UNCERTAIN DEMAND, CONSUMER LOSS AVERSION, AND FLAT-RATE TARIFFS

Fabian Herweg
University of Munich

Konrad Mierendorff
University of Zurich

Abstract
We consider a model of firm pricing and consumer choice, where consumers are loss averse and uncertain about their future demand. Possibly, consumers in our model prefer a flat rate to a measured tariff, even though this choice does not minimize their expected billing amount—a behavior in line with ample empirical evidence. We solve for the profit-maximizing two-part tariff, which is a flat rate if (a) marginal costs are not too high, (b) loss aversion is intense, and (c) there are strong variations in demand. Moreover, we analyze the optimal nonlinear tariff. This tariff has a large flat part when a flat rate is optimal among the class of two-part tariffs. (JEL: D11, D43, L11)

1. Introduction

Flat-rate tariffs that offer unlimited usage for a fixed amount of money are common practice in many industries, e.g., for telephone services, Internet access, car rental, car leasing, DVD rental, amusement parks, and many others. The prevalence of flat rates is hard to reconcile with orthodox economic theory. In industries where marginal costs are non-negligible, a marginal payment of zero leads to an inefficiently high level of consumption. A flat rate can nevertheless be optimal if measuring the actual usage of a consumer is costly (Sundararajan, 2004). Flat-rate contracts, however, are also found in industries with positive marginal costs where measurement costs are almost

The editor in charge of this paper was Stefano DellaVigna.

Acknowledgments: In preparing this paper, we have greatly benefited from comments made by the editor, Stefano DellaVigna, four anonymous referees, David Freeman, Paul Heidhues, David Laibson, Kristóf Madarász, Daniel Müller, Takeshi Murooka, Martin Peitz, Matthew Rabin, Klaus M. Schmidt, Urs Schweizer, Igor Sloev, and Philipp Weinschenk. We also thank participants at the Nordic Conference on Behavioral and Experimental Economics at Oslo (2009), at the IMEBE at Bilbao (2010), at the SFB Meeting at Caputh (2010), at the WZB Conference Biases in Markets, Berlin (2010), as well as seminar participants at University of California—Berkeley, ESMT Berlin, the University of Zurich, and the University of Bonn. All remaining errors are our own responsibility. Herweg gratefully acknowledges UC Berkeley's hospitality while a part of this paper was written, and financial support from the German Academic Exchange Service (DAAD) as well as from the Deutsche Forschungsgemeinschaft through SFB/TR-15. Mierendorff gratefully acknowledges financial support from the European Research Council (ERC Advanced Investigator Grant, ESEI-249433).

E-mail: fabian.herweg@lrz.uni-muenchen.de (Herweg); konrad.mierendorff@econ.uzh.ch (Mierendorff)

zero, e.g., for rental cars or telephone services.[1] For rental cars, a typical contract has a fixed price per day that does not depend on the mileage.[2] The costs for the car rental company are clearly higher if the car is used more heavily, for instance due to a higher wear of the tires. On the other hand, it is not very costly to determine the mileage of a customer.[3,4] Given these observations, what reasons do firms have to offer flat-rate contracts? We provide a theoretical answer which lies outside standard consumer behavior.[5]

There is plenty of evidence that consumers often do not select the tariff option that minimizes their expenditures for observed consumption patterns. In particular, consumers often prefer a flat-rate tariff even though they would save money with a measured tariff. Train (1991) referred to this phenomenon as the "flat-rate bias."[6] Given that consumers are willing to pay a "flat-rate premium," it is unsurprising that flat rates are widely used.

The literature classifies three potential causes of the flat-rate bias: the taxi-meter effect, overestimation, and an insurance motive (Lambrecht and Skiera, 2006). The first effect is discussed in the literature on mental accounting. If a measured tariff makes the link between payment and consumption very salient, it reduces the consumer's pleasure from the service. Put differently, during a taxi ride consumers dislike observing the meter running. According to mental accounting theory, the taxi-meter effect can be avoided if payments are decoupled from consumption (Prelec and Loewenstein, 1998; Thaler, 1999). In the industries we consider, however, consumers do not directly pay for consumption, but receive a bill at the end of the billing period. It is therefore unclear whether payments and consumption are sufficiently tightly coupled to create a strong taxi-meter effect.

Second, consumers may overestimate their future consumption and thus overvalue a contract with unlimited usage. DellaVigna and Malmendier (2006) provide evidence that many customers of health clubs overpredict their future usage. As

1. Despite conventional wisdom, the marginal cost of a telephone call is not zero (Faulhaber and Hogendorn, 2000). Moreover, telephone companies pay access charges on a per minute basis for off-net calls.

2. For instance, in Germany, the car rental companies Avis, Europcar, and Hertz (three major enterprises) offer flat-rate contracts. Another common contractual form for rental cars is a three-part tariff: the contract includes a mileage allowance and a charge per mile thereafter.

3. Typically, car rental companies record the car's mileage in the final bill even for contracts with unlimited mileage. We thank an anonymous referee for pointing this out.

4. Other well-fitting examples are the flat-rate contracts for leasing cars newly introduced by Ford and Citroën in Germany. These contracts cover—next to the usual services—also repairing of wear parts for a fixed amount per month that does not depend on the mileage.

5. We do not claim that our explanation is unique. For instance, with standard consumer behavior, marginal prices below marginal costs can be explained by crossing demand curves (Ng and Weissner, 1974).

6. The flat-rate bias was first documented for US household demand for telephone services (Train et al., 1987; Hobson and Spady, 1988; Kling and van der Ploeg, 1990; Mitchell and Vogelsang, 1991). Nunes (2000) provides evidence outside the telecommunications industry. For an overview see Lambrecht and Skiera (2006). Miravete (2003) finds no evidence of a flat-rate bias for telephone services.

the authors carefully expose, this misprediction could be caused by naive quasi-hyperbolic discounting. DellaVigna and Malmendier (2004) show that with naive quasi-hyperbolic discounters, the unit price of the optimal two-part tariff is below marginal costs for investment goods (health club attendance). For leisure goods like rental cars, however, naive quasi-hyperbolic discounters underestimate future demand and the optimal unit price exceeds marginal costs.

Third, consumers may prefer flat rates because they would like to be insured against payment variations that arise with measured tariffs if future consumption is uncertain. Early articles investigating the flat-rate bias, like Train et al. (1989), point out that "customers do not choose tariffs with complete knowledge of their demand, but rather choose tariffs [...] on the basis of the insurance provided by the tariff in the face of uncertain consumption patterns (p. 63)."[7] Standard risk aversion is not sufficient to explain the insurance motive, since the variations in payments are usually small compared to a consumer's income (Clay et al., 1992; Miravete, 2002). If one presumes that consumers are "narrow bracketers," risk aversion can explain why consumers have a preference for flat-rate tariffs but it does not explain why firms *offer* such tariffs. If demand is price sensitive, a slightly positive unit price creates an incentive for the consumers to partly internalize the firm's cost. This increases efficiency without imposing an unpleasant risk on the consumers.

In order to model the insurance motive, we posit that consumers are expectation-based loss averse.[8] A loss-averse consumer is first-order risk averse, i.e., he dislikes even small deviations from his reference point. We consider a model where a monopolistic firm offers a two-part tariff to consumers. When deciding whether to accept the contract, a consumer is uncertain about his future demand. After accepting the contract, the consumer learns his demand type, chooses a quantity, and makes a payment according to the two-part tariff. We assume that a consumer's demand is always satiated for a finite quantity. For example, consider a consumer who decides today whether or not to sign a contract with a car rental agency for his holidays in a few weeks. How many miles he will drive depends on the weather. If the sun is always shining, the consumer uses the car only to drive to the nearby beach. But if the weather is bad, he takes longer sight-seeing trips.

At the consumption stage, the consumer compares his actual bill to his reference bill. The consumer is disappointed if the actual bill exceeds the reference bill and tries to avoid this loss by reducing his consumption. The reference bill is determined by the consumer's lagged rational expectations about his billing amount, which he forms before accepting the contract.[9] Following Kőszegi and Rabin (2006, 2007), we

7. Evidence that the flat-rate bias is—at least partially—driven by the insurance effect is provided by Kridel et al. (1993), Lambrecht and Skiera (2006), and Lambrecht et al. (2007).

8. The marketing literature mentions consumer loss aversion as a potential explanation for the flat-rate bias (Lambrecht and Skiera, 2006). Loss aversion requires the implicit assumption of narrow bracketing. For an overview of the empirical evidence on loss aversion see Camerer (2004) and the references therein.

9. Abeler et al. (2011), Post et al. (2008), Crawford and Meng (forthcoming), Gill and Prowse (forthcoming), and Ericson and Fuster (forthcoming) provide evidence that reference points are determined

assume that the reference point is the full distribution of possible billing amounts, and that the consumer's expectations are self-fulfilling, i.e., the demand function is a personal equilibrium. At the contracting stage, the consumer anticipates the losses he will feel with a measured tariff. This increases his willingness to pay for contracts that insure against variations in payments. Put differently, the consumer's preferences are biased in favor of flat rates.

We abstract from loss aversion in the "good dimension," i.e., the consumer does not experience a loss if his expected gross utility from consumption is higher than his actual utility from consumption. Intuitively, a consumer does not feel a loss if the weather is nice and he uses the rental car less often than expected. The consumer feels a loss, however, if the rental price depends on his mileage and he used the car more often than expected.[10] This assumption also guarantees monotonicity of the demand function with respect to the demand type, which simplifies the characterization of personal equilibria.

We consider a profit-maximizing monopolist who offers a two-part tariff to the consumers. There is symmetric information at the contracting stage, neither the consumers nor the monopolist know the realized demand types. The monopolist can extract all expected surplus arising from the contract via the lump-sum fee. Therefore, his objective when setting the unit price is to maximize total expected surplus including the expected losses felt by the consumers. The monopolist faces a trade-off between maximizing standard efficiency which would require a positive unit price, and minimizing expected losses, which can be achieved by setting a unit price of zero. Minimizing losses is more important than maximizing standard efficiency and thus a flat-rate contract is optimal, if (i) the marginal cost is small, (ii) the consumer is sufficiently loss averse, and (iii) demand is sufficiently uncertain. Intuitively, if marginal costs are low, the quantities demanded under a flat rate are close to the efficient levels, so that overconsumption is not very costly. Moreover, the insurance value of a flat rate is high if either the variation in demand or the degree of loss aversion is high. Flat rates arise only when the consumer is sufficiently uncertain about his future consumption. Demand uncertainty is necessary for the insurance effect, but is not needed for the competing explanations of the flat-rate bias, the taxi-meter effect and overestimation of demand.[11]

While we focus on loss aversion according to Kőszegi and Rabin (2006), we discuss different notions of loss aversion in an illustrative example in the next section.

---

by expectations. See Berseghyan et al. (2010) for a paper that does not find evidence of expectation-based loss aversion.

10. One could defend this assumption also on the ground that there is one point in time where the consumer receives his bill and compares it with his expectations, whereas the potential losses regarding the consumption of the good are distributed over the whole billing period and therefore are less salient.

11. Uncertainty may also play a role in the mental accounting explanation. Prelec and Loewenstein (1998) argue that "mental prepayment" allows a consumer to enjoy consumption without thinking about the associated payments. If consumption is uncertain, flat rates facilitate mental prepayment whereas measured tariffs do not allow the consumer to mentally pay in advance because the exact billing amount is unknown beforehand. This could create a flat-rate bias if demand is uncertain.

The example demonstrates that a flat-rate tariff can also be optimal under an alternative notion of loss aversion. The relationship between the optimality of a flat rate and uncertainty of demand, however, only arises if consumers are expectation-based loss averse.

In a first extension, we relax the restriction to two-part tariffs and solve for the optimal nonlinear tariff. When the marginal cost is low, the optimal tariff consists of a flat part for intermediate quantities. The size of the flat part increases if (i) the consumers' degree of loss aversion increases, (ii) variations in preferences and thus in demanded quantities increase, and (iii) the marginal cost decreases.

Moreover, focusing on two-part tariffs, we inspect the robustness of our findings. First, we consider heterogeneity in the degree of loss aversion. We derive conditions such that the monopolist can screen differently loss-averse consumers without costs. In this case, it can be optimal to offer a menu of tariffs that includes a flat rate *and* a measured tariff. Second, we show that the structure of the optimal two-part tariff does not depend on the degree of competition. Finally, we relax the assumption that consumers do not feel losses in the good dimension. Under certain assumptions, a flat-rate contract is optimal under the same conditions as in the baseline model.

Related Literature

This paper is related to a recent and growing literature investigating how rational firms respond to consumer biases. In a seminal contribution, DellaVigna and Malmendier (2004) consider a market with time-inconsistent consumers and solve for the two-part tariff offered in equilibrium. A perfectly competitive market for credit cards with quasi-hyperbolic discounters is analyzed by Heidhues and Kőszegi (2010a). Using a different notion of time-inconsistency, Eliaz and Spiegler (2006) solve for the optimal menu of tariffs for a monopolist who faces consumers that differ in their degree of sophistication. Esteban et al. (2007) also analyze the optimal nonlinear pricing scheme for a monopolist who sells to consumers with self-control problems. Instead of assuming time-inconsistency, they model self-control problems using the concept of Gul and Pesendorfer (2001).[12]

In our model with rational consumers, the monopolist offers a flat-rate tariff because consumers are willing to pay a premium in order to be insured against unexpectedly high bills. The flat-rate tariff is not offered to exploit a cognitive bias of the consumers. This is in contrast to several models with boundedly rational consumers where firms design tariffs to exploit consumers' biases. See for instance Grubb (2009) for a model where consumers underestimate fluctuations in their demand, and Eliaz and Spiegler (2008) for a model with consumers who have biased beliefs.[13]

12. Similar results are obtained by Esteban and Miyagawa (2006) for a perfectly competitive market.

13. Uthemann (2005) studies a similar model where firms screen consumers with respect to their priors.

Closer to our paper, Hahn et al. (2010) adopt the model of Kőszegi and Rabin (2006) and analyze the product line of a monopolist who faces loss-averse consumers. After observing the offered product line, but before knowing their own types, consumers form expectations about the products they will purchase. The main finding is that the optimal product line contains fewer products than predicted by standard versioning models.

Expectation-based loss aversion according to Kőszegi and Rabin (2006, 2007) is also applied in other contexts. Heidhues and Kőszegi (2010b) apply this concept to provide an explanation why regular prices are sticky but sales prices are variable. Heidhues and Kőszegi (2008) introduce consumer loss aversion into a model of horizontally differentiated firms. They show that in equilibrium, asymmetric competitors charge identical focal prices for differentiated products.[14] Considering an agency model, Herweg et al. (2010) provide an explanation for the frequent usage of lump-sum bonus contracts. A repeated moral hazard model with a loss-averse agent is analyzed by Macera (2009).

The paper proceeds as follows: Section 2 presents a simple example that illustrates the main findings. In Section 3, the baseline model with two-part tariffs is introduced. The demand function of a loss-averse consumer is investigated in Section 4. Section 5 derives conditions for the optimality of a flat rate. General nonlinear pricing schemes are analyzed in Section 6. Section 7 discusses further extensions of the baseline model. Section 8 concludes. All proofs of the main analysis in Sections 3 to 5 are contained in the Appendix. The formal derivations of Sections 6 and 7 are developed in the Web Appendix.

2. Illustrative Example

Consider a monopolist who sells a good to a single consumer. The monopolist produces with constant marginal cost $0 < c < 6$. The take-it-or-leave-it offer of the monopolist is a two-part tariff $T(q) = L + pq$, where $q \geq 0$ is the quantity, p is the unit price, and L is the basic charge. The consumer's (intrinsic) consumption utility is quasi linear and given by: $u = \theta q - (1/2)q^2 - T$. The consumer's demand—his satiation point—depends on his demand type $\theta = 6, 10$. With probability $\alpha \in (0,1)$ the type is "low demand" ($\theta = 6$), and with probability $1-\alpha$ the type is "high demand" ($\theta = 10$). The average type is $\bar{\theta} = 6 + (1-\alpha)4$ and the variance is $\sigma^2 = 16\alpha(1-\alpha)$. At the contracting stage, neither the firm nor the consumer knows the demand type. After deciding whether or not to accept the monopolist's offer, the consumer privately observes his demand type and makes his purchasing decision. Moreover, we posit that the consumer is loss averse, in the sense that he incurs a loss when paying more than

14. Karle and Peitz (2010a,b) also study consumer loss aversion in a model of product differentiation.

his reference bill r. His total utility is $u-\mu$, with $\mu = \lambda \max\{T-r,0\}$ and $\lambda \in [0,1]$.[15] For $\lambda = 0$ the consumer has standard preferences, whereas for $\lambda > 0$ he is loss averse.

A key question in the literature on reference-dependent preferences is what determines the reference point. In the wake of Kahneman and Tversky (1979), the status quo is often assumed as the reference point. Here, the status quo could be the previous month's billing amount, which is a fixed amount $r \geq 0$, independent of the accepted contract.

We will argue that in this specification, a flat rate can only be optimal if the basic charge exactly coincides with the reference point. To see this, suppose first that the maximal amount L^F that the monopolist could charge for the flat-rate contract is strictly below the reference point ($L^F < r$). In this case, the monopolist could increase his profit by charging a small but positive unit price. If the unit price is sufficiently small, the payment of both demand types will remain below the reference point. Therefore, both types behave like a standard consumer with $\lambda = 0$ and a flat rate is not optimal.

Next, suppose that the reference point is strictly below the maximal charge for the flat rate ($L^F > r$). In this case, the monopolist would also profit from increasing the unit price to a small positive value. The intuition is that both demand types will now make payments that lead to losses compared to the reference point. It turns out that a loss-averse consumer with a low reference point behaves like a standard consumer with a transformed utility function $u = [\theta q - (1/2)q^2 + \lambda r]/[\lambda + 1]$, where the reference point is a constant shift in the consumer's utility function. Again, for such a consumer, a flat rate is not optimal.

To summarize, we have argued that a flat rate can only be optimal in the special case that the basic charge coincides exactly with the reference point.[16]

In light of this discussion, we will now analyze a reference point that depends on the contract offered by the monopolist. Sticking to the idea that the reference point is determined by the status quo, the basic charge of the two-part tariff seems like a natural choice, because the consumer has to pay the basic charge in any case.[17]

For a given demand type, the consumer chooses q in order to maximize $\theta q - (1/2)q^2 - pq - L - \lambda pq$. Hence, the optimal quantity is given by $q(\theta) = \theta - p(\lambda+1)$. Due to ex-ante contracting, the basic charge L is chosen such that the consumer's expected utility equals zero, i.e., the monopolist extracts the entire expected surplus. Thus, the optimal unit price p^* maximizes the joint surplus $S(p)$—expected utility plus profits,

$$S(p) = \alpha(1/2)(6-p(\lambda+1))^2 + (1-\alpha)(1/2)(10-p(\lambda+1))^2 \\ + (p-c)[\alpha(6-p(\lambda+1)) + (1-\alpha)(10-p(\lambda+1))].$$

15. We restrict λ to be less than 1 in order to guarantee concavity of the monopolist's objective function.

16. Moreover, $L^F = r$ is only a necessary condition for optimality of a flat rate.

17. Note that this is not a knife edge case. For example, we could also choose the payment of the lowest demand type as the reference point and get similar results.

For a consumer with standard preferences ($\lambda = 0$), the monopolist optimally sets the marginal price equal to the marginal cost, i.e. $p^* = c$, so that the consumer fully internalizes production costs. On the other hand, if the consumer is loss averse, it might be optimal to offer a flat-rate tariff, i.e. $p^* = 0$. A flat-rate tariff is optimal when $S'(p)|_{p=0} \leq 0$, which is equivalent to

$$c \leq \frac{\lambda}{\lambda+1}\bar{\theta}. \tag{1}$$

This shows that a flat rate is more likely to be optimal when marginal costs are small or the degree of loss aversion is high.

The empirical literature discussing the insurance effect as a possible explanation of the flat-rate bias stresses that this bias is driven by consumers' uncertainty about future consumption. In the above example, whether or not a flat rate is optimal is independent of the variation in the consumer's demand. In order to capture the effect of uncertainty, we apply a somewhat more complex model of loss aversion. Following Kőszegi and Rabin (2006), we posit that the consumer forms rational expectations about his future demand, which determine his reference point. The reference point is the full distribution of potential billing amounts. A given billing amount is compared to the billing amount the consumer expected to pay in the low-demand state and to the expected billing amount in the high-demand state. While the former comparison is weighted by α, the latter is weighted by $1-\alpha$. The consumer's reference point (distribution) is determined at the contracting stage and is fixed at the point in time when he chooses his actual consumption. For a given reference point $\langle q(6), q(10)\rangle$, the consumer chooses his consumption level q in order to maximize

$$\theta q-(1/2)q^2-pq-L-\lambda\left[\alpha\max\{p(q-q(6)),0\}+(1-\alpha)\max\{p(q-q(10)),0\}\right]$$

Since expected consumption influences actual consumption, Kőszegi and Rabin impose a consistency criterion called personal equilibrium. The expected consumption levels that form the consumer's reference point must coincide with the optimal consumption levels for the respective types. We now show that the type-dependent demand function given by $q(6) = 6 - p$ and $q(10) = 10 - p(1 + \alpha\lambda)$, is a personal equilibrium as long as p is not too high and thus $q(6) \leq q(10)$. Suppose that the consumer's reference point is $\langle q(6), q(10)\rangle$. If type $\theta = 6$ is realized and the consumer chooses a quantity $q \leq q(6)$, then he neither feels a loss compared to the bill for $q(6)$ nor to the bill for $q(10)$ and thus the optimal quantity is $q(6)$. If the consumer chooses a quantity $q(6) < q \leq q(10)$, then he feels a loss of $p(q-q(6))$ which is weighted by α. Due to this loss, the consumer prefers to choose a quantity lower than $q(6)$ and we are back in the former case. Now, suppose that $\theta = 10$ and that the consumer chooses a quantity $q(6) < q \leq q(10)$. Then his utility is given by

$$\underbrace{10q-(1/2)q^2-pq-L}_{\text{intrinsic utility}}-\underbrace{\alpha\lambda p[q-q(6)]}_{\text{loss utility}}.$$

With probability α, the consumer expected to pay only $L + pq(6)$ but his actual bill is $L + pq$. Comparing the actual and the expected bill leads to the sensation of a loss of $p(q-q(6))$. With probability $1-\alpha$, the consumer expected to pay $L+pq(10)$, but since his actual bill is lower, this comparison does not lead to the sensation of a loss. The above utility is maximized by $q=q(10)=10-p(1+\alpha\lambda)$. If the consumer chooses a quantity $q>q(10)$, additional losses arise from the comparison with the payment in the high demand state:

$$\underbrace{10q-(1/2)q^2-pq-L}_{\text{intrinsic utility}}-\underbrace{\alpha\lambda p[q-q(6)]-(1-\alpha)\lambda p[q-q(10)]}_{\text{loss utility}}.$$

Now, the consumer would prefer to choose a $q<q(10)$. Thus, we have shown that $q(6)$ and $q(10)$ indeed constitute a personal equilibrium.

Again, the monopolist chooses the unit price p in order to maximize the joint surplus including expected losses. The joint surplus is given by

$$\begin{aligned} S(p) = {} & \underbrace{\alpha(1/2)(6-p)^2+(1-\alpha)(1/2)(10-p(1+\alpha\lambda))(10-p(1-\alpha\lambda))}_{\text{expected intrinsic utility}} \\ & -\underbrace{\lambda\alpha(1-\alpha)p[4-p\alpha\lambda]}_{\text{expected loss}}+\underbrace{(p-c)[\alpha(6-p)+(1-\alpha)(10-p(1+\alpha\lambda))]}_{\text{expected profits from sales}}. \end{aligned}$$

For a measured tariff, i.e. $p>0$, the consumer expects to incur a loss which reduces his expected utility and in turn the joint surplus. The term $\lambda\alpha(1-\alpha)p[4-p\alpha\lambda]$ captures the "flat-rate premium" that the consumer is willing to pay for the insurance provided by the flat rate. The flat-rate premium vanishes if the unit price goes to zero or if there is no uncertainty in demand, i.e. $\alpha\to 0$ or $\alpha\to 1$. Intuitively, a loss-averse consumer dislikes fluctuations in his billing amount which reduces his willingness to pay for the measured tariff. A flat-rate tariff completely insures the consumer against such fluctuations.

For $\lambda=0$—no loss aversion—marginal cost pricing is optimal, i.e. $p^*=c$. A flat-rate tariff is optimal if $S'(p)|_{p=0}\leq 0$ which is equivalent to

$$c\leq 4\frac{\lambda\sigma^2}{16+\lambda\sigma^2}. \tag{2}$$

As in the case where the reference point equals the basic charge (cf. (1)), a flat rate is more likely to be optimal if marginal costs are small or if the consumer's degree of loss aversion is high. With expectation-based loss aversion, however, a flat rate is also more likely to be optimal if the consumer's demand varies significantly across types (σ^2 high), because the insurance value of a flat rate is higher in these cases. This comparative statics result is not predicted by the previous model in which the reference point is independent of the consumer's expectations about his future consumption choices.

Our model with expectation-based loss aversion predicts, for instance, that one observes flat-rate contracts for rental cars, in particular at vacation resorts where customers are unfamiliar with the network of roads. The model does not predict flat rates for heating oil. Typically, the demand for heating oil is uncertain but the marginal costs are high.

3. Monopolistic Market with Homogeneous Consumers

3.1. Players and Timing

We consider a market where a monopolist produces a single good at constant marginal cost $c > 0$ and without fixed costs. The monopolist offers a two-part tariff to a continuum of ex ante homogeneous consumers of measure one. The tariff is given by $T(q) = L + pq$, where $q \geq 0$ is the quantity, L denotes the basic charge, and p denotes the unit price. At the contracting stage, a consumer does not know his future demand type $\theta \in \Theta := [\underline{\theta}, \bar{\theta}]$. Consumers' demand types are independently and identically distributed according to the commonly known and twice differentiable cumulative distribution function $F(\theta)$ with strictly positive density function $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

The sequence of events is as follows: (1) The monopolist makes a take-it-or-leave-it offer (L, p) to consumers. (2) Each consumer forms expectations about his demand and decides whether or not to accept the offered two-part tariff. (3) Each consumer privately observes his demand type θ. (4) Each consumer who accepted the offer chooses a quantity that maximizes his utility and makes a payment according to the concluded contract.

3.2. Consumers' Preferences

We assume that consumers are loss averse in the sense that a consumer is disappointed if the payment he has to make exceeds his reference payment. Intuitively, a consumer feels a loss if at the end of the month the invoice from his telephone provider is higher than expected. We apply the approach of reference-dependent preferences developed by Kőszegi and Rabin (2006, 2007). First, this concept posits that overall utility has two additively separable components, consumption utility (intrinsic utility) and gain-loss utility. Second, the consumer's reference point is determined by his lagged rational expectations about outcomes. Third, a given outcome is evaluated by comparing it to each possible outcome, where each comparison is weighted with the ex-ante probability of the alternative outcome. Finally, actual choices must coincide with expected choices.

Intrinsic utility is quasi linear in money and given by $u(q, \theta) - T(q)$. For the markets we have in mind, like rental cars or telecommunication services, even if the price per unit is zero, demand is bounded. Therefore, we assume that there exists a satiation point, $q^S(\theta)$, and that overconsumption is harmless, i.e., free disposal is possible. Additionally, we assume that a higher demand type is associated with a stronger need for the good. These assumptions imply that a higher demand type corresponds to a strictly higher satiation point. Heterogeneity of the satiation points is a necessary condition for the optimality of flat rates, or more generally flat parts in a general tariff, as we will elaborate later.

Formally, a consumer's utility without free disposal is given by the function $\hat{u}(q, \theta)$. For each demand type $\theta \in \Theta$, $\hat{u}(., \theta)$ has a unique maximum at the satiation point $q^S(\theta) > 0$. Since we allow for free disposal, the consumer's gross utility is given by:

$$u(q,\theta) = \begin{cases} \hat{u}(q,\theta), & \text{if } q \leq q^S(\theta), \\ \hat{u}(q^S(\theta),\theta) & \text{if } q > q^S(\theta). \end{cases}$$

ASSUMPTION 1.

(i) $\hat{u}(q,\theta)$ satisfies $\partial_q \hat{u}(q,\theta) > 0$ for $q < q^S(\theta)$, and $\partial_{qq}\hat{u}(q,\theta) < -\rho$, and $\partial_{q\theta}\hat{u}(q,\theta) > \kappa$, for all $\theta \in \Theta$ and all $q \geq 0$, where $\rho, \kappa > 0$.

(ii) $\hat{u}(q,\theta)$ is three times continuously differentiable with bounded first, second and third derivatives.

(iii) We normalize utility such that $E_\theta[u(0,\theta)] = 0$ and assume that $\partial_q u(0,\theta)$ is sufficiently big to ensure positive demand whenever necessary.

For example, the quadratic utility function $\hat{u}(q,\theta) = \theta q - (1/2)q^2$, which we used in Section 2, satisfies Assumption 1.[18]

By Assumption 1, the satiation point $q^S(\theta)$ is given by $\partial_q \hat{u}(q^S(\theta),\theta) = 0$. The property that the cross derivative is positive—even when evaluated at the satiation point—guarantees that the satiation point is strictly increasing in the demand type:

$$\frac{dq^S(\theta)}{d\theta} = -\frac{\partial_{q\theta}\hat{u}(q^S(\theta),\theta)}{\partial_{qq}\hat{u}(q^S(\theta),\theta)} > 0. \quad (3)$$

In order to rule out arbitrarily high demand under a flat-rate tariff, we assume that a consumer who is indifferent between two or more quantities, always chooses the lowest of these quantities. Alternatively, one could assume that overconsumption is not harmless.

For simplicity, we depart from the concept of Kőszegi and Rabin (2006, 2007) by assuming that the consumer feels losses only in the money dimension.[19] Relaxing this assumption does not change our main findings as is demonstrated below. The loss function is assumed to be piece-wise linear, since the main driver of loss aversion—in particular for small stakes—is the kink in the value function and not its diminishing sensitivity. If a consumer pays T, but expected to pay $\hat{T}$, then his loss utility is given by

$$\mu(T - \hat{T}) = \max\{\lambda(T - \hat{T}), 0\},$$

with $\lambda \geq 0$. For $\lambda > 0$ the consumer's preferences exhibit loss aversion, whereas $\lambda = 0$ corresponds to the standard case without loss aversion.[20]

18. To ensure positive demand for all types (Assumption 1.iii), we need to assume that $\underline{\theta}$ is not too small.

19. This assumption is also imposed by Spiegler (forthcoming).

20. Our findings are qualitatively robust toward applying the usual two-piece linear gain-loss function with a slope of $\tilde{\eta} > 0$ for gains and a slope of $\tilde{\lambda}\tilde{\eta} > \tilde{\eta}$ for losses. For a more detailed comparison we refer to Section F of the Web Appendix.

Consider a consumer who signed a given contract. His expected demand schedule fully determines the distribution of his expected payments, and thus his reference point. If his demand type is φ and his expected consumption is $\langle q(\theta)\rangle_{\theta\in\Theta}$, his overall utility from this contract when purchasing q units, is given by

$$U(q|\varphi,\langle q(\theta)\rangle) = u(q,\varphi) - T(q) - \int_{\underline{\theta}}^{\bar{\theta}} \mu(T(q) - T(q(\theta)))\, f(\theta)\, d\theta \,. \tag{4}$$

Observe that for $p > 0$, a higher quantity increases the number of demand types compared to which the consumer feels a loss.

To deal with the resulting interdependence between actual consumption and expected consumption, we apply the personal equilibrium concept, which requires that the reference point is given by rational (self-fulfilling) expectations about the consumption decision (Kőszegi and Rabin, 2006, 2007).

DEFINITION 1 (Personal Equilibrium). *For a given unit price p, the demand function* $\langle \hat{q}(\theta,p)\rangle_{\theta\in\Theta}$ *is a personal equilibrium if for all* $\varphi \in \Theta$*,*

$$\hat{q}(\varphi,p) \in \arg\max_{q\geq 0} U(q|\varphi,\langle \hat{q}(\theta,p)\rangle) \,.$$

4. The Demand Function

4.1. Personal Equilibria

In this section, we analyze the personal equilibrium demand function of a consumer who accepted a two-part tariff (p,L). We can restrict attention to nonnegative unit prices, $p \geq 0$. A negative unit price cannot be optimal since overconsumption is harmless. We now show that for low unit prices, there is a unique personal equilibrium. In this personal equilibrium, demand is strictly increasing in the type. For higher unit prices, there are multiple personal equilibria, all of which involve bunches in the demand schedule.

By Assumption 1, higher demand types have a stronger (intrinsic) preference for the good ($\partial_{q\theta} u > 0$). Adding loss utility only in the money dimension does not destroy this property. The consumer's loss utility only depends on how much he consumes and is thus independent of his demand type. For any given reference point $\langle q(\theta)\rangle$, higher demand types still have a stronger preference for the good. Hence, any personal equilibrium demand schedule must be weakly increasing in the demand type.[21]

In a personal equilibrium $\langle \hat{q}(\theta,p)\rangle$, a consumer who consumes q feels losses compared to all types who consume less than q. If $\hat{q}(\theta,p)$ is strictly increasing in

21. The assumption that losses are felt only in the money dimension is crucial. If the consumer also feels losses in the good dimension, personal equilibrium demand need not be monotone.

θ, he therefore feels losses compared to all types below a cutoff type $\alpha(q)$ defined by $\hat{q}(\alpha(q), p) = q$.[22] With this notation, the utility of a consumer with demand type θ who consumes q can be expressed as

$$U(q|\theta, \langle \hat{q}(\varphi, p) \rangle) = u(q, \theta) - pq - L - \lambda p \int_{\underline{\theta}}^{\alpha(q)} [q - \hat{q}(\varphi, p)] f(\varphi) \, d\varphi \,. \tag{5}$$

The utility function (5) is strictly concave and thus the optimal quantity is characterized by the first-order condition. Taking the first-order condition and using that in equilibrium a consumer with demand type θ feels losses compared to all types below θ, i.e. $\alpha(q) = \theta$, we obtain the following condition which characterizes the demand function of a strictly increasing personal equilibrium:

$$\partial_q u(\tilde{q}(\theta, p), \theta) = p[1 + \lambda F(\theta)]. \tag{6}$$

The solution to this equation is a candidate for a personal equilibrium and shall be denoted as $\tilde{q}(\theta, p)$. For a standard consumer without loss aversion ($\lambda = 0$), $\tilde{q}(\theta, p)$ equates marginal utility and unit price. For $p = 0$, each type demands his satiation point independently of his degree of loss aversion, i.e., $\tilde{q}(\theta, 0) = q^S(\theta)$. For $p > 0$, a loss-averse consumer perceives a loss compared to lower demand types which are paying lower bills. Thus, loss aversion leads to a downward distortion of demand for all but the lowest demand type and this distortion is increasing in the type. Finally, $\tilde{q}(\theta, p)$ has the reasonable property that it is decreasing in the unit price.

The candidate $\tilde{q}(\cdot)$ constitutes a personal equilibrium only if it is strictly increasing in the demand type, which is equivalent to the following condition:

CONDITION 1. *For all $\theta \in \Theta$,*

$$p < \frac{\partial_{q\theta} u(\tilde{q}(\theta, p), \theta)}{\lambda f(\theta)} \,. \tag{C1}$$

In the appendix, we show that every personal equilibrium must be strictly increasing if $\tilde{q}(\cdot)$ is strictly increasing in the demand type. This enables us to show our first result.

PROPOSITION 1. *Suppose that Condition 1 holds. Then there exists a unique personal equilibrium, which is given by $\langle \tilde{q}(\theta, p) \rangle_{\theta \in \Theta}$.*

Proposition 1 shows existence and uniqueness of the strictly increasing personal equilibrium $\tilde{q}(\cdot)$ if the unit price is not too high. In particular, the personal equilibrium

22. Lemma A.3 in the Appendix shows that q must be continuous in the demand type. The proof is straightforward because two-part tariffs are continuous. The absence of continuity of demand in the case of general tariffs is a major complication of the generalized model analyzed in Section 6.

is unique if $p < \bar{p} := \min_\theta\{\kappa/[\lambda f(\theta)]\}$, where κ is the lower bound of $\partial_{q\theta}u$. The range of unit prices for which Condition 1 holds depends (i) on the degree of loss aversion, and (ii) on the heterogeneity of preferences. If preferences vary significantly with the demand type, maximizing intrinsic utility requires an increasing demand schedule. Minimizing losses, on the other hand, requires a flat demand schedule, in particular if the unit price is high. Therefore, Condition 1 is more likely to be satisfied if loss aversion is not very strong (λ is small), and there is sufficient heterogeneity of preferences measured by $\partial_{q\theta}u(q,\theta)/f(\theta)$.[23] Nevertheless, around a flat-rate tariff the personal equilibrium is always unique.

If Condition 1 is violated, there are multiple personal equilibria. We know that any personal equilibrium demand schedule is weakly increasing in the demand type. If there are multiple personal equilibria, then all of these equilibria consist of strictly increasing parts that coincide with $\tilde{q}(\theta,p)$, and flat parts that connect the strictly increasing parts (cf. Lemmas A.3 and A.4 in the Appendix). It can be shown that every demand function that satisfies these properties is a personal equilibrium.[24]

In order to specify consumer behavior in the case of multiple equilibria, we assume that consumers behave according to the *preferred personal equilibrium (PPE)*, defined by Kőszegi and Rabin (2006, 2007). Here, the PPE is the consumption plan among all credible plans—personal equilibria—that maximizes the consumer's expected utility. We denote the PPE by $q^{\mathrm{PPE}}(\theta,p)$.

LEMMA 1. *The preferred personal equilibrium is given by*

$$q^{PPE}(\theta,p) = \max_{\varphi\le\theta} \tilde{q}(\varphi,p).$$

4.2. Participation and Flat-Rate Bias

Whether a consumer accepts the offered two-part tariff depends on the utility he expects to enjoy with this contract. The consumer's expected utility from accepting the two-part tariff (p,L)—taking the personal equilibrium into account—is,

$$\begin{aligned}\mathbb{E}_\theta[U(\hat{q}(\theta,p)|\theta,\langle\hat{q}(\theta,p)\rangle)] = &\int_{\underline{\theta}}^{\bar{\theta}}[u(\hat{q}(\theta,p),\theta)-p\hat{q}(\theta,p)]f(\theta)\,d\theta - L\\ &-\lambda p\int_{\underline{\theta}}^{\bar{\theta}}\int_{\underline{\theta}}^{\theta}[\hat{q}(\theta,p)-\hat{q}(\varphi,p)]f(\varphi)f(\theta)\,d\varphi d\theta\,. \quad (7)\end{aligned}$$

The first term of this expression represents expected intrinsic utility. The second term is the ex ante expected loss, which is weighted by λ. The expected loss vanishes if the

23. The term $\partial_{q\theta}u(q,\theta)/f(\theta)$ is a measure of heterogeneity of preferences. It increases if types are more dispersed, and if the marginal utility depends more strongly on the demand type.

24. If $\tilde{q}(\theta,p)$ is decreasing in θ all personal equilibria are constant and the demanded quantity can take any value from the range of $\tilde{q}$.

unit price goes to zero. Moreover, the expected loss is relatively low if demand does not vary significantly across different demand types. If, on the other hand, demand is highly uncertain, then a loss-averse consumer who subscribed to a measured tariff expects to incur severe losses, which reduces his willingness to pay for the contract.

The expected losses that a loss-averse consumer incurs with a measured tariff make him biased in favor of flat-rate tariffs. This bias can be so severe that the consumer favors a flat rate over a measured tariff, although the average bill under the measured tariff for satiated consumption is lower than the basic charge of the flat rate. We define a consumer's preferences as "flat-rate biased" if, for any flat-rate tariff ($p = 0$) there exists a measured tariff ($p > 0$) such that (i) ex ante the consumer prefers the flat-rate contract, and (ii) the expected bill for satiated consumption under the measured tariff is lower than the basic charge of the flat rate.

PROPOSITION 2 (Flat-Rate Bias). *A loss-averse consumer ($\lambda > 0$) with uncertain demand has flat-rate biased preferences.*

Proposition 2 shows that for each flat-rate contract, there is some measured tariff, such that the consumer prefers the flat rate although the expected payment under the measured tariff is lower. A natural question is whether this flat-rate bias can also arise if we restrict attention to two-part tariffs that are offered by a profit-maximizing monopolist. In Section 7.1, we provide an example of a monopolist who faces two groups of consumers. The optimal menu of two-part tariffs induces one group of consumers to choose a flat rate, even though the flat rate is more expensive than satiated consumption under the alternative measured tariff.

5. The Optimality of Flat-Rate Tariffs

The monopolist maximizes expected revenues minus expected costs subject to the constraint that consumers voluntarily accept the two-part tariff:

$$\max_{L,p\geq 0} \; L + (p-c)\int_{\underline{\theta}}^{\bar{\theta}} \hat{q}(\theta,p)f(\theta)d\theta$$
$$\text{subject to } \; \mathbb{E}_\theta\left[U(\hat{q}(\theta,p)|\theta,\langle\hat{q}(\varphi,p)\rangle)\right] \geq 0.$$

For any unit price p, the optimal fixed fee is determined by the binding participation constraint. Thus, the monopolist's tariff choice problem can be restated as a problem of choosing only the unit price p. Since there is no asymmetric information at the contracting stage, the monopolist can extract the entire expected gains from trade, net of expected losses. The optimal unit price p^* maximizes the joint surplus of the two contracting parties which we denote by $S(p)$:

$$S(p) = \int_{\underline{\theta}}^{\bar{\theta}} \left\{ u(\hat{q}(\theta,p),\theta) - c\hat{q}(\theta,p) - p\lambda \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,p) - \hat{q}(\varphi,p)]f(\varphi)d\varphi \right\} f(\theta)\, d\theta. \tag{8}$$

Without loss aversion ($\lambda = 0$), the joint surplus equals the consumers' expected intrinsic utility minus the firm's expected costs of production. Ex ante, a loss-averse consumer expects to feel a loss if tariff payments depend on the purchased quantities and demand differs across types. This expected loss reduces the joint surplus. When choosing the optimal unit price, the monopolist faces a trade-off between maximizing standard efficiency and minimizing the consumers' expected losses. Intuitively, for a high degree of loss aversion, minimizing expected losses is more important whereas for a high marginal cost, maximizing standard efficiency is more important.

We can use equation (6) in order to write the derivative of the joint surplus as

$$S'(p) = \Psi(p) + \lambda p \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{t(\theta,p)} \partial_p \hat{q}(\varphi, p) f(\varphi) d\varphi f(\theta) d(\theta), \tag{9}$$

where

$$\Psi(p) = (p-c) \int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta, p) f(\theta) d\theta - \lambda \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta, p) - \hat{q}(\varphi, p)] f(\varphi) f(\theta) d\varphi d\theta.$$

In this expression, $t(\theta, p)$ is the lowest type that consumes the same quantity as θ. If there is a unique and strictly increasing personal equilibrium, we have $t(\theta, p) = \theta$. On the other hand, if θ is a type who consumes the same amount as some adjacent types in the personal equilibrium, then $t(\theta, p) < \theta$.

If the consumer has standard preferences ($\lambda = 0$), then the first-order condition $S'(p) = 0$ is satisfied for marginal cost pricing, i.e. $p^* = c$. In any case, for $p > c$ the joint surplus is strictly decreasing in p.[25] Thus, $p^* \in [0, c]$. Notice that the marginal joint surplus is bounded from above by $\Psi(p)$. We impose the following assumption:[26]

ASSUMPTION 2.
(i) $\lambda \leq 1$ *(no dominance of loss utility).*
(ii) $\partial_{qqq} u(q, \theta) \geq 0$ *(convex demand).*

While Assumption 2 is by far not necessary for flat-rate tariffs to be optimal, it guarantees that $\Psi(p)$ is non-increasing for $p \leq c$. This implies that $S'(p)|_{p=0} \leq 0$ is a necessary and sufficient condition for the optimality of flat rates. This condition is easy to interpret and allows us to make statements about comparative statics.

In order to guarantee that $S'(p)|_{p=0} \leq 0$ is a sufficient condition, we have to rule out that a higher unit price leads to a reduction in expected losses, which may happen due to a highly compressed demand profile. A higher unit price has two effects. On the one hand, it increases expected losses due to increased variations in payments for

25. For a given θ, $q^{\mathrm{PPE}}(\theta, p)$ as a function of p, is the maximum of strictly decreasing functions with $\partial_p \tilde{q}(\theta, p) \leq \varepsilon < 0$ by Assumption 1. Therefore, $q^{\mathrm{PPE}}(\theta, p)$ is strictly decreasing in p.

26. No dominance of loss utility is also imposed, for instance, by Herweg et al. (2010). In our setup, $\lambda = 1$ corresponds to the conventional estimate of two-to-one loss aversion.

a given demand function. On the other hand, consumers react to the higher unit price by choosing a more compressed demand function, which reduces expected losses. In summary, Assumption 2 ensures that the direct effect on expected losses is always stronger than the indirect effect.

To cut back on our lengthy formulas we define

$$\Sigma(\lambda) := \lambda \, \frac{\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [q^S(\theta) - q^S(\varphi)] f(\varphi) f(\theta) d\varphi d\theta}{- \int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta, 0) f(\theta) d\theta} \, .$$

Note that $\hat{q}(\theta, p)$ also depends on λ. Obviously, $\Sigma(0) = 0$. Moreover, it can be shown that $\Sigma(\lambda)$ is strictly increasing in λ and thus $\Sigma(\lambda) > 0$ for $\lambda > 0$. Noting that $S'(p)|_{p=0} \leq 0$ is equivalent to $\Sigma(\lambda) \geq c$, we state the main result of this section.

PROPOSITION 3. *Suppose that Assumption 2 holds. Then, the monopolist optimally offers a flat-rate tariff, i.e.* $p^* = 0$, *if and only if* $\Sigma(\lambda) \geq c$. *Moreover,* $\Sigma'(\lambda) > 0$.

According to Proposition 3, if the consumer is loss averse, then a flat-rate tariff is optimal for sufficiently low marginal costs, since $\Sigma(\lambda) > 0$. On the one hand, a flat-rate tariff eliminates losses felt by the consumer, which increases his willingness to pay for the contract. On the other hand, a flat-rate tariff leads to an inefficiently high level of consumption which is costly to the monopolist. If the degree of loss aversion is high and marginal costs are low, the positive effect due to minimized losses outweighs the negative effect of higher costs due to overconsumption, and thus a flat-rate tariff is optimal.

The range of marginal cost values for which a flat-rate tariff is optimal is increasing in the consumer's degree of loss aversion and in the uncertainty of demand.[27] A flat-rate contract can be optimal only if demand is sufficiently uncertain. The numerator of $\Sigma(\lambda)$ is a measure for the degree of demand variation. Intuitively, if all types have very similar preferences, the monopolist can set a positive unit price such that all types consume close to the efficient quantity. At the same time, expected losses are small because the efficient quantity varies very little. Thus, if all types have similar preferences, a measured tariff is optimal. Conversely, a strong variation in satiation points increases $\Sigma(\lambda)$ and makes a flat rate optimal for a wider range of marginal cost levels.[28]

Finally, a measured tariff is optimal if demand reacts sensitively to price changes. The denominator of $\Sigma(\lambda)$ measures how strong on average a consumer's demand reacts to an increase of the unit price slightly above zero. A positive unit price reduces costly overconsumption compared to a flat rate. On the other hand, it introduces losses

27. Note that $\Sigma(\lambda)$ is not linear in λ. The example at the end of this section illustrates this point.

28. Equation (3) shows that the difference between two satiation points is increasing in the cross derivative of the intrinsic utility function.

felt by the consumer. When demand reacts strongly to price changes, the reduction of overconsumption dominates and a measured tariff is more likely to be optimal.

EXAMPLE 1. To illustrate the optimality of flat-rate tariffs and in particular to highlight the importance of demand uncertainty, suppose that $\theta \sim U[\mu-\sigma,\mu+\sigma]$ with $\mu+\sigma > \mu-\sigma > \varepsilon > 0$. The mean of the demand type distribution is $\mathbb{E}[\theta] = \mu$ and the variance is $\mathbb{V}[\theta] = (1/3)\sigma^2$. Let the intrinsic utility function for the good be given by $\hat{u}(q,\theta) = \theta q - (1/2)q^2$. Here, the only parameter that affects demand uncertainty is the size of the support of the type distribution (given by σ).[29]

For $p < \bar{p} = 2\sigma/\lambda$, the personal equilibrium is unique and demand is strictly increasing in the demand type. Solving (6) yields

$$\hat{q}(\theta,p) = \theta - p\left[1+\lambda\frac{1}{2\sigma}(\theta-\mu+\sigma)\right] = \left[1-\frac{p}{\bar{p}}\right]\theta - p\left[1+\frac{\sigma-\mu}{\bar{p}}\right]. \tag{10}$$

For $p > \bar{p}$, Condition 1 is violated and there are multiple personal equilibria. In all of these personal equilibria, all demand types $\theta \in [\mu-\sigma,\mu+\sigma]$ consume the same amount $\bar{q}$, with $(\mu+\sigma) - p(1+\lambda) \leq \bar{q} \leq (\mu-\sigma) - p$. By Lemma 1, the preferred personal equilibrium is given by the upper bound, i.e., $q^{\mathrm{PPE}}(p) = \max\{(\mu-\sigma)-p, 0\}$.

In this example, the joint surplus is a quasi-concave function for all unit prices and all degrees of loss aversion as long as demand is positive. Thus, a flat-rate tariff is optimal if and only if $S'(p)|_{p=0} \leq 0$, which is equivalent to

$$c \leq \frac{2}{3}\,\frac{\lambda\sigma}{2+\lambda} = \Sigma(\lambda). \tag{11}$$

Note that $\Sigma'(\lambda) > 0$ and $\lim_{\lambda\to\infty}\Sigma(\lambda) = (2/3)\sigma$. ■

RESULT 1. *Consider the specifications of the example. A flat-rate tariff is optimal if either the marginal cost is sufficiently low or demand is sufficiently uncertain. A high degree of loss aversion makes it more "likely" that a flat-rate tariff is optimal. An arbitrarily high degree of loss aversion, however, is not sufficient to ensure optimality of a flat-rate tariff.*

6. General Nonlinear Tariffs

In this section, we relax the restriction to two-part tariffs. The monopolist can now offer a general tariff $T(q)$ to the consumers. We find that the main insights from the analysis of two-part tariffs are robust. A flat part in the optimal tariff arises if (i) the consumers are sufficiently loss averse, (ii) there is sufficient variation in preferences

29. Keeping σ constant and varying the cross derivative of $\hat{u}$ would have the same qualitative effects.

and thus sufficient demand uncertainty, and (iii) the marginal cost is not too high. On the other hand, we show that a monopolist who is not restricted to two-part tariffs does not offer a fully flat tariff. The optimal tariff is increasing at the lower end and at the upper end of the range of quantities demanded by consumers. Flat parts arise for intermediate quantities.

The formal derivation of the results can be found in the Web Appendix. To simplify the analysis, we augment Assumption 1 by

ASSUMPTION 3. $\partial_{qq\theta}\hat{u}(q,\theta) < 0$, *for all* $\theta \in \Theta$ *and all* $q \geq 0$.

Applying the revelation principle, we can restrict the monopolist's offer to the class of direct mechanisms $\langle q(\theta), P(\theta)\rangle_{\theta\in\Theta}$ for which truth-telling is a personal equilibrium.[30] As in the case of two-part tariffs, the personal equilibrium constraints imply that $q(\theta)$ and $P(\theta)$ must be non-decreasing.

We define $V(\theta,\varphi)$ as the utility of a consumer with true type φ who pretends to be of type θ, given he expected ex ante to reveal his type truthfully. By monotonicity, this can be written as

$$V(\theta,\varphi) = u(q(\theta),\varphi) - P(\theta) - \lambda \int_{\underline{\theta}}^{\theta} [P(\theta) - P(z)] f(z)\, dz.$$

The monopolist's problem is then given by:

$$\begin{aligned} &\max_{\langle q(\theta),P(\theta)\rangle_{\theta\in\Theta}} && \int_{\underline{\theta}}^{\bar{\theta}} [P(\theta) - cq(\theta)] f(\theta)\, d\theta \\ &\text{subject to} && \mathbb{E}_\theta[V(\theta,\theta)] \geq 0, && \text{(IR)} \\ &\text{and} && \varphi \in \arg\max_{\theta\in\Theta} V(\theta,\varphi),\ \forall\, \varphi \in \Theta. && \text{(PE)} \end{aligned}$$

The monopolist maximizes expected revenues minus production costs. The individual rationality constraint (IR) ensures that consumers voluntarily accept the mechanism at the contracting stage. (PE) ensures that truth-telling is a personal equilibrium.

We can express local personal equilibrium constraints, i.e. $V_1(\theta,\theta) = 0$, by the following "revenue-equivalence" formula:[31]

$$P'(\theta) = \frac{\partial_q u(q(\theta),\theta)}{1+\lambda F(\theta)} q'(\theta). \tag{12}$$

Given this, we observe that flat parts in the payment schedule can arise (a) due to bunching in the quantity schedule ($q'(\theta) = 0$) or (b) due to a quantity schedule that

30. The tariff $T(q)$ can be recovered from a mechanism $\langle q(\theta), P(\theta)\rangle$: $T(q) = P(\inf\{\theta | q(\theta) \geq q\})$.

31. Note that (6) is a special case of (12) since $T'(q(\theta)) = P'(\theta)/q'(\theta)$. Lemma B.1 in the Web Appendix shows that the optimal mechanism is absolutely continuous and that (12) is necessary and sufficient for (PE). Revenue equivalence my fail if we introduce loss aversion in the good dimension and the reference distribution has mass points (see Carbajal and Ely (2010) for an example).

coincides with the satiation point ($\partial_q u(q(\theta),\theta)=0$). We focus on the latter case which has the feature that not only the payment schedule as a function of the type ($P(\theta)$) is flat, but also the tariff that maps quantities into payments ($T(q)$) has a flat part.

If quantity-bunching does not occur, the optimal solution can be characterized by the following first-order condition,

$$\partial_q u(q(\theta),\theta)=\frac{(1+\lambda F(\theta))^2}{\lambda+1}c-F(\theta)(1-F(\theta))\frac{\lambda(1+\lambda F(\theta))}{\lambda+1}\frac{\partial_{q\theta}u(q(\theta),\theta)}{f(\theta)}. \tag{13}$$

By Assumption 3, this equation has a solution—there exists a $\tilde{q}(\theta)<q^S(\theta)$ that solves (13)—if and only if

$$c-\frac{\lambda F(\theta)(1-F(\theta))}{1+\lambda F(\theta)}\frac{\partial_{q\theta}\hat{u}(q^S(\theta),\theta)}{f(\theta)}>0. \tag{14}$$

Assumption 3 also implies uniqueness of the solution if it exists. If (14) is violated, the optimal solution is given by $q^S(\theta)$. We define

$$q^*(\theta)=\begin{cases}\tilde{q}(\theta), & \text{if (14) is fulfilled for } \theta,\\ q^S(\theta), & \text{otherwise.}\end{cases}$$

Moreover, we define $P^*(\theta)$ as the payment rule that satisfies the (IR) constraint with equality and equation (12).[32] Now we can state the main result of this section.

PROPOSITION 4. *If q^* is strictly increasing, then (q^*,P^*) is an optimal solution to the monopolist's problem.*

Equations (13) and (14) reveal several observations about the location of flat parts in the tariff, and about the distortions of the optimal quantity schedule compared to the first-best solution ($\partial_q u(q^{\text{FB}}(\theta),\theta)=c$), which is optimal in the absence of loss aversion. By equation (14), a flat part is more likely for intermediate quantities, i.e., quantities which are chosen by types in the middle of the type space. Conversely, for $\theta\in\{\underline{\theta},\bar{\theta}\}$, (14) is never violated and we have

$$\begin{aligned}\partial_q u(q^*(\underline{\theta}),\underline{\theta}) &= \frac{1}{\lambda+1}\,c \qquad \Rightarrow \qquad q^*(\underline{\theta})\in(q^{\text{FB}}(\underline{\theta}),q^S(\underline{\theta})),\ \text{if } \lambda>0,\\ \partial_q u(q^*(\bar{\theta}),\bar{\theta}) &= (\lambda+1)\,c \qquad \Rightarrow \qquad q^*(\bar{\theta})<q^{\text{FB}}(\bar{\theta}),\ \text{if } \lambda>0.\end{aligned}$$

The tariff is never flat for quantities chosen by very low or very high demand types. We also observe that the coefficient of c in the first-order condition is increasing in

32. Dividing (13) by $(1+\lambda F(\theta))$ and using $T'=P'/q'$, we can show that $T'(q(\theta))<c$ for $\theta<\bar{\theta}$ and $T'(q(\bar{\theta}))=c$. This is reminiscent of the result that $p\leq c$ in the optimal two-part tariff.

the type if $\lambda > 0$. This implies that (ignoring the effect of the second term), the quantity schedule becomes flatter and more compressed if the consumer is loss averse. Intuitively, this decreases the variation in payments (cf. equation (12)) which decreases ex ante expected losses.

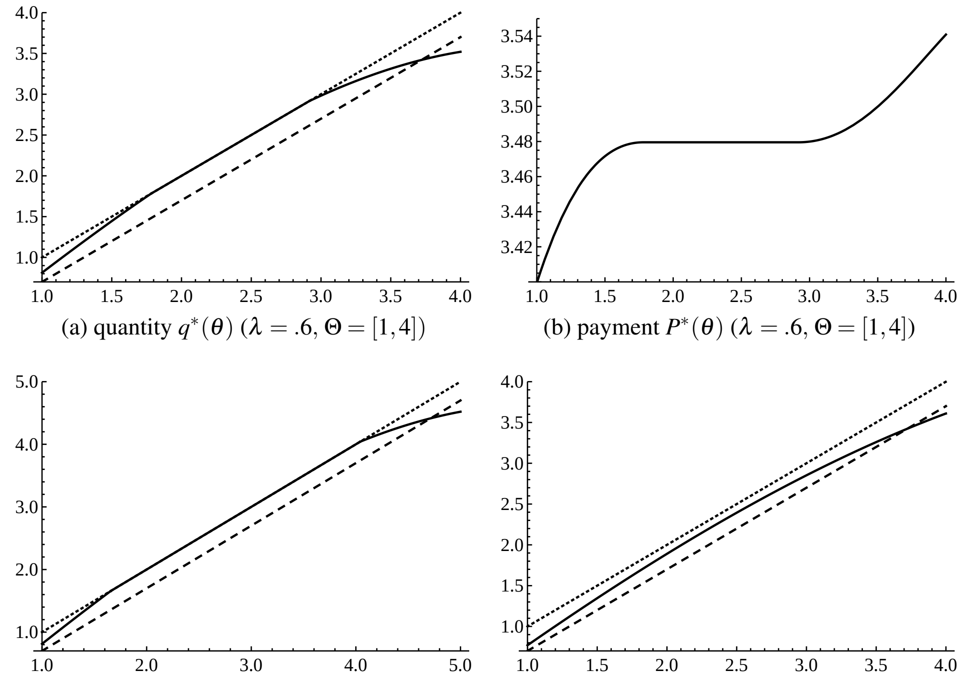
The second term in the first-order condition leads to an increase in the quantity consumed by intermediate types if $\lambda > 0$. If (14) is violated, this increase is so large that the optimal quantity coincides with the satiation point and the optimal tariff becomes flat. A flat part is particularly likely if c is small and consumers' preferences are sufficiently heterogeneous at the satiation point (i.e. $\partial_{q\theta}u(q^S(\theta),\theta)/f(\theta)$ is large). Intuitively, the upward distortion of the quantity schedule arises because the variation in payments depends on the marginal utility of consumption (cf. equation (12)). Increasing the quantity consumed by a type θ to the satiation point, where the marginal utility is zero, therefore makes the payment schedule flat around θ. Compared to a situation where $P(\theta)$ increases steeply around θ, this reduces the losses that types above θ feel compared to types below θ. In the ex-ante surplus, the reduction in losses is weighted by the mass of types that feel less losses $(1-F(\theta))$, times the weight of the types below θ in the loss function $(F(\theta))$. Hence, a flat part is more likely in the middle of the type-space, where $F(\theta)(1-F(\theta))$ is large. The following parametric example illustrates the main insights of this section.

EXAMPLE 2. Let $\hat{u}(q,\theta) = q\theta - (1/2)q^2$ and thus $q^S(\theta) = \theta$. Let $\theta \sim U[\underline{\theta}, \underline{\theta}+1/d]$ so that $F(\theta) = d(\theta - \underline{\theta})$ and $f(\theta) = d$. The optimal quantity schedule is now given by

$$q^*(\theta) = \min\left\{\theta\,,\,\theta - \frac{(1+\lambda d(\theta-\underline{\theta}))^2}{\lambda+1}c + \frac{\lambda}{\lambda+1}(1+\lambda d(\theta-\underline{\theta}))(1-d(\theta-\underline{\theta}))(\theta-\underline{\theta})\right\},$$

and we have $q(\bar{\theta}) = \bar{\theta} - (\lambda+1)c$ and $q(\underline{\theta}) = \underline{\theta} - c/(\lambda+1)$. In Figure 1, the optimal quantity schedules are plotted for different combinations of loss aversion $\lambda \in \{.3, .6\}$ and preference heterogeneity $d \in \{.25, .33\}$. Moreover, we set $c = .3$ and $\underline{\theta} = 1$. In panel (a), loss aversion is high ($\lambda = .6$) and preference heterogeneity is low ($d = .33$). We see that about 2/5 of the types are satiated. For the quantities consumed by these types, the tariff is flat (panel (b)). If we increase preference heterogeneity by stretching the support of θ (panel (c)), this fraction increases to about 3/5. Conversely, if we keep preference heterogeneity fixed and decrease loss aversion to $\lambda = .3$, the flat part in the tariff vanishes and no demand type is satiated.

Table 1 shows the surplus under the optimal tariff (row 1). The second row shows the surplus under a flat rate. For the first two parameter sets, the flat rate is not the optimal two-part tariff. Nevertheless, the loss in surplus compared to the optimal tariff is very low. For the last parameter set, the flat rate is the optimal two-part tariff and the difference to the surplus under the optimal tariff almost vanishes. In contrast, there is a significant loss of surplus if the monopolist offers a cost-based tariff (row 3). ■



(a) quantity $q^*(\theta)$ ($\lambda = .6$, $\Theta = [1,4]$)

(b) payment $P^*(\theta)$ ($\lambda = .6$, $\Theta = [1,4]$)

(c) quantity $q^*(\theta)$ ($\lambda = .6$, $\Theta = [1,5]$)

(d) quantity $q^*(\theta)$ ($\lambda = .3$, $\Theta = [1,4]$)

FIGURE 1. Quantity schedules: satiation (dotted), first-best (dashed), optimal (solid).

The analysis of the optimal general tariff is limited in two ways. First, we only consider the case that the optimal solution $q^*(\theta)$ is increasing in θ. For high values of λ, it can be the case that the solution of (13) is decreasing at the top. We conjecture that in this case, the optimal solution prescribes a constant quantity for an interval of types at the top. This could be approximated by a flat rate with a quantity limit.

Second, we have implicitly assumed that the consumers play the truth-telling personal equilibrium. We cannot rule out that there are other personal equilibria that yield a higher expected utility for the consumers. Therefore, the surplus figures from Table 1 should be interpreted as an upper bound on the achievable surplus. Notice, however, that the multiplicity problem disappears if the monopolist offers a flat rate. Table 1 therefore suggests that the monopolist can ensure a surplus very close to the upper bound.

$c = .3$	$\Theta = [1,4]$, $\lambda = .6$		$\Theta = [1,4]$, $\lambda = .3$		$\Theta = [1,5]$, $\lambda = .6$	
	surplus	loss	surplus	loss	surplus	loss
optimal tariff	8.28	-	8.30	-	17.09	-
flat rate	8.25*	.32%	8.25*	.60%	17.07	.15%
cost-based tariff	7.51	9.26%	7.96	4.05%	15.22	10.96%

*The flat rate is not the optimal two-part tariff for these parameters.

TABLE 1. Expected surplus under different tariff choices.

7. Extensions and Robustness

In this section, we go back to the setting of Sections 3–5 in which the monopolist is restricted to two-part tariffs and discuss several extensions.

7.1. Heterogeneous Consumers

First, we relax the assumption of an ex ante homogeneous group of consumers. Experimental evidence shows that people differ significantly with respect to their degrees of loss aversion, see for instance Choi et al. (2007). In line with our previous analysis, we posit that the degree of loss aversion is a stable preference parameter that is known to consumers already at the contracting stage. We suppose that there are two groups of consumers denoted by $j = 1,2$, with $0 \leq \lambda_1 < \lambda_2$. Consumers from group one are less loss averse than consumers from group two. The degree of loss aversion is private information. Otherwise, the model stays the same. In particular, we abstract from heterogeneity in the distribution of demand types. The monopolist now offers a menu of two-part tariffs, $T_j(q) = L_j + p_j q$, $j = 1,2$, where L_j is the basic charge, and p_j is the unit price of the tariff intended for group j. The timing of the game is the same as before.

To analyze the robustness of our findings regarding the optimality of flat-rate contracts, we focus on cases where the monopolist would offer a flat-rate tariff under symmetric information, i.e., we assume that $\Sigma(\lambda_2) \geq c$. In these cases, the monopolist can screen differently loss-averse consumers at no cost.

Since the degree of loss aversion is private information, the monopolist has to ensure that each type selects the contract that is intended for him. Suppose that the monopolist offers the tariffs T_1 and T_2 that would be optimal for the respective groups in the absence of self-selection constraints (i.e. the optimal tariffs from Section 5). If a flat rate is optimal for both groups, T_1 and T_2 are identical and self-selection constraints are trivially fulfilled. It turns out that the latter is also the case if T_1 is a measured tariff.

LEMMA 2. *Consider a two-part tariff (p,L). Then,*

$$\frac{d}{d\lambda}\left[\mathbb{E}_\theta[U(\hat{q}(\theta,p)|\theta,\langle q(\theta,p)\rangle)]\right] \leq 0.$$

According to Lemma 2, consumers from group two derive lower expected utility from choosing T_1 than consumers from group one because they are more loss averse. Intuitively, higher loss aversion increases the losses felt by a consumer and leads to a more distorted quantity choice under a measured tariff. Both effects reduce the ex ante expected utility. As the optimal contract T_1 leaves no rent to consumers from group one, Lemma 2 implies that consumers from group two must strictly prefer the flat rate. On the other hand, the expected utility from signing a flat-rate contract is independent of the degree of loss aversion and therefore equals zero for both consumer groups. This shows that the less loss-averse consumers are indifferent between the two tariffs.

In summary, the monopolist can screen consumers with respect to their degrees of loss aversion at no cost.

PROPOSITION 5. *Suppose that Assumption 2 holds for both consumer groups.*

(i) If $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$, then the monopolist offers a measured tariff (p_1^*, L_1^*) next to a flat-rate contract $(0, L^F)$. The measured tariff is signed by consumers of group one while consumers of group two sign the flat-rate contract.

(ii) If $c \leq \Sigma(\lambda_1) < \Sigma(\lambda_2)$, then the monopolist offers only a flat-rate contract $(0, L^F)$ which is signed by consumers of both groups.

The tariffs (p_1^*, L_1^*) and $(0, L^F)$ satisfy: $S'(p_1^*|\lambda_1) = 0$, $L_1^* = S(p_1^*|\lambda_1) + (c - p_1^*) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}(\theta, p_1^*) f(\theta) d\theta$, and $L^F = S(0) + c \int_{\underline{\theta}}^{\bar{\theta}} q^S(\theta) f(\theta) d\theta$, with $p_1^* \in (0, c]$ and $L_1^* < L^F$.

Part (i) of Proposition 5 identifies a case in which the monopolist offers a flat-rate contract next to a measured tariff. Thus, consumer heterogeneity with respect to the degree of loss aversion provides one possible answer to the question why firms offer flat rates next to measured tariffs. If the degree of loss aversion of both groups is above the threshold given by $\Sigma(\lambda) = c$, then the monopolist offers only a flat-rate contract.

One could also relax the assumption that, ex ante, consumers differ only in their degrees of loss aversion. If consumer groups also differ in their demand-type distributions, the monopolist faces a sequential screening problem with loss-averse consumers.[33] To illustrate the complications that can arise in this setting, suppose that there are two groups of consumers. For simplicity, assume that both groups have the same degree of loss aversion. Suppose further, that the demand of group one is low on average but fairly uncertain such that it would be optimal to offer a flat-rate tariff to these consumers. Finally, suppose that the demand of group two is high on average but rather certain, so that it would be optimal for the monopolist to offer a measured tariff to consumers of group two. If consumers are privately informed about their demand type distributions, this menu of tariffs would not be incentive compatible. The high demand consumers of group two would prefer the flat rate intended only for consumers of group one. If, on the other hand, the correlation between average demand and demand uncertainty is reversed, offering a flat rate next to a measured tariff can again be optimal.

In summary, whether the optimal menu of two-part tariffs comprises a flat-rate option depends on the precise nature of heterogeneity. In the following example, consumers are heterogeneous both with respect to the demand type distribution and the degree of loss aversion. In this example it is optimal to offer a flat-rate tariff to consumers with low demand and high uncertainty because they are sufficiently more loss averse than the consumers from the other group. This shows that the simple intuition from the discussion above can break down if there is heterogeneity also in

33. See Courty and Li (2000) for the seminal paper on sequential screening in the standard framework.

the degree of loss aversion. Moreover, the example exhibits the empirically observed flat-rate bias.

EXAMPLE 3. We extend Example 1 and assume that there are two groups of consumers indexed by $j = 1,2$. The utility function and the parametrization of the type distribution are the same, except that we choose different parameters for the two groups. Let $(\mu_1, \sigma_1) = (1.3, .1)$ and $(\mu_2, \sigma_2) = (1.17, 1)$. Moreover, we assume that loss aversion is low in group one, $\lambda_1 = .3$, and high in group two, $\lambda_2 = .6$. The marginal cost of the monopolist is $c = .1$. We calculate the optimal two-part tariff for each group in isolation (see Section 5). For group one, the optimal tariff has a positive unit price and is given by $(L_1, p_1) \approx (.744, .081)$. For group two, it is optimal to offer a flat rate with basic charge $L_2 \approx .851$. With this menu of tariffs, consumers from group one strictly prefer the measured tariff and consumers from group two strictly prefer the flat rate. This example demonstrates the flat-rate bias. For consumers of group two, the expected payment for satiated consumption under the measured tariff is approximately .839 which is lower than the basic charge of the flat rate. Consumers from group two pay a premium in order to be insured against variable payments that arise under the measured tariff. ■

7.2. Competition

Now, we briefly investigate the effect of competition on the profit-maximizing two-part tariff.[34] Analyzing the impact of competition on the structure of the optimal contract is insightful, because other findings regarding firms' pricing strategies when facing loss-averse consumers crucially depend on the degree of competition.[35] Competition takes place at the contracting stage and affects a consumer's outside option. With competition, the outside option does not yield zero expected utility but rather $\bar{u} > 0$, the expected utility arising from the best alternative offer. Competing firms offer alternative two-part tariffs for the good, which increases $\bar{u}$. A higher degree of competition corresponds to a higher $\bar{u}$. Since the outside option is only a positive constant in the participation constraint of the firm's optimization problem, the optimal unit price p^* does not depend on the degree of competition. Competition only affects the basic charge L, which decreases if competition becomes more intense. In the limit—under perfect competition—L is determined by a zero profit condition.

RESULT 2. *Suppose that Assumption 2 holds and consider a perfectly competitive market with homogeneous consumers. The equilibrium contract (p^*, L^*) is a flat-rate tariff if and only if $\Sigma(\lambda) \geq c$, with $p^* = 0$ and $L^* = c\int_{\underline{\theta}}^{\bar{\theta}} q^S(\theta) f(\theta) d\theta$.*

34. A formal model of imperfect competition is analyzed in the Web Appendix.

35. See, for instance, Heidhues and Kőszegi (2008) or Heidhues and Kőszegi (2010b).

7.3. Loss Aversion in Both Dimensions

In this section, we assume that the consumers have reference-dependent preferences in the good dimension and in the money dimension. We posit that a consumer is disappointed if his intrinsic utility from consumption, $u(q,\theta)$, falls short of his reference point. The universal loss function for both dimensions is $\mu(x) = \lambda x$, for $x > 0$ and zero otherwise. Moreover, we modify Assumption 1 by positing that the intrinsic utility for the good, evaluated at the satiation point, is constant:[36]

$$\forall\, \theta \in \Theta \qquad u(q^S(\theta),\theta) \equiv \bar{u}. \tag{15}$$

This implies that higher demand types achieve lower utility levels, i.e., $\partial_\theta \hat{u}(q,\theta) < 0$ for $q < q^S(\theta)$. The rest of Assumption 1 remains unchanged. In particular, higher types still have stronger preferences for the good and the satiation point is strictly increasing in the demand type. The utility function $\hat{u}(q,\theta) = \theta q - (1/2)q^2 - (1/2)\theta^2 + \chi = -(1/2)[q^S(\theta) - q]^2 + \chi$, where $q^S(\theta) = \theta$ and χ is chosen such that $\mathbb{E}_\theta[u(0,\theta)] = 0$, is of the form described above.

Intuitively, (15) means that the consumer's utility does not depend on his absolute consumption level, but rather on the difference between his ideal and his actual consumption level. For instance, a customer of a telephone service provider may want to make an uncertain number of calls in each billing period. If he can make all calls he wants to, he achieves a constant level of happiness, independent of the actual number of calls. Here, a higher θ corresponds to an inferior demand type, a state where the consumer has to make many telephone calls and suffers a lot if he cannot do so.

Suppose that the consumer accepted a two-part tariff (p,L) and that his expected demand is $\hat{q}(\cdot,p)$. If demand type θ is realized and the consumer demands quantity q, then he feels a loss in the money dimension compared to the demand types in the set $X(q) = \{z \in \Theta | q > \hat{q}(z,p)\}$. Similarly, the set of demand types compared to which the consumer feels a loss in the good dimension is denoted by $Y(q,\theta) = \{z \in \Theta | u(q,\theta) < u(\hat{q}(z,p),z)\}$. Obviously X is increasing in q while Y is decreasing in q. Moreover, Y is increasing in θ. If demand type θ is realized and the consumer demands quantity q, then his utility is

$$\begin{aligned} U(q|\theta,\langle\hat{q}(\theta,p)\rangle) = u(q,\theta) - pq - L - \lambda \int_{Y(q,\theta)} \big[u(\hat{q}(z,p),z) - u(q,\theta)\big] f(z)\, dz \\ - \lambda \int_{X(q)} p\big[q - \hat{q}(z,p)\big] f(z)\, dz. \end{aligned} \tag{16}$$

The analysis of losses in both dimensions is more challenging because the set Y depends on the demand type θ. It is intricate to narrow down the set of potential personal equilibria or to show uniqueness.[37] Since we are mainly interested in the

36. Without loss aversion in the good dimension, this normalization does not affect consumer behavior.

37. In a slightly different setting with a binary type space, Hahn et al. (2010) provide an example for a personal equilibrium in which the high type consumes less than the low type.

robustness of our findings, we construct only one personal equilibrium demand function with reasonable properties. The demand function should be continuous in θ, since the utility function, the type space, and the two-part tariff are continuous. Additionally, we posit that (i) demand is strictly increasing in the type, and (ii) higher types do not consume so much more that they achieve a higher intrinsic utility than lower types. Given that these two presumptions hold, the personal equilibrium is characterized by

$$\partial_q u(\hat{q}(\theta, p), \theta) = p. \tag{17}$$

The demand function characterized by (17) is increasing in the demand type. Whether the second hypothesis, that higher types achieve lower utility levels in equilibrium, is also satisfied depends on the utility function. We restrict attention to the case that $u(q, \theta)$ fulfills this hypothesis.

Given that the personal equilibrium is characterized by (17), a flat rate is optimal under the same conditions as in the case without losses in the good dimension.

PROPOSITION 6. *Suppose that Assumption 2 holds and that the personal equilibrium demand function is characterized by* (17). *Then, the monopolist optimally offers a flat-rate tariff if and only if* $\Sigma(\lambda) \geq c$.

In contrast to the case of loss aversion only in the money dimension, the demand function is independent of the degree of loss aversion and thus $\Sigma(\lambda)$ is linear in λ.

EXAMPLE 4. Let the demand type be uniformly distributed $\theta \sim U[\mu - \sigma, \mu + \sigma]$ with $\mu - \sigma > 0$. Suppose that intrinsic utility is quadratic and given by $\hat{u}(q, \theta) = \theta q - (1/2)q^2 - (1/2)\theta^2 + \chi$, where $\chi = (1/6)[3\mu^2 + \sigma^2]$. The personal equilibrium demand function defined by equation (17) is

$$\hat{q}(\theta, p) = \theta - p. \tag{18}$$

It can easily be verified that the demand function (18) indeed constitutes a personal equilibrium. In this personal equilibrium, the intrinsic utility in the good dimension equals $u(\hat{q}(\theta, p), \theta) = \chi - (1/2)p^2$ and is independent of the demand type. As long as all types demand a positive quantity, the joint surplus is strictly concave without additional assumptions on the degree of loss aversion. Hence, a flat rate is optimal when $S'(p)|_{p=0} \leq 0$, which is equivalent to

$$c \leq (1/3)\lambda\sigma = \Sigma(\lambda)\,. \tag{19}$$

As in the main part of the paper, the optimal two-part tariff is a flat rate when (i) the marginal cost of production is low, (ii) the consumers are loss averse, and (iii) there is enough variation in demand. ■

8. Conclusion

We developed a model of firm pricing and consumer choice, where consumers are loss averse and uncertain about their own future demand. We showed that loss-averse consumers are biased in favor of flat-rate contracts: a loss-averse consumer may prefer a flat-rate contract to a measured tariff before learning his preferences even though the expected consumption would be cheaper with the measured tariff than with the flat rate. Moreover, we analyzed the optimal pricing strategy of a monopolist when consumers are loss averse. The optimal two-part tariff is a flat-rate contract if marginal costs are low and if consumers value the insurance provided by the flat-rate contract sufficiently. A flat-rate contract insures a loss-averse consumer against fluctuations in his billing amount, and this insurance is particularly valuable if loss aversion is intense or demand is highly uncertain. Thus, this paper provides one possible explanation for the prevalence of flat-rate tariffs. If the contract is not restricted to the class of two-part tariffs, the optimal tariff is not fully flat. The optimal general tariff has a large flat part for intermediate quantities if marginal costs are low and demand of the loss-averse consumers is highly uncertain.

Appendix

Preliminary Considerations. Without the assumption of a strictly increasing personal equilibrium, the cutoff type $\alpha(q)$ is given by $\alpha(q) = \inf\{\theta | \hat{q}(\theta,p) \geq q\}$. With this new definition, the consumer's utility function at the consumption stage is formally unchanged:

$$U(q|\theta,\langle\hat{q}(\varphi,p)\rangle) = u(q,\theta) - pq - L - \lambda p \int_{\underline{\theta}}^{\alpha(q)} [q - \hat{q}(\varphi,p)] f(\varphi) d\varphi.$$

We need two lemmas to prove Proposition 1.

LEMMA A.3. *Let $\langle\hat{q}(\theta,p)\rangle$ be a personal equilibrium. Then $\hat{q}$ is a continuous function of θ.*

Proof. The utility function $U(q|\theta,\langle\hat{q}(\varphi,p)\rangle)$ is continuous in (q,θ). By Assumption 1, it is strictly concave in q for $q \leq q^S(\theta)$ and strictly decreasing for $q \geq q^S(\theta)$. Hence it has a unique maximum which is continuous in θ by the theorem of the maximum.[38] □

By continuity of $\hat{q}$, α is strictly increasing for $q \in (\hat{q}(\underline{\theta},p), \hat{q}(\bar{\theta},p))$.

38. We thank an anonymous referee for suggesting this proof.

LEMMA A.4. *Let $\langle \hat{q}(\theta,p)\rangle$ be a personal equilibrium. Suppose that $\hat{q}(\theta,p)=\bar{q}$ for all θ in a maximal interval $I=[a,b]$. Then $\tilde{q}(b,p)\leq \bar{q}\leq \tilde{q}(a,p)$.*

Proof. Fix $\theta \in (a,b)$. Since $U(q|\theta,\langle \hat{q}(\varphi,p)\rangle)$ is concave it has left and right derivatives at $\bar{q}$ which are given by

$$\begin{aligned} &\partial_q^- U(\bar{q}|\theta,\langle \hat{q}(\varphi,p)\rangle) = \partial_q u(\bar{q},\theta) - p(1+\lambda F(a)),\\ \text{and}\qquad &\partial_q^+ U(\bar{q}|\theta,\langle \hat{q}(\varphi,p)\rangle) = \partial_q u(\bar{q},\theta) - p(1+\lambda F(b)). \end{aligned}$$

Suppose by contradiction that $\bar{q}>\tilde{q}(a,p)$. Then, by concavity of $u(q,\theta)$ as a function of q,

$$\partial_q u(\bar{q},a) - p(1+\lambda F(a)) < \partial_q u(\tilde{q}(a,p),a) - p(1+\lambda F(a)) = 0.$$

By Assumption 1, $\partial_q^- U(\bar{q}|\theta,\langle \hat{q}(\varphi,p)\rangle) \approx \partial_q u(\bar{q},a) - (1+\lambda F(a))$ for θ close to a. Hence, all types θ close to a strictly prefer to consume less than $\bar{q}$. This is a contradiction. Supposing that $\bar{q}<\tilde{q}(b,p)$ leads to a similar contradiction. □

Proof of Proposition 1. If Condition 1 holds, $\tilde{q}(\theta,p)$ is strictly increasing in θ. Therefore, by Lemma A.4, every personal equilibrium $\hat{q}(\theta,p)$ is strictly increasing in θ. Hence, as argued in the text, $\hat{q}(\theta,p)$ must satisfy (6) for almost every θ. By continuity of $\tilde{q}(\theta,p)$ (by Assumption 1) and $\hat{q}(\theta,p)$ (by Lemma A.3), we have $\hat{q}(\theta,p)=\tilde{q}(\theta,p)$. Since $U(q|\theta,\langle \hat{q}(\varphi,p)\rangle)$ is strictly concave, the first-order condition (6) is also sufficient, which proves existence. □

Proof of Lemma 1. From (6) and the expressions for the left and right derivatives of $U(q|\theta,\langle q^{\mathrm{PPE}}(\varphi,p)\rangle)$ (see the proof of Lemma A.4), it follows that $q^{\mathrm{PPE}}(\theta,p)$ is a personal equilibrium. To show that it is the preferred personal equilibrium, suppose that $\hat{q}(\theta,p)$ is a personal equilibrium that differs from $q^{\mathrm{PPE}}(\theta,p)$. By definition of $q^{\mathrm{PPE}}(\theta,p)$, we have $\hat{q}(\theta,p)\leq q^{\mathrm{PPE}}(\theta,p)$ for all θ with strict inequality for some θ.

Now let a denote the highest type for which $\hat{q}$ coincides with q^{PPE} on the whole interval $(\underline{\theta},a]$, i.e., $a := \max\left\{\theta | \forall s\in[\underline{\theta},\theta] : \hat{q}(s,p)=q^{\mathrm{PPE}}(s,p)\right\}$ if $\hat{q}(\underline{\theta},p) = q^{\mathrm{PPE}}(\underline{\theta},p)$, and $a=\underline{\theta}$, otherwise. a is the left endpoint of a flat part of $\hat{q}$. Let c be the right endpoint of the flat part, $c:=\max\{\theta|\hat{q}(\theta,p)=\hat{q}(a,p)\}$.

We want to increase $\hat{q}$ on $[a,c]$. This will extend the flat part to the right of c unless $c=\bar{\theta}$. But we only want to increase so much that we do not merge with another flat part. Hence, if $c<\bar{\theta}$ the new flat part must end at $d\leq \bar{d} := \max\{\theta\,|\,\forall\varphi\in[c,\theta] : \hat{q}(\varphi,p)=\tilde{q}(\varphi,p)\}$, and we set $\bar{d}=\bar{\theta}$ if $c=\bar{\theta}$. The lower end of the flat part will move to some type $b\geq a$ and $\tilde{q}$ must be non-decreasing on $[a,b]$. Hence, $b\leq \bar{b} := \max\{\theta|\tilde{q}$ non-decreasing on $[a,\theta]\}$.

We now construct a new personal equilibrium $\check{q}(\theta,p)$ by increasing the flat part of $\hat{q}(\theta,p)$ to $\bar{q} := \min\left\{\tilde{q}(\bar{b},p),\tilde{q}(\bar{d},p)\right\}$ if $\bar{d}<\bar{\theta}$ and to $\bar{q} := \tilde{q}(\bar{b},p)$ otherwise.

Therefore, we have $b = \min\{\theta \geq a | \tilde{q}(\theta,p) = \bar{q}\}$ if $\bar{b} > a$ and $b = 0$ otherwise; and $d = \min\{\theta > c | \tilde{q}(\theta,p) = \bar{q}\}$ if $\bar{d} < \bar{\theta}$ and $d = \bar{\theta}$ otherwise. The new personal equilibrium is given by $\check{q}(\theta,p) = \hat{q}(\theta,p)$ if $\theta \notin [a,d]$, $\check{q}(\theta,p) = \tilde{q}(\theta,p)$ if $\theta \in [a,b]$, and $\check{q}(\theta,p) = \bar{q}$ if $\theta \in [b,d]$.

Now we show that the utility of type θ in the new equilibrium, $\check{V}(\theta)$, exceeds the utility $\hat{V}(\theta)$ in the old equilibrium. For $\theta \in (a,b]$ the loss under the new equilibrium is

$$\check{\mathscr{L}}(\theta) = p\int_{\underline{\theta}}^{a} (\tilde{q}(\theta) - \hat{q}(\varphi))f(\varphi)d\varphi + p\int_{a}^{\theta} (\tilde{q}(\theta) - \tilde{q}(\varphi))f(\varphi)d\varphi.$$

Inserting $\tilde{q}(\theta) = \hat{q}(\theta) + (\tilde{q}(\theta) - \hat{q}(\theta))$ in the first integral and using $\tilde{q}(\varphi) > \hat{q}(\theta)$ in the second we get

$$\check{\mathscr{L}}(\theta) < p\int_{\underline{\theta}}^{a} (\hat{q}(\theta) - \hat{q}(\varphi))f(\varphi)d\varphi + p(\tilde{q}(\theta) - \hat{q}(\theta))F(\theta).$$

Hence, $\check{\mathscr{L}}(\theta) < \hat{\mathscr{L}}(\theta) + p(\tilde{q}(\theta) - \hat{q}(\theta))F(\theta)$, and thus

$$\begin{aligned}\check{V}(\theta) - \hat{V}(\theta) &= u(\tilde{q}(\theta),\theta) - u(\hat{q}(\theta),\theta) - p(\tilde{q}(\theta) - \hat{q}(\theta)) - p\lambda(\check{\mathscr{L}}(\theta) - \hat{\mathscr{L}}(\theta)) \\ &\geq \int_{\hat{q}(\theta)}^{\tilde{q}(\theta)} \{\partial_q u(q,\theta) - p(1+\lambda F(\theta))\}\,dq > 0,\end{aligned}$$

where the last inequality follows from $\partial_q u(q,\theta) - p(1+\lambda F(\theta)) > \partial_q u(\tilde{q}(\theta),\theta) - p(1+\lambda F(\theta)) = 0$ for $q < \tilde{q}(\theta)$. Since $\hat{\mathscr{L}}(\theta)$ and $\check{\mathscr{L}}(\theta)$ are constant on $(b,c]$, and $\check{V}(b) \geq \hat{V}(b)$, $\partial_{\theta q} u > 0$ implies that $\check{V}(\theta) > \hat{V}(\theta)$ for all $\theta \in (b,c]$. Finally, for $\theta \in (c,d]$ we have

$$\check{\mathscr{L}}(\theta) = \int_{\underline{\theta}}^{\theta} (\bar{q} - \check{q}(\varphi))\,f(\varphi)d(\varphi) \leq \int_{\underline{\theta}}^{b} (\bar{q} - \hat{q}(\varphi))\,f(\varphi)d(\varphi)$$

and $$\hat{\mathscr{L}}(\theta) = \int_{\underline{\theta}}^{\theta} (\tilde{q}(\theta) - \hat{q}(\varphi))\,f(\varphi)d(\varphi) \geq \int_{\underline{\theta}}^{b} (\tilde{q}(\theta) - \hat{q}(\varphi))\,f(\varphi)d(\varphi).$$

Hence,

$$\check{V}(\theta) - \hat{V}(\theta) \geq \int_{\tilde{q}(\theta)}^{\bar{q}} \{\partial_q u(q,\theta) - p(1+\lambda F(b))\}\,dq > 0,$$

because $\bar{q} = \tilde{q}(b)$ and for $q < \bar{q}$, $\partial_q u(q,\theta) - p(1+\lambda F(b)) > \partial_q u(\tilde{q}(b),\theta) - p(1+\lambda F(b)) = 0$ by $\partial_{qq} u < 0$. This shows that $\hat{q}$ is not the preferred personal equilibrium. □

Proof of Proposition 2. Let $(L^F,0)$ be a flat-rate tariff and let (L,p) with $0 < p < \bar{p}$ be a measured tariff. The measured tariff leads to lower expenditures than the flat rate if

$$L^F - L - p\int_{\underline{\theta}}^{\bar{\theta}} q^S(\theta)f(\theta)\,d\theta > 0. \tag{A.1}$$

The consumer prefers the flat-rate tariff ex ante if

$$\int_{\underline{\theta}}^{\bar{\theta}}[u(\hat{q}(\theta,p),\theta)-p\hat{q}(\theta,p)]f(\theta)\,d\theta-L-\lambda p\chi(p)<\int_{\underline{\theta}}^{\bar{\theta}}u(q^S(\theta),\theta)f(\theta)\,d\theta-L^F, \tag{A.2}$$

$$\text{where} \qquad \chi(p)=\int_{\underline{\theta}}^{\bar{\theta}}\int_{\underline{\theta}}^{\theta}[\hat{q}(\theta,p)-\hat{q}(\varphi,p)]f(\varphi)f(\theta)\,d\varphi d\theta>0. \tag{A.3}$$

Equation (A.2) can be rearranged to

$$\begin{aligned}&\frac{1}{p}\left(L^F-L-p\int_{\underline{\theta}}^{\bar{\theta}}q^S(\theta)f(\theta)\,d\theta\right)<\lambda\chi(p)+\\ &\frac{1}{p}\int_{\underline{\theta}}^{\bar{\theta}}[u(q^S(\theta),\theta)-u(\hat{q}(\theta,p),\theta)]f(\theta)\,d\theta-\int_{\underline{\theta}}^{\bar{\theta}}[q^S(\theta)-\hat{q}(\theta,p)]f(\theta)\,d\theta.\end{aligned} \tag{A.4}$$

Let L be such that the expenditure savings from the measured tariff are $p\varepsilon$ with $\varepsilon>0$, i.e., $L^F-L-p\int_{\underline{\theta}}^{\bar{\theta}}q^S(\theta)f(\theta)\,d\theta=p\varepsilon$. By construction, for all $p>0$ and $\varepsilon>0$ the consumer would save money with the measured tariff. Thus, if we can always find positive values for p and ε such that (A.4) is fulfilled, then the consumer's preferences are flat-rate biased. The left-hand side of inequality (A.4) equals ε by definition. For $p\to 0$, the right-hand side is at least as great as $\lambda\int_{\underline{\theta}}^{\bar{\theta}}\int_{\underline{\theta}}^{\theta}[q^S(\theta)-q^S(\varphi)]f(\varphi)f(\theta)\,d\varphi d\theta>0$. Thus, we can always find $p>0$ and $\varepsilon>0$—both sufficiently small—such that (A.4) holds if $\lambda>0$. □

Proof of Proposition 3. First, we show for $\hat{q}=q^{\mathrm{PPE}}$, that $p^*=0$ if and only if $S'(0)\leq 0$. By the definition of q^{PPE}, we have $\hat{q}(\theta,p)=\tilde{q}(t(\theta,p),p)$, where $t(\theta,p)=\max\{\arg\max_{\varphi\leq\theta}\tilde{q}(\varphi,p)\}$. Using this, the marginal surplus can be written as

$$\begin{aligned}S'(p)=\int_{\underline{\theta}}^{\bar{\theta}}\Big\{(\partial_q u(\tilde{q}(t(\theta,p),p),\theta)-c)\,\partial_q\hat{q}(\theta,p)-\lambda\int_{\underline{\theta}}^{\theta}(\hat{q}(\theta,p)-\hat{q}(\varphi,p))\,f(\varphi)d\varphi\\ -\lambda p\int_{\underline{\theta}}^{t(\theta,p)}(\partial_q\hat{q}(\theta,p)-\partial_p\hat{q}(\varphi,p))\,f(\varphi)d\varphi\Big\}f(\theta)d(\theta).\end{aligned}$$

Using $\partial_q u(\tilde{q}(t(\theta,p),p),\theta)=p(1+\lambda F(t(\theta,p)))$ we get (9).

Our assumptions guarantee that $\partial_p\tilde{q}(\theta,p)\leq\varepsilon<0$ and

$$\partial_{pp}\tilde{q}(\theta,p)=-\frac{(1+\lambda F(\theta))\,\partial_{qqq}u(\hat{q}(\theta,p),\theta)\,\partial_p\hat{q}(\theta,p)}{(\partial_{qq}u(\hat{q}(\theta,p),\theta))^2}\geq 0.$$

As the maximum of decreasing and convex functions, $\hat{q}$ is also strictly decreasing and convex as a function of p . This implies that $S'(p)<0$ for $p>c$. Clearly $S'(p)=\Psi(p)$ at $p=0$ and $S'(p)<\Psi(p)$ for $p>0$ and $\lambda>0$. If $\Psi(p)$ is non-increasing for $p\in[0,c]$, we can conclude that (i) $S'(0)\leq 0$ implies that $S'(p)<0$ for $p\in(0,c]$, and (ii) $S'(0)>0$ implies that there exist a $p^*\in(0,c]$ that maximizes $S(p)$ and this price satisfies $S'(p^*)=0$. It remains to show that $\Psi(p)$ is indeed non-increasing for

$p \in [0,c]$. By taking the derivative of $\Psi(p)$ with respect to p and using integration by parts, we obtain

$$\Psi'(p) = \int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta,p)[1+\lambda-2\lambda F(\theta)] f(\theta) d\theta + (p-c)\int_{\underline{\theta}}^{\bar{\theta}} \partial_{pp}\hat{q}(\theta,p) f(\theta) d\theta. \tag{A.5}$$

Thus, $\Psi'(p) \leq 0$ for $p \in [0,c]$. The first term is negative because $\lambda \leq 1$ and $\hat{q}$ is strictly decreasing. The second term is non-positive because $\hat{q}$ is convex.

Next, we show that $S'(0) \leq 0$ is equivalent to $\Sigma(\lambda) \geq c$. By evaluating (9) at $p=0$, it is obvious that $S'(0) \leq 0$ if and only if

$$-c\int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta,0) f(\theta) d\theta - \lambda \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [\hat{q}(\varphi,0)-\hat{q}(\theta,0)] f(\varphi) f(\theta)\, d\varphi d\theta \leq 0\,.$$

Rearranging this inequality yields $c \leq \Sigma(\lambda)$.

Finally, we show that $\Sigma'(\lambda) > 0$.[39] To simplify notation, we define $Z(\lambda)$ and $N(\lambda)$ as the numerator and the denominator, respectively, of the fraction of $\Sigma(\cdot)$. Thus,

$$Z(\lambda) \equiv \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,0)-\hat{q}(\varphi,0)] f(\varphi) f(\theta) d\varphi d\theta, \tag{A.6}$$

and

$$N(\lambda) \equiv -\int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta,0) f(\theta) d\theta\,. \tag{A.7}$$

Observe that $Z(\lambda) > 0$, since $\hat{q}$ is increasing in θ. $N(\lambda) > 0$ follows from

$$\partial_p \hat{q}(\theta,p) = \frac{1+\lambda F(\theta)}{\partial_{qq} u(\hat{q}(\theta,p),\theta)} < 0 \tag{A.8}$$

With this notation the derivative of $\Sigma(\cdot)$ with respect to λ can be written as

$$\Sigma'(\lambda) = \frac{Z(\lambda)}{N(\lambda)} + \lambda \frac{Z'(\lambda)N(\lambda) - N'(\lambda)Z(\lambda)}{[N(\lambda)]^2}. \tag{A.9}$$

In order to show that $\Sigma'(\lambda) > 0$, we analyze the parts separately. Since $\partial_\lambda \hat{q}(\theta,0) = 0$, $Z'(\lambda) = 0$. Taking the derivative of (A.8) with respect to λ, and using $\partial_\lambda \hat{q}(\theta,0) = 0$, yields $\partial_{\lambda p}\hat{q}(\theta,0) = F(\theta)/\partial_{qq}u(\hat{q}(\theta,0),\theta)\ (<0)$. Thus,

$$N'(\lambda) = -\int_{\underline{\theta}}^{\bar{\theta}} \frac{F(\theta)}{\partial_{qq} u(\hat{q}(\theta,0),\theta)} f(\theta) d\theta\,.$$

Since $Z'(\lambda) = 0$, equation (A.9) simplifies to

$$\Sigma'(\lambda) = \frac{Z(\lambda)}{[N(\lambda)]^2}\left[N(\lambda) - \lambda N'(\lambda)\right].$$

39. For $p=0$ there is a unique and increasing personal equilibrium. Hence, $\hat{q}(\theta,0) = \tilde{q}(\theta,0)$.

Since $Z(\lambda) > 0$, it remains to show that $N(\lambda) - \lambda N'(\lambda) > 0$, which is equivalent to

$$-\int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta,0) f(\theta) d\theta + \lambda \int_{\underline{\theta}}^{\bar{\theta}} \frac{F(\theta)}{\partial_{qq} u(\hat{q}(\theta,0),\theta)} f(\theta) d\theta > 0\,,$$

$$\iff \qquad \int_{\underline{\theta}}^{\bar{\theta}} -\frac{f(\theta)}{\partial_{qq} u(\hat{q}(\theta,0),\theta)} d\theta > 0.$$

The last inequality is satisfied since $u(\cdot)$ is a strictly concave function in q for $q \leq q^S(\theta)$. $\square$

References

Abeler, Johannes, Armin Falk, Lorenz Götte, and David Huffman (2011). "Reference Points and Effort Provision." *American Economic Review*, 101(2), 470–492.

Berseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum (2010). "The Nature of Risk Preferences: Evidence from Insurance Choices." Working paper, Cornell University.

Camerer, Colin F. (2004). "Prospect Theory in the Wild: Evidence from the Field." In *Advances in Behavioural Economics*, edited by Colin F. Camerer, George Loewenstein, and Matthew Rabin, chap. 5, pp. 148–161. Princeton University Press, Princeton, New Jersey.

Carbajal, Juan Carlos and Jeffrey Ely (2010). "Mechanism Design without Revenue Equivalence." Working paper, Northwestern University.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv (2007). "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review*, 97(5), 1921–1938.

Clay, Karen B., David S. Sibley, and Padmanabhan Srinagesh (1992). "Ex Post vs. Ex Ante Pricing: Optimal Calling Plans and Tapered Tariffs." *Journal of Regulatory Economics*, 4(2), 115–138.

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

Crawford, Vincent P. and Juanjuan Meng (forthcoming). "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income." *American Economic Review*.

DellaVigna, Stefano and Ulrike Malmendier (2004). "Contract Design and Self-Control: Theory and Evidence." *Quarterly Journal of Economics*, 119(2), 353–402.

DellaVigna, Stefano and Ulrike Malmendier (2006). "Paying Not to Go to the Gym." *American Economic Review*, 96(3), 694–719.

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

Eliaz, Kfir and Ran Spiegler (2008). "Consumer Optimism and Price Discrimination." *Theoretical Economics*, 3(4), 459–497.

Ericson, Keith M. M. and Andreas Fuster (forthcoming). “Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments.” *Quarterly Journal of Economics*.

Esteban, Susanna and Eiichi Miyagawa (2006). “Temptation, Self-Control, and Competitive Nonlinear Pricing.” *Economics Letters*, 90, 348–355.

Esteban, Susanna, Eiichi Miyagawa, and Matthew Shum (2007). “Nonlinear Pricing with Self-Control Preferences.” *Journal of Economic Theory*, 135, 306–338.

Faulhaber, Gerald R. and Christiaan Hogendorn (2000). “The Market Structure of Broadband Telecommunications.” *Journal of Industrial Economics*, 48(3), 305–329.

Gill, David and Victoria L. Prowse (forthcoming). “A Structural Analysis of Disappointment Aversion in a Real Effort Competition.” *American Economic Review*.

Grubb, Michael D. (2009). “Selling to Overconfident Consumers.” *American Economic Review*, 99, 1770–1807.

Gul, Faruk and Wolfgang Pesendorfer (2001). “Temptation and Self Control.” *Econometrica*, 60, 1293–1330.

Hahn, Jong-Hee, Jinwoo Kim, Sang-Hyun Kim, and Jihong Lee (2010). “Screening Loss Averse Consumers.” Working Paper, Yonsei University.

Heidhues, Paul and Botond Kőszegi (2008). “Competition and Price Variation when Consumers are Loss Averse.” *American Economic Review*, 98, 1245–1268.

Heidhues, Paul and Botond Kőszegi (2010a). “Exploiting Naivete about Self-Control in the Credit Market.” *American Economic Review*, 100(5), 2279–2303.

Heidhues, Paul and Botond Kőszegi (2010b). “Regular Prices and Sales.” Working paper, University of California, Berkeley.

Herweg, Fabian, Daniel Müller, and Philipp Weinschenk (2010). “Binary Payment Schemes: Moral Hazard and Loss Aversion.” *American Economic Review*, 100(5), 2451–2477.

Hobson, Michael and Richard H. Spady (1988). “The Demand for Local Telephone Service Under Optimal Local Measured Service.” Bellcore Economics Discussion Paper, No. 50.

Kahneman, Daniel and Amos Tversky (1979). “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica*, 47(2), 263–292.

Karle, Heiko and Martin Peitz (2010a). “Consumer Loss Aversion and the Intensity of Competition.” Working Paper, Université Libre de Bruxelles.

Karle, Heiko and Martin Peitz (2010b). “Pricing and Information Disclosure in Markets with Loss-Averse Consumers.” CEPR Discussion Paper, No. 7785.

Kőszegi, Botond and Matthew Rabin (2006). “A Model of Reference-Dependent Preferences.” *Quarterly Journal of Economics*, 121(4), 1133–1165.

Kőszegi, Botond and Matthew Rabin (2007). “Reference-Dependent Risk Attitudes.” *American Economic Review*, 97(4), 1047–1073.

Kling, John P. and Stephen S. van der Ploeg (1990). “Estimating Local Elasticities with a Model of Stochastic Class of Service and Usage.” In *Telecommunications Demand Modeling: An Integrated View*, edited by Alain de Fontenay, Mary H.

Shugard, and David S. Sibley, pp. 119–136. North Holland, Amsterdam.

Kridel, Donald J., Dale E. Lehman, and Dennis L. Weisman (1993). "Option Value, Telecommunication Demand, and Policy." *Information Economics and Policy*, 5(2), 125–144.

Lambrecht, Anja, Katja Seim, and Bernd Skiera (2007). "Does Uncertainty Matter? Consumer Behavior Under Three-Part Tariffs." *Marketing Science*, 26(5), 698–710.

Lambrecht, Anja and Bernd Skiera (2006). "Paying Too Much and Being Happy About It: Existence, Causes, and Consequences of Tariff-Choice Biases." *Journal of Marketing Research*, 63(3), 212–223.

Macera, Rosario (2009). "Intertemporal Incentives under Loss Aversion." Working paper, University of California, Berkeley.

Miravete, Eugenio J. (2002). "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans." *Review of Economic Studies*, 69, 943–971.

Miravete, Eugenio J. (2003). "Choosing the Wrong Calling Plan? Ignorance and Learning." *American Economic Review*, 93(1), 297–310.

Mitchell, Bridger M. and Ingo Vogelsang (1991). *Telecommunications Pricing: Theory and Practice*. Cambridge University Press, Cambridge.

Ng, Yew-Kwang and Mendel Weissner (1974). "Optimal Pricing with a Budget Constraint: The Case of Two-Part Tariffs." *Review of Economic Studies*, 41, 337–345.

Nunes, Joseph C. (2000). "A Cognitive Model of People's Usage Estimations." *Journal of Marketing Research*, 37, 397–409.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler (2008). "Deal Or No Deal? Decision Making Under Risk in a Large-Payoff Game Show." *American Economic Review*, 98(1), 38–71.

Prelec, Drazen and George Loewenstein (1998). "The Red and the Black: Mental Accounting of Savings and Dept." *Marketing Science*, 17, 4–28.

Spiegler, Ran (forthcoming). "Monopoly Pricing when Consumers are Antagonized by Unexpected Price Increases: A "Cover Version" of the Heidhues-Kőszegi-Rabin Model." *Economic Theory*.

Sundararajan, Arun (2004). "Nonlinear Pricing of Information Goods." *Management Science*, 50(12), 1660–1673.

Thaler, Richard H. (1999). "Mental Accounting Matters." *Journal of Behavioral Decision Making*, 12, 183–206.

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

Train, Kenneth E., Moshe Ben-Akiva, and Terry Atherton (1989). "Consumption Patterns and Self-Selecting Tariffs." *Review of Economics and Statistics*, 71(1), 62–73.

Train, Kenneth E., Daniel L. McFadden, and Moshe Ben-Akiva (1987). "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choice." *RAND Journal of Economics*, 18(1), 109–123.

Uthemann, Andreas (2005). "Competitive Screening of Customers with Non-Common Priors." Working paper, University College London.