{x \geqslant 0}{\arg\max}\{V(x) - C(x)\}. \tag{11}$$

Then the optimal firm plan is

$$F^{\text{opt}} = V(x_{\text{V},c}^{\max}), \quad T^{\text{opt}} = x_{\text{V},c}^{\max}, \quad p^{\text{opt}} \geqslant p_{\text{c}},$$

where

$$p_{\text{c}} := \max_{x \geqslant x_{\text{V},c}^{\max}} \left\{ \frac{V(x) - V(x_{\text{V},c}^{\max})}{x - x_{\text{V},c}^{\max}} \right\} \tag{12}$$

is the minimal optimal overage price. In addition,

1. *the consumer talks $x_{\text{V},c}^{\max}$ minutes, where $0 < x_{\text{V},c}^{\max} < x_{\text{V}}^{\max}$;*
2. *the consumer utility is zero;*
3. *the firm revenue is $V(x_{\text{V},c}^{\max}) - C(x_{\text{V},c}^{\max})$.*

Proof. See web appendix.

Thus, when the firm offers an unlimited plan ($T=\infty$), the maximal fixed fee that a consumer who wants to talk x minutes is willing to pay is $F = V(x)$. In this case, the firm's revenue is $F - C(x) = V(x) - C(x)$. Therefore, from the firm perspective, the maximal revenue is attained where the consumer talks $x_{\text{V},c}^{\max}$ minutes; see (11). From the consumer perspective, however, her maximal utility is attained when she talks $x_{\text{V}}^{\max}$ minutes; see (2). Since $x_{\text{V},c}^{\max} < x_{\text{V}}^{\max}$, the firm has to "convince" the consumer to use exactly $x_{\text{V},c}^{\max}$ minutes. To do that, the firm sets $T = x_{\text{V},c}^{\max}$, so that the consumer pays no overage fee when she uses $x = x_{\text{V},c}^{\max}$ minutes, and pays an overage fee when she uses $x > x_{\text{V},c}^{\max}$. In addition, the firm sets the minimal overage price p_{c} so that for any $x > x_{\text{V},c}^{\max}$, the overage payment will be greater than the additional valuation gained from exceeding $x_{\text{V},c}^{\max}$, i.e., so that $p(x - x_{\text{V},c}^{\max}) > V(x) - V(x_{\text{V},c}^{\max})$. This guarantees that consumers will not benefit from exceeding $x_{\text{V},c}^{\max}$.

If $V(x)$ is concave, then by (12), the mean value theorem, and the concavity of $V(x)$,

$$p_c = V'(x^{\max}_{V,c}). \tag{13}$$

In other words, p should be greater than the marginal valuation at $x^{\max}_{V,c}$. In particular, if $C(x) = cx$, then by (11) and (13),

$$p_c = V'(x^{\max}_{V,c}) = C'(x^{\max}_{V,c}) = c. \tag{14}$$

We recall that when the firm costs are negligible, the firm only uses one out of three levers possible under the three-part tariff contract. Thus, the contract is effectively reduced to a fixed-price contract where consumers can use as many minutes as they desire. In contrast, in the case of variable firm costs, the firm uses all three levers: the fixed fee F, the number of free minutes T, and a sufficiently large overage price p. Note that even when the firm incurs variable costs, it still extracts all of the consumer's utility via the fixed fee.

3.1. Parametric Example

The quadratic valuation function

$$V(x) := \alpha_1 x - \alpha_2 x^2 \tag{15}$$

is common in the three-part tariff literature. The maximum of $V(x)$ is attained at $x^{\max}_V = \alpha_1/(2\alpha_2)$ and is given by $V^{\max} := V(x^{\max}_V) = \alpha_1^2/(4\alpha_2)$. We use the values $\alpha_1 = 37 \cdot 10^{-2}$ dollars/minute and $\alpha_2 = 4.14 \cdot 10^{-4}$ dollars/minute^2, which were estimated by Iyengar et al. (2008) from a conjoint study.

We begin with the case of constant firm costs. By Proposition 1, the optimal firm plan is

$$F^{\text{opt}} = \frac{\alpha_1^2}{4\alpha_2} = \$83, \quad T^{\text{opt}} \geqslant \frac{\alpha_1}{2\alpha_2} = 447 \text{ minutes}, \quad p^{\text{opt}} \geqslant 0.$$

Hence, the optimal firm revenue is $\Pi(p^{\text{opt}}, T^{\text{opt}}, F^{\text{opt}}) = F^{\text{opt}} = \$83$.

To include variable firm costs, we consider a linear cost function $C(x) = cx$. It is easy to check that

$$x^{\max}_{V,c} = \arg\max\{\alpha_1 x - \alpha_2 x^2 - cx\} = \frac{\alpha_1 - c}{2\alpha_2} = 447 - 1{,}208c.$$

Therefore, $V(x^{\max}_{V,c}) = (\alpha_1^2 - c^2)/(4\alpha_2) = 83 - 604c^2$. In addition, by (14), $p_c = c$. Therefore, by Proposition 2, the optimal firm plan is

$$F^{\text{opt}} = \$(83 - 604c^2), \qquad T^{\text{opt}} = (447 - 1{,}208c) \text{ minutes}, \quad p^{\text{opt}} \geqslant c, \tag{16}$$

and the income derived from the optimal firm plan is

$$\Pi(p^{\text{opt}}, T^{\text{opt}}, F^{\text{opt}}) = F^{\text{opt}} - cx^{\max}_{V,c} = \frac{(\alpha_1 - c)^2}{4\alpha_2} = \$(\sqrt{83} - \sqrt{604}c)^2.$$

As expected, the firm revenue decreases with c.

4. Heterogeneous Consumers

To analyze the effect of consumers heterogeneity, we consider a market that consists of n_L light users with utility $U_L = V_L - \pi - S_L$ and n_H heavy users with utility $U_H = V_H - \pi - S_H$. We assume that in an unlimited plan ($T = \infty$), heavy users want to use more minutes than the light ones, i.e.,

$$x^{\max}_{V,L} < x^{\max}_{V,H}, \tag{17}$$

where $x^{\max}_{V,i} = \arg\max_{x \geqslant 0} V_i(x)$ and $i = $ L,H. We also assume that the maximal valuation of the light users is smaller than that of the heavy ones, i.e.,

$$V^{\max}_L < V^{\max}_H, \tag{18}$$

where $V^{\max}_i = V_i(x^{\max}_{V,i}) = \max_{x \geqslant 0} V_i(x)$. The psychological cost of the light and heavy users satisfy (4). In addition, we assume that the psychological cost of the heavy users is of the form

$$S_H(x, p, T) = \int_T^x s_H(y, p)\, dy, \quad x \geqslant T, \tag{19}$$

and that the marginal psychological cost s_H is positive, independent of T and F, and satisfies the relation $\lim_{p \to 0} s_H(y, p) = 0$.

4.1. Optimal Single Plan

The optimal plan $(p^{\text{opt}}, T^{\text{opt}}, F^{\text{opt}})$ is the one that maximizes the average firm revenue per consumer

$$\Pi(p, T, F) = \gamma_H \Pi_H(p, T, F) + (1 - \gamma_H)\Pi_L(p, T, F),$$

where $\gamma_H = n_H/(n_L + n_H)$ is the fraction of heavy users, $\Pi_H(p, T, F)$ is defined by (9) with $U^{\text{opt}} = U^{\text{opt}}_H = \max_{x \geqslant 0} U_H$ and $x^{\text{opt}}_U = x^{\text{opt}}_{U,H} := \arg\max_{x \geqslant 0} U_H$, and similarly for $\Pi_L(p, T, F)$.

Lemma 3. *There is no three-part tariff plan that extracts the maximal revenues from both light and heavy users. In other words, for any plan (p, T, F),*

$$\Pi(p, T, F) < \gamma_H V^{\max}_H + (1 - \gamma_H) V^{\max}_L.$$

Proof. The only way to extract the maximal revenue from each segment is through the fixed fee (Lemma 2). Since $V^{\max}_L < V^{\max}_H$, however, this is not possible.

One possible firm strategy is to focus on the heavy users:

Lemma 4. *The optimal firm plan that maximizes revenue from heavy users is to allow them to talk as much as they want and then extract all of their utility through the fixed fee, i.e.,*

$$F^{\text{opt}}_H = V^{\max}_H, \quad T^{\text{opt}}_H \geqslant x^{\max}_{V,H}, \quad p^{\text{opt}}_H \geqslant 0.$$

In this case we have the following:

1. *Heavy users sign up to the plan and use $x^{\max}_{V,H}$ minutes (i.e., as much as they would in an unlimited plan). Their utility is zero.*

2. *Light users do not sign up to the plan.*
3. *The firm revenue per consumer is*

$$\Pi_{\text{H-only}} := \Pi(p_{\text{H}}^{\text{opt}}, T_{\text{H}}^{\text{opt}}, F_{\text{H}}^{\text{opt}}) = \gamma_{\text{H}} V_{\text{H}}^{\max}. \tag{20}$$

Proof. The optimal firm strategy follows from Proposition 1. Since $V_{\text{L}}^{\max} < V_{\text{H}}^{\max} = F_{\text{H}}^{\text{opt}}$, light users will not sign up to the plan.

Another possible firm strategy is to focus on the light users:

Lemma 5. *The optimal firm plan that maximizes revenue from light users is to allow them to talk as much as they want, extract all their utility through the fixed fee, and then maximize the revenue from the heavy users by a proper choice of p and T, i.e.,*

$$F_{\text{L}}^{\text{opt}} = V_{\text{L}}^{\max}, \quad T_{\text{L}}^{\text{opt}} = x_{\text{V},L}^{\max}, \\ p_{\text{L}}^{\text{opt}} = \arg\max_{p \geqslant 0} \{ p(\tilde{x}_{\text{U},H}^{\text{opt}}(p) - x_{\text{V},L}^{\max}) \} > 0, \tag{21}$$

where $\tilde{x}_{\text{U},H}^{\text{opt}}(p) = \arg\max_{x \geqslant 0} U_{\text{H}}(x, p, T_{\text{L}}^{\text{opt}}, F_{\text{L}}^{\text{opt}})$. In this case we have the following:

1. *Light users sign up to the plan and use $x_{\text{V},L}^{\max}$ minutes (i.e., as much as they would in an unlimited plan). Their utility is zero.*
2. *Heavy users sign up to the plan and use $x_{\text{U},H}^{\text{opt}} = \arg\max_{x \geqslant 0} U_{\text{H}}(x, p_{\text{L}}^{\text{opt}}, T_{\text{L}}^{\text{opt}}, F_{\text{L}}^{\text{opt}})$ minutes, where $x_{\text{V},L}^{\max} < x_{\text{U},H}^{\text{opt}} < x_{\text{V},H}^{\max}$. Thus, they pay for overage usage, and do not use as many minutes as they would in an unlimited plan. Their utility is positive.*
3. *The firm revenue per consumer is*

$$\Pi_{\text{L-mainly}} := \Pi(p_{\text{L}}^{\text{opt}}, T_{\text{L}}^{\text{opt}}, F_{\text{L}}^{\text{opt}}) \\ = V_{\text{L}}^{\max} + \gamma_{\text{H}} p_{\text{L}}^{\text{opt}} (x_{\text{U},H}^{\text{opt}} - x_{\text{V},L}^{\max}). \tag{22}$$

In particular, $\Pi_{\text{L-mainly}} > V_{\text{L}}^{\max}$.

Proof. See web appendix.

Thus, the firm maximizes its profits from the light users by setting T to be at least the number of minutes they want to talk, and extracting all their utility through the fixed fee. Unlike the optimal plan for homogeneous light users (Proposition 1), however, the firm sets p and T not only to maximize its revenues from the light users, but also to maximize its revenues from heavy users. Hence, the firm sets T to be *equal* to the number of minutes that light users want to talk, since a larger T will allow the heavy users to talk more minutes without paying for them. In addition, p cannot be any positive price, because it should maximize the revenue from heavy users.

Thus, the firm's revenue consists of the fixed fee $V_{\text{L}}^{\max}$ that both light and heavy users pay, and the overage payment $p_{\text{L}}^{\text{opt}}(x_{\text{U},H}^{\text{opt}} - x_{\text{V},L}^{\max})$ of the heavy users for exceeding T. Note that the firm fails to extract all of the surplus from heavy users, who are thus subsidized by the light users.

A priori, one might think that when they are "few" heavy users, the optimal plan is given by (21). We now show, however, that maximizing the revenues from the light users is never an optimal strategy:

Proposition 3. *Any optimal plan $(p^{\text{opt}}, T^{\text{opt}}, F^{\text{opt}})$ that targets both heavy and light users satisfies $F^{\text{opt}} < V_{\text{L}}^{\max}$ and $T^{\text{opt}} < x_{\text{V,L}}^{\max}$. Hence, plan (21) cannot be optimal.*

Proof. Since light users sign up to the plan, $F^{\text{opt}} \leqslant V_{\text{L}}^{\max}$. Assume by negation that $F^{\text{opt}} = V_{\text{L}}^{\max}$. In that case, the optimal plan is given by Lemma 5. In particular, $T^{\text{opt}} = x_{\text{V,L}}^{\max}$. To show that this plan is not optimal, we now show that the firm revenues increase if T and F decrease to $T^- := x_{\text{V,L}}^{\max} - \Delta T$ and $F^- := V_{\text{L}}(x_{\text{V,L}}^{\max} - \Delta T)$, respectively, for $\Delta T \ll 1$ sufficiently small.

Under plan $(p^{\text{opt}}, T^-, F^-)$, if the light users will use $x = x_{\text{V,L}}^{\max} - \Delta T$, their utility will be $U_{\text{L}}(x = x_{\text{V,L}}^{\max} - \Delta T, p^{\text{opt}}, T^-, F^-) = V_{\text{L}}(x_{\text{V,L}}^{\max} - \Delta T) - F^- = 0$. Hence, they will sign up to the plan. Regardless of whether they talk more than $x_{\text{V,L}}^{\max} - \Delta T$, the firm revenue from them will be at least F^+. Hence, $\Pi_{\text{L}}(p^{\text{opt}}, T^-, F^-) \geqslant V_{\text{L}}(x_{\text{V,L}}^{\max} - \Delta T)$. Since $x_{\text{V,L}}^{\max} = \arg\max V_{\text{L}}(x)$, then $V_{\text{L}}'(x_{\text{V,L}}^{\max}) = 0$ and $V_{\text{L}}''(x_{\text{V,L}}^{\max}) < 0$. Therefore,

$$V_{\text{L}}^{\max} - F^- = V_{\text{L}}(x_{\text{V,L}}^{\max}) - V_{\text{L}}(x_{\text{V,L}}^{\max} - \Delta T) \sim -\frac{V_{\text{L}}''(x_{\text{V,L}}^{\max})}{2} (\Delta T)^2.$$

Hence, the decrease of the firm revenue from a light user due to the changes in F and T is $O((\Delta T)^2)$.

The heavy users will still sign up to the plan, since their utility is positive. In addition, as in the proof of Lemma 5, the change in T and F does not affect the number of minutes they use. Therefore, the firm revenue from overage usage by the heavy users will increase by $p\Delta T$. Since the firm revenue decrease by $O((\Delta T)^2)$ and increase by $O(\Delta T)$, for ΔT sufficiently small the net firm revenue will increase. See also Section 4.2 for an example.

The choice between targeting only the heavy users versus targeting all users depends on the firm revenue under each strategy. Since the revenues under the (suboptimal) plan (21) are at least $V_{\text{L}}^{\max}$, the firm should target all users when $\gamma_{\text{H}} V_{\text{H}}^{\max} < V_{\text{L}}^{\max}$. When $\gamma_{\text{H}} V_{\text{H}}^{\max} \gg V_{\text{L}}^{\max}$, however, the firm should target the heavy users exclusively.

4.2. Parametric Example

Consider a market that consists of n_{H} heavy users and n_{L} light users with valuations

$$V_{\text{H}}(x) := \alpha_1 x - \alpha_2 x^2, \quad V_{\text{L}}(x) := \lambda \alpha_1 x - \alpha_2 x^2, \tag{23}$$

respectively, where $0 < \lambda < 1$ captures the reduction in light users' valuation, compared to heavy users valuation.

1. If the firm focuses on the heavy users (Lemma 4) then, as in Section 3.1,

$$\Pi_{\mathrm{H-only}} = \gamma_{\mathrm{H}} V_{\mathrm{H}}^{\max} = \gamma_{\mathrm{H}} \frac{\alpha_1^2}{4\alpha_2}. \tag{24}$$

2. If the firm focuses on the light users (Lemma 5) then, as in Section 3.1,

$$F_{\mathrm{L}}^{\mathrm{opt}} = V_{\mathrm{L}}^{\max} = \frac{\lambda^2\alpha_1^2}{4\alpha_2}, \quad T_{\mathrm{L}}^{\mathrm{opt}} = x_{\mathrm{L},V}^{\max} = \frac{\lambda\alpha_1}{2\alpha_2}. \tag{25}$$

The overage price is computed from

$$p_{\mathrm{L}}^{\mathrm{opt}} \stackrel{(21)}{=} \arg\max_{p\geqslant 0} \Pi_{\mathrm{H}}(p, T_{\mathrm{L}}^{\mathrm{opt}}, F_{\mathrm{L}}^{\mathrm{opt}}) \stackrel{(6)}{=} \arg\max_{p>0} \{p(x_{\mathrm{U},H}^{\mathrm{opt}} - T_{\mathrm{L}}^{\mathrm{opt}})\}. \tag{26}$$

To proceed, we need to compute $x_{\mathrm{U},H}^{\mathrm{opt}}$, the number of minutes that heavy users consume when they exceed T. For simplicity, we assume that psychological costs are negligible. Then by (6) and (23), $U_{\mathrm{H}}'(x) = V_{\mathrm{H}}'(x) - p = \alpha_1 - 2\alpha_2 x - p$ and $U_{\mathrm{H}}''(x) = -2\alpha_2 < 0$. Therefore,

$$x_{\mathrm{U},H}^{\mathrm{opt}} = \frac{\alpha_1 - p}{2\alpha_2}. \tag{27}$$

Substituting (25) and (27) in (26) yields

$$p_{\mathrm{L}}^{\mathrm{opt}} = \frac{\alpha_1(1-\lambda)}{2}. \tag{28}$$

Based on (22), (25), (27), and (28), the optimal revenue per consumer is

$$\begin{aligned}\Pi_{\text{L-mainly}} &= V_{\mathrm{L}}^{\max} + \gamma_{\mathrm{H}} p_{\mathrm{L}}^{\mathrm{opt}} (x_{\mathrm{U},H}^{\mathrm{opt}} - T_{\mathrm{L}}^{\mathrm{opt}}) \\ &= \frac{\lambda^2\alpha_1^2}{4\alpha_2} + \gamma_{\mathrm{H}} \frac{\alpha_1(1-\lambda)}{2}\left(\frac{\alpha_1 + \lambda\alpha_1}{4\alpha_2} - \frac{\lambda\alpha_1}{2\alpha_2}\right) \\ &= \frac{\alpha_1^2}{4\alpha_2}\left(\lambda^2 + \gamma_{\mathrm{H}}\frac{(1-\lambda)^2}{2}\right).\end{aligned} \tag{29}$$

If λ is close to 1, light and heavy users are almost identical. Therefore, the optimal strategy is to offer a plan that targets both segments. If λ is close to 0, heavy users are much more valuable to the firm. Hence, the firm should offer a plan that targets only heavy users. To find the threshold value of λ at which the optimal strategy changes, let λ^* be such that $\Pi_{\text{L-mainly}} = \Pi_{\mathrm{H-only}}$. By (24) and (29), $\Pi_{\text{L-mainly}} = \Pi_{\mathrm{H-only}}(\lambda^2/\gamma_{\mathrm{H}} + (1-\lambda)^2/2)$. Therefore,

$$\lambda^* = \frac{\gamma_{\mathrm{H}} + \sqrt{2\gamma_{\mathrm{H}}^2 + 2\gamma_{\mathrm{H}}}}{2+\gamma_{\mathrm{H}}}. \tag{30}$$

Consequently,

1. If $\lambda < \lambda^*$, $\Pi_{\text{L-mainly}} < \Pi_{\mathrm{H-only}}$ and so the firm is better off targeting only the heavy users' segment.

2. If $\lambda > \lambda^*$, $\Pi_{\text{L-mainly}} > \Pi_{\mathrm{H-only}}$ and so the firm is better off targeting mainly the light consumers' segment, i.e., selling to both segments while extracting all profits from the light users' segment.

As noted, allowing light users to talk as much as they want is always a suboptimal strategy. We now compute the optimal plan when the firm targets both light and heavy users. By Proposition 3, the optimal plan is attained for some $T < x_{\mathrm{V,L}}^{\max}$. Since it is always better to extract money from consumers using the fixed fee, the firm should set $F = V_{\mathrm{L}}(T)$. In this case, light users pay $V_{\mathrm{L}}(T)$ and heavy users pay $V_{\mathrm{L}}(T) + p(x_{\mathrm{U},H}^{\mathrm{opt}}(p) - T)$, where $x_{\mathrm{U},H}^{\mathrm{opt}}$ is given by (27). Therefore, the firm revenue is

$$\Pi_{\mathrm{L}+H} = V_{\mathrm{L}}(T) + \gamma_{\mathrm{H}} p(x_{\mathrm{U},H}^{\mathrm{opt}} - T).$$

To compute the optimal p and T, we differentiate $\Pi_{\mathrm{L}+H}$ with respect to p and T. This yields

$$\frac{\partial \Pi_{\mathrm{L}+H}}{\partial T} = V_{\mathrm{L}}'(T) - \gamma_{\mathrm{H}} p = 0,$$
$$\frac{\partial \Pi_{\mathrm{L}+H}}{\partial p} = (x_{\mathrm{U},H}^{\mathrm{opt}}(p) - T) + p\frac{d}{dp} x_{\mathrm{U},H}^{\mathrm{opt}}(p) = 0.$$

Substituting (23) and (27) yields

$$p = \frac{V_{\mathrm{L}}'(T)}{\gamma_{\mathrm{H}}} = \frac{1}{\gamma_{\mathrm{H}}}(\lambda\alpha_1 - 2\alpha_2 T),$$
$$T = x_{\mathrm{U},H}^{\mathrm{opt}} + p\frac{d}{dp} x_{\mathrm{U},H}^{\mathrm{opt}} = \frac{\alpha_1 - 2p}{2\alpha_2}.$$

The solution of these linear equations is

$$T^{\mathrm{opt}} = \frac{\lambda\alpha_1}{2\alpha_2}\left(1 - \frac{\gamma_{\mathrm{H}}}{2-\gamma_{\mathrm{H}}}\frac{1-\lambda}{\lambda}\right), \quad p^{\mathrm{opt}} = \frac{1-\lambda}{2-\gamma_{\mathrm{H}}}\alpha_1. \tag{31}$$

Note that

$$T^{\mathrm{opt}} = T_{\mathrm{L}}^{\mathrm{opt}}\left(1 - \frac{\gamma_{\mathrm{H}}}{2-\gamma_{\mathrm{H}}}\frac{1-\lambda}{\lambda}\right) < x_{\mathrm{V,L}}^{\max},$$
$$p^{\mathrm{opt}} = p_{\mathrm{L}}^{\mathrm{opt}} \frac{2}{2-\gamma_{\mathrm{H}}} > p_{\mathrm{L}}^{\mathrm{opt}}.$$

Thus, as predicted in Proposition 3, the optimal plan that targets both heavy and light users satisfies $T^{\mathrm{opt}} < x_{\mathrm{V,L}}^{\max}$. The decrease in the firm revenues from the fixed fee is offset by the increase in the overage price, since $p^{\mathrm{opt}} > p_{\mathrm{L}}^{\mathrm{opt}}$. Finally, some additional manipulations show that the optimal revenue per consumer is

$$\begin{aligned}\Pi_{\mathrm{L}+H} &= V_{\mathrm{L}}(T^{\mathrm{opt}}) + \gamma_{\mathrm{H}} p^{\mathrm{opt}}(x_{\mathrm{U},H}^{\mathrm{opt}} - T^{\mathrm{opt}}) \\ &= \frac{\alpha_1^2}{4\alpha_2}\frac{2\lambda^2 - 2\lambda\gamma_{\mathrm{H}} + \gamma_{\mathrm{H}}}{2-\gamma_{\mathrm{H}}} \\ &= \Pi_{\mathrm{H-only}}\frac{2\lambda^2 - 2\lambda\gamma_{\mathrm{H}} + \gamma_{\mathrm{H}}}{\gamma_{\mathrm{H}}(2-\gamma_{\mathrm{H}})}.\end{aligned}$$

We use the values of α_1 and α_2 from Section 3.1. In addition, we use $\lambda = 0.51$, $n_H = 50{,}000$ and $n_L = 125{,}000$. Thus, $\gamma_H = 50/175 \approx 0.286$ and $\lambda^* = 0.5$, see (30). Since $\lambda > \lambda^*$, the firm is better off targeting both light and heavy consumers.

Table 1 presents the two potential policies that target all consumers. Under policy I which was analyzed in Lemma 5, the firm extracts all the surplus from the light users setting the allowance to be exactly the number of minutes they wish to talk (228 minutes).

Table 1. Comparison of Two Policies That Target Both Light and Heavy Users

	Policy I	Policy II
Fixed fee (F)	\$22	\$21
Usage allowance (T)	228 minutes	191 minutes
Overage price (p)	9 cents/minute	11 cents/minute
Usage of light consumers ($x^{\text{opt}}_{U,L}$)	228 minutes	191 minutes
Usage of heavy consumers ($x^{\text{opt}}_{U,H}$)	337 minutes	319 minutes
Firm revenue per consumer	$\Pi_{\text{L-mainly}} = \$24.3$	$\Pi_{L+H} = \$24.8$

Therefore, the light users use 228 minutes and do not pay any overage fee. The firm sets a fixed fee of \$22, which is equal to the valuation of the light consumers when talking 228 minutes. To maximize the revenue from the heavy users, the firm sets an overage price of 9 cents per-minute. The heavy users use 337 minutes, out of which 228 are free and 337 − 228 = 109 are being charged for. Hence, they pay an overage fee of $p(x-T) = 0.09 \cdot 109 = \$9.81$. Overall, the firm revenue per consumer is \$24.3.

Policy II is the optimal policy that targets all consumers, which was calculated earlier in this subsection. Thus, p and T are given by (31). Under this policy the firm offers less free minutes ($T = 191$). The profits from the light users are lower, since they now use 191 minutes, and so their valuation is lower. Since they still do not pay any overage fee, the fixed fee reduces to $F = V_L(T) = \$21$. While the heavy users also use less minutes than in policy I (319 instead of 337), they pay for more minutes, since 319 − 191 = 128. In addition, they pay 2 cents per minute more for exceeding their monthly allowance. Overall, their overage fee increases dramatically to $p(x-T) = 0.11 \cdot 128 = \$14.08$. Overall, the firm revenue per consumer is \$24.8. The difference between the profit under the two policies is close to 2%. While it might not look that large, this 2% are net addition to the firm profit, since they do not increase the firm costs.

4.3. Optimal Two Plans

The firm can try to further increase its revenues by offering two three-part tariff plans (p_1, T_1, F_1) and (p_2, T_2, F_2) that target the light and heavy consumers, respectively. Since consumers choose the plan that maximizes their utility, the heavy consumers choose the plan

$$(p_{\text{H}}, T_{\text{H}}, F_{\text{H}}) := \begin{cases} (p_1, T_1, F_1), & \text{if } U^{\text{opt}}_{\text{H}}(p_1, T_1, F_1) \\ & \quad > \max\{U^{\text{opt}}_{\text{H}}(p_2, T_2, F_2), 0\}, \\ (p_2, T_2, F_2), & \text{if } U^{\text{opt}}_{\text{H}}(p_2, T_2, F_2) \\ & \quad > \max\{U^{\text{opt}}_{\text{H}}(p_1, T_1, F_1), 0\}, \\ \text{do not sign up}, & \text{otherwise}, \end{cases} \tag{32}$$

where $U^{\text{opt}}_{\text{H}}$ is defined by (8) with $U = U_{\text{H}}$. Similarly, the light consumers choose the plan $(p_{\text{L}}, T_{\text{L}}, F_{\text{L}})$. The firm revenues from heavy and light users are $n_{\text{H}}\Pi^{\text{opt}}_{\text{H}}(p_{\text{H}}, T_{\text{H}}, F_{\text{H}})$ and $n_{\text{L}}\Pi^{\text{opt}}_{\text{L}}(p_{\text{L}}, T_{\text{L}}, F_{\text{L}})$, respectively. Hence, the firm optimization problem reads

$$\{(p^{\text{opt}}_1, T^{\text{opt}}_1, F^{\text{opt}}_1), (p^{\text{opt}}_2, T^{\text{opt}}_2, F^{\text{opt}}_2)\} = \arg\max \Pi_{\text{two plans}}(p_1, T_1, F_1, p_2, T_2, F_2),$$

where $\Pi_{\text{two plans}} = \gamma_{\text{H}}\Pi^{\text{opt}}_{\text{H}}(p_{\text{H}}, T_{\text{H}}, F_{\text{H}}) + (1-\gamma_{\text{H}})\Pi^{\text{opt}}_{\text{L}}(p_{\text{L}}, T_{\text{L}}, F_{\text{L}})$ is the average firm revenue per consumer.

Ideally, the firm would like to extract the maximal revenue from all consumers, i.e., $\gamma_{\text{H}} V^{\max}_{\text{H}}$ from the heavy consumers and $(1-\gamma_{\text{H}})V^{\max}_{\text{L}}$ from the light ones. In Lemma 3 we showed that this is not possible with a single plan. Whether this is possible with two plans depends on the valuation of the heavy users at the optimal usage level of the light users:

Proposition 4. *Two three-part tariff plans can extract the maximal revenues from both light and heavy users if and only if $V_{\text{H}}(x^{\max}_{\text{V,L}}) \leqslant V_{\text{L}}(x^{\max}_{\text{V,L}})$, i.e., if the heavy users have a negative utility when joining the optimal plan of the light users. In other words, if $V_{\text{H}}(x^{\max}_{\text{V,L}}) > V_{\text{L}}(x^{\max}_{\text{V,L}})$, then for any two plans (p_1, T_1, F_1) and (p_2, T_2, F_2), the average firm revenue per consumer satisfies*

$$\Pi_{\text{two plans}} < \gamma_{\text{H}} V^{\max}_{\text{H}} + (1-\gamma_{\text{H}})V^{\max}_{\text{L}}.$$

Proof. See web appendix.

In general, one would expect that $V_{\text{H}}(x^{\max}_{\text{V,L}}) > V_{\text{L}}(x^{\max}_{\text{V,L}})$. This, however, is not always the case. For example, a residential light user might value a few megabites of Internet, while a heavy user might have no value for the Internet unless it can be used for business.

In Lemma 4 we saw that if the firm insists on maximizing the revenue from the heavy users, the light consumers will not sign up to this plan. If the firm adds a second plan but makes sure that it would be unattractive to the heavy users, the light users will not sign up to the second plan if and only if the valuation of the heavy users is always larger than that of the light ones:

Lemma 6. *There are no two plans (p_1, T_1, F_1) and (p_2, T_2, F_2) that extract the maximal revenue from the heavy consumers and also extract some revenues from the light consumers, if and only if*

$$V_{\text{L}}(x) < V_{\text{H}}(x), \quad x \geqslant 0. \tag{33}$$

Proof. See web appendix.

Thus, if the firm wants to attract the light users, it has to give up some of the potential revenues from the heavy ones. We note that Proposition 4 and Lemma 6 remain valid if we increase the number of plans. For example, assume that there are three segments of consumers: light, medium, and heavy. Then with three three-part tariff plans, the firm can extract the maximal revenues from the light, medium, and heavy users, if and only if the medium users have a negative utility when joining the optimal plan of the light users, and the heavy users have a negative utility when joining the optimal plans of the light users or of the medium users.

In Lemma 5 we saw that if the firm offers a single plan that extracts the maximal revenue from the light consumers, it can increase its revenues by maximizing the overage charges from the heavy consumers with an optimal choice of p and T. In that case, the firm profit was denoted by $\Pi_{\text{L-mainly}}$. We now show that the firm can further increase its profits by adding a second plan for the heavy users:

Lemma 7. *Let $V''_{\mathrm{H}}(x) < 0$. Then out of all the two plans (p_1, T_1, F_1) and (p_2, T_2, F_2) that maximize the revenues from the light users, the ones that maximize the overall profits are*

$$F_1 = V_{\mathrm{L}}^{\max}, \quad T_1 = x_{\mathrm{V,L}}^{\max}, \quad p_1 \geqslant V'_{\mathrm{H}}(x_{\mathrm{V,L}}^{\max}) \tag{34a}$$

for the light users, and

$$F_2 = V_{\mathrm{H}}^{\max} - (V_{\mathrm{H}}(x_{\mathrm{V,L}}^{\max}) - V_{\mathrm{L}}^{\max}), \quad T_2 \geqslant x_{\mathrm{V,H}}^{\max}, \quad p_2 \geqslant 0 \tag{34b}$$

for the heavy ones. In this case we have the following:

1. *Light users sign up to the plan and use $x_{\mathrm{V,L}}^{\max}$ minutes (i.e., as many minutes as they would in an unlimited plan). Their utility is zero.*
2. *Heavy users sign up to the plan and use $x_{\mathrm{V,H}}^{\max}$ minutes, (i.e., as many minutes as they would in an unlimited plan). They do not pay for overage usage. Their utility is positive.*
3. *The firm revenue per consumer is*

$$\Pi_{\text{two plans}} = V_{\mathrm{L}}^{\max} + \gamma_{\mathrm{H}}(V_{\mathrm{H}}^{\max} - V_{\mathrm{H}}(x_{\mathrm{V,L}}^{\max})).$$

In particular, it is higher than when the firm offers a single plan that extracts the maximal revenue from the light consumers, i.e., $\Pi_{\text{two plans}} > \Pi_{\text{L-mainly}}$, where $\Pi_{\text{L-mainly}}$ is given by (22).

Proof. See web appendix.

While adding a second plan for the heavy users increases the firm revenue, maximizing the revenue from the light users remains a suboptimal strategy:

Lemma 8. *Any optimal two plans $(p_1^{\mathrm{opt}}, T_1^{\mathrm{opt}}, F_1^{\mathrm{opt}})$ and $(p_2^{\mathrm{opt}}, T_2^{\mathrm{opt}}, F_2^{\mathrm{opt}})$ that target the light and heavy users, respectively, satisfy $F_1^{\mathrm{opt}} < V_{\mathrm{L}}^{\max}$ and $T_1^{\mathrm{opt}} < x_{\mathrm{V,L}}^{\max}$. Hence, the two plans given by (34) cannot be optimal.*

Indeed, if we set $T_1^{\mathrm{opt}} = x_{\mathrm{V,L}}^{\max} - \Delta T$ where $0 < \Delta T \ll 1$, the revenue loss from the light users is quadratic in ΔT, since their utility is maximized at $x_{\mathrm{V,L}}^{\max}$. The additional revenue gain from the heavy users, however, is linear in ΔT. Therefore, the net revenue increases as T decreases from $x_{\mathrm{V,L}}^{\max}$. See Appendix F for further details.

5. Stochastic Demand

In practice, consumers cannot predict exactly how many minutes they will use. This is especially true in the United States where consumers pay for incoming calls, which are harder to predict and control. Therefore, when a consumer plans to talk x minutes, he ends up talking X_x minutes, where X_x is a random variable. The randomness of X_x can be additive (i.e., $X_x = x + Z_1$), multiplicative (i.e., $X_x = x(1+Z_2)$) or both (i.e., $X_x = x(1+Z_1)+Z_2$), where Z_1 and Z_2 are random variables. To allow for all of these possibilities, we assume that for any x, X_x is a random variable that attains its value in $[0, M(x)]$ with probability 1, where $0 \leqslant M(x) < \infty$.[1] We also denote the density distribution of X_x by g_x. We assume that both the consumer and the firm know the distribution of X_x.

The expected firm revenue where the consumer plans to talk x minutes is

$$\bar{\pi}(x,p,T,F) := \mathbb{E}[\pi(X_x,p,T,F)] = \int_0^{M(x)} \pi(y,p,T,F)\, g_x(y)\, dy,$$

where π is defined by (3). Therefore,

$$\bar{\pi}(x,p,T,F) = \begin{cases} F, & \text{if } M(x) \leqslant T, \\ F + p\int_T^{M(x)} (y-T) g_x(y)\, dy, & \text{if } M(x) > T. \end{cases} \tag{35}$$

The consumer expected valuation where he plans to talk x minutes is

$$\bar{V}(x) := \mathbb{E}[V(X_x)] = \int_0^{M(x)} V(y) g_x(y)\, dy, \tag{36}$$

where V is defined by (1). We denote by $\bar{V}^{\max}$ the maximum of $\bar{V}(x)$ and by $x_{\bar{\mathrm{V}}}^{\max}$ the number of minutes that maximizes $\bar{V}(x)$, i.e.,

$$x_{\bar{\mathrm{V}}}^{\max} = \arg\max_{x \geqslant 0} \bar{V}(x), \quad \bar{V}^{\max} = \bar{V}(x_{\bar{\mathrm{V}}}^{\max}). \tag{37}$$

Thus, $x_{\bar{\mathrm{V}}}^{\max}$ is the number of minutes that a rational stochastic consumer plans to talk when he signs up to an unlimited plan ($T = \infty$).

The consumer expected psychological cost is

$$\bar{S}(x,p,T) := \mathbb{E}[S(X_x,p,T)] = \begin{cases} 0, & \text{if } M(x) \leqslant T, \\ \int_T^{M(x)} S(y,p,T)\, g_x(y)\, dy, & \text{if } M(x) > T, \end{cases}$$

where S is defined by (4). The consumer expected utility when he plans to talk x minutes is

$$\bar{U}(x,p,T,F) := \mathbb{E}[U(X_x,p,T,F)] \overset{(5)}{=} \bar{V}(x) - \bar{\pi}(x,p,T,F) - \bar{S}(x,p,T). \tag{38}$$

For a given plan (p,T,F), a rational consumer plans to talk $x_{\bar{U}}^{\text{opt}}$ minutes, where

$$x_{\bar{U}}^{\text{opt}}(p,T,F) := \arg\max_{x\geqslant 0} \bar{U}(x,p,T,F). \tag{39}$$

The consumer signs up to the plan if his maximal expected utility is nonnegative, i.e.,

$$\bar{U}^{\text{opt}}(p,T,F) := \bar{U}(x_{\bar{U}}^{\text{opt}}(p,T,F),p,T,F) \geqslant 0. \tag{40}$$

Otherwise, he does not sign up to the plan. Therefore, the firm expected revenue is

$$\bar{\Pi}(p,T,F) := \begin{cases} \bar{\pi}(x_{\bar{U}}^{\text{opt}}(p,T,F),p,T,F), & \text{if } \bar{U}^{\text{opt}}(p,T,F)\geqslant 0, \\ 0, & \text{otherwise.} \end{cases} \tag{41}$$

In the case of constant firm costs, the firm optimization problem reads

$$(p^{\text{opt}},T^{\text{opt}},F^{\text{opt}}) = \arg\max_{p,T,F\geqslant 0} \bar{\Pi}(p,T,F).$$

The following proposition characterizes the optimal three-part tariff when the demand is stochastic:

Proposition 5. *The optimal firm plan when consumers are homogeneous, firm costs are constant, and consumers have stochastic demand is*

$$F^{\text{opt}} = \bar{V}^{\max}, \quad T^{\text{opt}} \geqslant M(x_{\bar{V}}^{\max}), \quad p^{\text{opt}} \geqslant 0, \tag{42}$$

where $\bar{V}^{\max}$ and $x_{\bar{V}}^{\max}$ are defined in (37). In addition,
1. *the consumer plans to talk $x_{\bar{V}}^{\max}$ minutes,*
2. *the consumer expected utility is 0,*
3. *the expected firm revenue is $\bar{V}^{\max}$.*

Proof. Because the optimization problem is nonsmooth, and because we do not assume explicit forms for V, S, and X_x, it cannot be solved using the first-order condition approach. Therefore, we solve the optimization problem by obtaining an upper bound on the firm revenue under any three-part plan, see Equation (43), and then showing that plan (42) attains this bound. We first show that the expected firm revenue is bounded by the maximal expected consumer valuation, i.e.,

$$\bar{\Pi}(p,T,F) \leqslant \bar{V}^{\max}. \tag{43}$$

Indeed, for any firm plan (p,T,F) such that the maximal utility of the consumer $\bar{U}^{\text{opt}}(p,T,F)$ is negative, the consumer does not sign up to the plan. Therefore, the firm's revenue is zero. In particular, $\bar{\Pi}(p,T,F) = 0 < \bar{V}^{\max}$.

If $\bar{U}^{\text{opt}}(p,T,F) \geqslant 0$, the consumer signs up to the plan. Hence, by (4), (37), and (38),

$$\begin{aligned} 0 \leqslant \bar{U}^{\text{opt}}(p,T,F) &= \bar{V}(x_{\bar{U}}^{\text{opt}}) - \bar{\Pi}(p,T,F) - \bar{S}(x_{\bar{U}}^{\text{opt}}) \\ &< \bar{V}^{\max} - \bar{\Pi}(p,T,F). \end{aligned} \tag{44}$$

We now show that if the firm plan satisfies (42), then $\bar{\Pi}(p^{\text{opt}},T^{\text{opt}},F^{\text{opt}}) = \bar{V}^{\max}$. Indeed, for any $T^{\text{opt}} \geqslant M(x_{\bar{V}}^{\max})$, if a consumer signs up to the plan, he will plan to use $x_{\bar{V}}^{\max}$ minutes. By (6), his utility is $\bar{U}(x_{\bar{V}}^{\max}, p^{\text{opt}},T^{\text{opt}},F^{\text{opt}}) = \bar{V}^{\max} - F^{\text{opt}} = 0$. Therefore, he chooses to sign up to the plan. In this case, the firm revenue is $\bar{\pi}(x_{\bar{V}}^{\max},p^{\text{opt}},T^{\text{opt}},F^{\text{opt}}) = F^{\text{opt}} = \bar{V}^{\max}$.

Thus, as in the deterministic case (Proposition 1), the solution of this nonsmooth optimization problem is to let consumers talk as much as they want and then extract all their utility through the fixed fee. Similarly, the result of Proposition 2 for homogeneous consumers with variable firm costs extends almost "as is" to the case of stochastic demand (see Proposition 7 in the web appendix). The results for heterogeneous consumers also extend to the stochastic case almost "as is." In that case, we assume that for any x, $X_{x,\text{H}}$ and $X_{x,\text{L}}$ are random variables that attain their values with probability 1 in $[0,M_{\text{H}}(x)]$ and $[0,M_{\text{L}}(x)]$, respectively. For example, the following lemma shows the extension of Lemma 7 to the case of stochastic demand:

Lemma 9. *Let $\bar{V}_{\text{H}}''(x) < 0$. Then out of all the two plans (p_1,T_1,F_1) and (p_2,T_2,F_2) which maximize the revenues from the stochastic light users, the ones that maximize the overall profits are*

$$F_1 = \bar{V}_{\text{L}}^{\max}, \quad T_1 = M_L(x_{\bar{V},\text{L}}^{\max}), \quad p_1 \geqslant \bar{V}_{\text{H}}'(x_{\bar{V},\text{L}}^{\max}) \tag{45a}$$

for the stochastic light users, and

$$F_2 = \bar{V}_{\text{H}}^{\max} - (\bar{V}_{\text{H}}(x_{\bar{V},\text{L}}^{\max}) - \bar{V}_{\text{L}}^{\max}), \quad T_2 \geqslant M_H(x_{\bar{V},\text{H}}^{\max}), \quad p_2 \geqslant 0 \tag{45b}$$

for the stochastic heavy users.

The proof is identical to the deterministic case, with the obvious changes $V \to \bar{V}$, $x_{\text{V,L}}^{\max} \to x_{\bar{V},\text{L}}^{\max}$, $U_{\text{H}}^{\text{opt}} \to \bar{U}_{\text{H}}^{\text{opt}}$, etc.

Similarly, the extension of Lemma 8 to the stochastic case reads as follows:

Lemma 10. *Any optimal two plans $(p_1^{\text{opt}},T_1^{\text{opt}},F_1^{\text{opt}})$ and $(p_2^{\text{opt}},T_2^{\text{opt}},F_2^{\text{opt}})$ that target the stochastic light and heavy users, respectively, satisfy $F_1^{\text{opt}} < \bar{V}_{\text{L}}^{\max}$ and $T_1^{\text{opt}} < x_{\bar{V},\text{L}}^{\max}$. Hence, the two plans given by (45) cannot be optimal.*

5.1. Stochastic Influence

In this section we discuss how the firm's optimal revenue is affected by the stochastic demand X_x. We first compare consumers with stochastic and deterministic demand.

Lemma 11. *The maximal expected valuation of consumers with stochastic demand is always less than that of consumers with deterministic demand ($\bar{V}^{\max} < V^{\max}$). Therefore, the optimal firm's revenue from consumers with deterministic demand is greater than from consumers with stochastic demand ($\bar{\Pi} < \Pi$).*

In general, as the variance of the consumer's monthly usage increases, his expected utility decreases. Therefore, the firm's optimal revenue also decreases. We next prove this result for the case of additive randomness.

Proposition 6. *Suppose that $V'' < 0$, let the stochastic demand be given by $X_x^w = x + wZ$, where Z is a bounded random variable, and denote by $\bar{\Pi}(w)$ the corresponding optimal firm revenue. Then $\bar{\Pi}(w)$ decreases as w increases.*

5.2. Parametric Example

We extend the parametric example from Section 3.1 to the case of homogeneous consumers with stochastic demand. Let $X_x = x + Z$, where Z is a bounded random variable with zero mean and a variance of σ^2. By Proposition 5, the optimal firm plan is

$$F^{\mathrm{opt}} = \bar{V}(x_{\bar{\mathrm{V}}}^{\max}), \quad T^{\mathrm{opt}} \geqslant x_{\bar{\mathrm{V}}}^{\max} + \max Z, \quad p^{\mathrm{opt}} \geqslant 0,$$

and the maximal expected firm revenue is

$$\bar{\Pi}(p^{\mathrm{opt}}, T^{\mathrm{opt}}, F^{\mathrm{opt}}) = F^{\mathrm{opt}} = \bar{V}(x_{\bar{\mathrm{V}}}^{\max}).$$

Since $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] = \sigma^2$, the expected consumer valuation is, see (15),

$$\begin{aligned}\bar{V}(x) &= \mathbb{E}[V(x+Z)]\\ &= \alpha_1(x + \mathbb{E}[Z]) - \alpha_2(x^2 + 2x\mathbb{E}[Z] + \mathbb{E}[Z^2])\\ &= V(x) - \alpha_2\sigma^2. \end{aligned} \tag{46}$$

Since $\alpha_2\sigma^2$ does not depend on x, then $\bar{x}_{\mathrm{V}}^{\max} = x_{\mathrm{V}}^{\max}$, and so

$$\bar{V}(\bar{x}_{\mathrm{V}}^{\max}) = V(x_{\mathrm{V}}^{\max}) - \alpha_2\sigma^2 = \$(83 - \alpha_2\sigma^2).$$

Therefore,

$$F = \$(83 - \alpha_2\sigma^2), \quad T \geqslant (447 + \max Z)\ \text{minutes}, \quad p \geqslant 0.$$

and

$$\bar{\Pi}(p, T, F) = \$(83 - \alpha_2\sigma^2).$$

In particular, the firm revenue decreases with σ^2, in agreement with Proposition 6.

6. Conclusions

Services play an ever larger role in the modern economy. Nonlinear pricing plans are ubiquitous in the service industry, primarily as three-part tariff plans. Nevertheless, prior research on three-part tariffs was limited, because the standard mathematical approach (which is based on first-order conditions) is not suitable for this nonsmooth nested optimization problem. To overcome this obstacle, we adopted an alternative approach that is based on finding tight bounds. This novel approach allows us to explicitly calculate the optimal three-part tariff contract under general conditions. Our approach may be suitable to other optimization problems in marketing and management, since many of these problems are inherently nonsmooth (because, e.g., of the different response of consumers to "gains" and "losses," or the existence of a threshold price).

When consumers are homogeneous and the firm costs are constant, the optimal three-part tariff plan is to allow consumers to use as many minutes as they want and extract all their surplus through the fixed fee. In that case, the monthly allowance only needs to be "sufficiently high," and the value of the per-minute overage price can be arbitrary. In practice, however, cellular firms often offer plans with a limited number of minutes, and consumers often pay for exceeding their monthly allowance. Our analysis reveals that firms may adopt this strategy when its costs depend on the usage level and/or when consumers are heterogeneous. In the latter case, the firm should use all three levers of the tariff plan (fixed fees, unit allowances, and overage fees) to discriminate among consumer segments.

When the market consists of two segments of light and heavy users, then depending on the relative size of each segment and its attractiveness in terms of potential revenue, the firm may either serve the heavy users exclusively, or serve both segments. In the latter case, one could expect that the optimal firm policy would be to extract the maximal surplus from the light users (by allowing them to use as many minutes as they want), and then set the overage price to maximize the profits from the heavy users. This strategy, however, turns out to be always suboptimal. Rather, the optimal policy is to a lower monthly allowance, a lower monthly fixed fee, and a higher overage price. Thus, the reduction of the monthly allowance reduces the revenues from the light users, since they are willing to pay a lower fixed fee. This reduction is more than compensated by the increase in the overage charges paid by the heavy users, who pay for more minutes and pay more for each minute. Interestingly, under both policies, the light users subsidize the heavy users, in the sense that the firm extracts all of the surplus from the light users, while leaving a positive surplus to the heavy users.

In closing, we acknowledge that our analysis considers a monopoly service provider who sells to a market that consists of at most two segments of consumers that are risk neutral. The focus of this study is on computing and characterizing the optimal three-part-tariff contract under different considerations (variable firm's costs, heterogeneous or homogeneous consumers, deterministic or stochastic demand, one or two three-part tariff plans). There are several important issues that remain open. The most obvious one is to allow for *competition*. Another interesting research avenue to consider is more *general multipart tariff plans*. For example, water and electricity are often priced using *four-part tariff plans* in which consumers pay a fixed monthly fee F, a price p_1 for each unit consumed below a threshold T, and a (higher or lower) price p_2 for each unit above T. Briefly, whenever the optimal three-part tariff plan in our model extracts maximum utility from consumers (e.g., in the homogeneous case with or without firm costs and in the heterogeneous case when the firm targets the heavy users), adding levers will, at best, match (and might reduce) the profit. Therefore, for example, an optimal four-part tariff plan for homogeneous consumers is one in which $p_1 = 0$. Furthermore, even when the optimal three-part tariff plan does not extract maximum utility from consumers, adding levers is not always profitable. For example, consider the optimal three-part tariff plan that targets light and heavy users; see Proposition 3. Charging price p_1 for each unit below T will not increase the firm's profit since the additional revenue $(p_1 T)$ must be offset by an identical reduction in the fixed fee. Adding levers can increase the firm's profit when there are more than two types of heterogeneous consumers.

Another assumption that can be challenged concerns the psychological costs. While we allowed for a general psychological cost function associated with overage, we did not take into account the *psychological costs associated with leaving minutes on the table (underage)*. In the deterministic case, allowing for psychological underage costs has a limited effect on our results. Indeed, in most of our results, consumers use their allowance (see, e.g., Proposition 2 and Lemma 5). In such cases, allowing for underage costs does not change the results. Consumers may experience underage costs in cases such as Proposition 1, where the optimal firm strategy is $T^{\text{opt}} \geqslant x_{\text{V}}^{\max}$. In such cases, the effect of introducing underage costs is to change the optimal strategy to $T^{\text{opt}} = x_{\text{V}}^{\max}$. In the stochastic case, the situation is more subtle. Briefly, including underage costs will result in lower expected utility for a given plan, as consumers incur psychological costs if the realized consumption is below the plan's free minutes T. As a result, the firm will offer plans with a lower T.[2] We leave all these open questions for future research.

Finally, we acknowledge that our analysis suggests that in most cases, consumers do not (choose to) exceed their monthly allowance, which is inconsistent with evidence generated by some of the empirical literature (Lambrecht et al. 2007, Iyengar et al. 2007, and Grubb 2009) that consumers use more minutes than the number of minutes included in their monthly plan. For example, Grubb (2009) states in Figure 2 that this happens about 17% of the time. One reason for such inconsistency may be that not all consumers are strategic as we assume in our model and analysis. Relaxing this assumption may lead to results that will be more consistent with the empirical evidence. We also note that our analysis suggests that if the firm targets the low users, then strategic heavy users will exceed their monthly allowance (Lemma 5).

Endnotes

[1] The assumption that the demand shock is bounded follows from our assumption that the consumer has a finite budget (Section 2).

[2] The consumers' utility function has some commonality with the (producer/retailer) newsvendor problem. Under the newsvendor problem, a firm that has to produce (order) units and faces uncertain demand has to take into account the costs of selling less than the produced quantity (underage costs) or demand that exceeds the produced quantity (overage).

References

Ascarza E, Lambrecht A, Vilcassim N (2012) When talk is free: The effect of tariff structure on usage under two- and three-part tariffs. *J. Marketing Res.* 49(6):882–899.
Bagh A, Bhargava HK (2013) How to price discriminate when tariff size matters. *Marketing Sci.* 32(1):111–126.
Cachon GP, Feldman P (2011) Pricing services subject to congestion: Charge per-use fees or sell subscriptions? *Manufacturing Service Oper. Management* 13(2):244–260.
Desai P, Purohit D, Zhou B (2016) Allowing Consumers to Bundle Themselves: The Profitability of Family Plans. Working paper, Duke University.
Essegaier S, Gupta S, Zhang ZJ (2002) Pricing access services. *Marketing Sci.* 21(2):139–159.
Grubb MD (2009) Selling to overconfident consumers. *Amer. Econom. Rev.* 99:1770–1807.
Huang C-I (2008) Estimating demand for cellular phone service under nonlinear pricing. *Quant. Marketing Econom.* 6:371–413.
Iyengar R, Ansari A, Gupta S (2007) A model of consumer learning for service quality and usage. *J. Marketing Res.* XLIV:529–544.
Iyengar R, Jedidi K, Kohli R (2008) A conjoint approach to multipart pricing. *J. Marketing Res.* XLV:195–210.
Johnson JP, Myatt DP (2003) Multiproduct quality competition: Fighting brands and product line pruning. *Amer. Econom. Rev.* 93(3):748–774.
Kim Y, Telang R, Vogt WB, Krishnan R (2010) An empirical analysis of mobile voice service and SMS: A structural model. *Management Sci.* 56(2):234–252.
Koenigsberg O, Kohli R, Montoya R (2010) Package size decisions. *Management Sci.* 56(3):485–494.
Lambrecht A, Seim K, Skiera B (2007) Does uncertainty matter? Consumer behavior under three-part tariffs. *Marketing Sci.* 26(5): 698–710.
Mas-Colell A, Whinston MD, Green JR (1995) *Microeconomic Theory* (Oxford University Press, New York).
Moorthy SK (1984) Market segmentation, self-selection, and product line design. *Marketing Sci.* 3(4):288–305.
Mussa M, Rosen S (1978) Monopoly and product quality. *J. Econom. Theory* 18(2):301–317.
Villas-Boas MJ (2004) Communication strategies and product line design. *Marketing Sci.* 23(3):304–316.

Gadi Fibich is a professor of applied mathematics at Tel Aviv University. His research focuses on bringing applied math techniques to problems in marketing (nonlinear pricing, diffusion of new products) and economics (auction theory).

Roy Klein is a product manager in NICE LTD. He has a long experience in the high-tech industry and his research focuses on obtaining marketing insights using advanced mathematical techniques. He earned a a M.Sc. in Applied Mathematics from Tel-Aviv University.

Oded Koenigsberg is the London Business School Term Associate Professor of Marketing. His primary research areas include pricing and marketing-manufacturing interface.

Eitan Muller has a joint appointment at the Arison School of Business, the Interdisciplinary Center, Herzliya and the Stern School of Business, New York University. He has published extensively, mostly on innovation diffusion, with more than 16,000 citations in Google Scholar.