University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN1664-705X(online)

Working Paper No. 12

# Uncertain Demand, Consumer Loss Aversion, and Flat-Rate Tariffs

Fabian Herweg and Konrad Mierendorff

March 2011

Uncertain Demand, Consumer Loss Aversion, and Flat-Rate Tariffs*

Fabian Herweg† and Konrad Mierendorff‡

March 23, 2011

**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 classification: D11; D43; L11*

*Keywords: Consumer Loss Aversion; Flat-Rate Tariffs; Nonlinear Pricing; Uncertain Demand*

## 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 zero, e.g. for rental cars or telephone services.[1] For rental cars, a typical contract has a fixed price per day that

*In preparing this paper, we have greatly benefited from comments made by 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), the IMEBE at Bilbao (2010), the SFB Meeting at Caputh (2010), the WZB Conference Biases in Markets at Berlin (2010), as well as seminar participants at the University of California—Berkeley and the University of Zurich. 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. All remaining errors are our own responsibility. No-one else can be blamed for any shortcomings.

†University of Bonn, Department of Economics, Chair of Economic Theory II, Adenauerallee 24-42, D-53113, Bonn, Germany, E-mail address: fherweg@uni-bonn.de (corresponding author)

‡University of Zurich, Department of Economics, Chair of Organizational Design, Blümlisalpstrasse 10, CH-8006 Zürich, Switzerland, E-mail address: mierendorff@iew.uzh.ch

[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.

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 expenditure 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 the authors carefully expose, this misprediction could be caused by naïve quasi-hyperbolic discounting. DellaVigna and Malmendier (2004) show that with naïve 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, naïve 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.

[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 wear repairings 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 for a flat-rate bias for telephone services.

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 consumer.

In order to model the insurance motive, we posit that consumers are 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 rental car company 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 amounts, which he forms before accepting the contract.[9] Following Kőszegi and Rabin (2006, 2007), we assume that the reference point is the full distribution of possible billing amounts, and that the consumer's expectations are a 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

[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. (forthcoming), Post et al. (2008), Crawford and Meng (forthcoming), Gill and Prowse (forthcoming), and Ericson and Fuster (2010) provide evidence that reference points are determined by expectations. See Berseghyan et al. (2010) for a paper that does not find evidence for expectation-based loss aversion.

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 contact 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]

We extend our analysis and solve for the optimal nonlinear tariff, which relaxes the restriction to two-part tariffs. 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 consumer's degree of loss aversion increases, (ii) variations in preferences and thus in demanded quantities increase, and (iii) the marginal cost decreases. The optimal general tariff is increasing at least for some quantities at the extremes of the quantity schedule.

Moreover, focusing on two-part tariffs, we inspect the robustness of our findings. First, we consider heterogeneity among consumers with respect to their degree of loss aversion. We derive conditions such that the monopolist can screen differently loss-averse consumers without costs. In this case, it can 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 if their consumption utility is lower than expected. With

[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 among 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.

this generalization consumers form rational expectations that serve as a reference point not only about the payment but also about the utility from consumption. Under certain assumptions, a flat-rate contract is optimal under the same conditions as in the baseline model.

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 unexpected 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) or Eliaz and Spiegler (2008).

**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 (2010b). 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]

Another strand of the literature analyzes optimal selling strategies for overconfident consumers. The optimal menu of nonlinear tariffs for consumers who underestimate fluctuations in their future demand is analyzed by Grubb (2009). The optimal menu is similar to a menu of three-part tariffs which is common for cellular phone services.[13] Focusing on only two states of the world, Eliaz and Spiegler (2008) solve the tariff design problem of a monopolist who faces consumers with biased beliefs.

Closer to our paper, Hahn et al. (2010) adopt the model of Kőszegi and Rabin (2006, 2007) and analyze the optimal 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 (2010a) 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

[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.

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. The baseline model with two-part tariffs is introduced in Section 3. The demand function of a loss-averse consumer who accepted a two-part tariff is investigated in Section 4. Section 5 derives conditions for a flat-rate tariff to be optimal within the class of two-part tariffs. General nonlinear pricing schemes are analyzed in Section 6 while 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 derivation of the extensions in Sections 6 and 7 are developed in the web appendix.

## 2. Illustrative Example

Consider a monopolist who sells one 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 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. First, we assume that the reference point equals the basic charge, i.e. $r = L$, if the consumer accepts the contract, and zero otherwise. This reference point could be viewed as a forward looking status quo, since the consumer needs to pay the basic charge in any case. 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

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

[15] We restrict $\lambda$ to be less than 1 in order to guarantee concavity of the monopolist's objective function.

optimal unit price $p^*$ maximizes the joint surplus $S(p)$—expected utility plus profits,

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

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. A flat-rate tariff, i.e. $p^* = 0$, is optimal when $S'(p)|_{p=0} \leq 0$, which is equivalent to

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

This example illustrates the basic insight that a preference for flat-rate tariffs can be explained by loss aversion. Moreover, 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 documenting the flat-rate bias points out 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 at the contracting stage, 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 $\alpha$, 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 $q$ to maximize

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

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 $q(6) = 6-p$ and $q(10) = 10 - p(1+\alpha\lambda)$ constitutes a personal equilibrium, as long as $p$ is not too large and thus $q(6) \leq (10)$. 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 $\alpha$. 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 $\alpha$, 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 demanding $q = q(10) = 10 - p(1 + \alpha\lambda)$. If the consumer chooses a quantity $q > q(10)$, additional losses arise from the comparison with the payment in the high demand state:

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

Therefore, the consumer prefers 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 marginal price $p$ in order to maximize the joint surplus including expected loss utility. 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 variations in the billing amount.

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

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

Hence, a flat rate is optimal when three criteria are satisfied: (i) the consumer is loss averse, i.e. $\lambda > 0$ , (ii) the marginal cost is not too high, and (iii) the consumption pattern is sufficiently uncertain, i.e., $\sigma^2$ not too small. The model 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.

A further interesting insight can be obtained by considering the case where the monopolist sells to two ex ante heterogeneous consumers. Suppose that both consumers have identically distributed demand types. One has standard preferences while the other one is loss averse. The degree of loss aversion is private information of the consumers. Suppose further, that under complete information it is optimal to offer a flat-rate tariff to the loss-averse consumer. What is the optimal menu of two-part tariffs if the consumers are privately informed about their degree of loss aversion? The monopolist optimally offers a cost-based tariff and a flat rate as if she could the observe consumers' types. The basic charge of the cost-based tariff is chosen such that the consumer with standard preferences has an expected utility of zero. Therefore, the loss-averse consumer strictly prefers the flat rate. Intuitively, loss aversion reduces the expected utility a consumer obtains from accepting a measured tariff, so that the loss-averse consumer has a negative expected utility from this contract. On the other hand, the consumer with standard preferences weakly prefers the measured tariff. His expected utility equals zero under both contracts because the expected utility from signing a flat-rate contract is independent of a consumer's degree of loss aversion.[16] Thus, the monopolist can screen differently loss-averse consumers at no cost and it is optimal to offer a cost-based tariff next to a flat rate.

## 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 $\theta$. (4) Each consumer who accepted the offer chooses a quantity

[16] In a perfectly competitive market not only the loss-averse consumer strictly prefers the flat rate, but also the standard consumer strictly prefers the cost-based tariff. In a perfectly competitive market the basic charge is determined by a zero profit condition.

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, consumers feel a loss if at the end of the month the invoice from their telephone provider is higher than expected. For the situations we have in mind, it seems natural that the reference point incorporates lagged expectations. Therefore, 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.[17]

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

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:

$$\frac{dq^S(\theta)}{d\theta} = -\frac{\partial_{q\theta} \hat{u}(q^S(\theta), \theta)}{\partial_{qq} \hat{u}(q^S(\theta), \theta)} > 0\,. \tag{1}$$

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.[18] 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.[19]

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 $\phi$ 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

$$U(q|\phi, \langle q(\theta) \rangle) = u(q, \phi) - T(q) - \int_{\underline{\theta}}^{\bar{\theta}} \mu(T(q) - T(q(\theta)))\, f(\theta)\, d\theta\,. \tag{2}$$

Observe that for $p \geq 0$, 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).

[18] This assumption is also imposed by Spiegler (2010). The implications of not assuming a universal gain-loss function for both dimensions are investigated by Karle and Peitz (2010b).

[19] 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. The optimality of flat-rate contracts is driven by the consumer's ex ante preferences and ex ante only net losses matter. We refrain from applying this gain-loss function because it introduces a time inconsistency problem caused by assuming loss aversion only in the money dimension. Without loss aversion ($\tilde{\lambda} = 1$) the marginal utility of money equals $1 + \tilde{\eta}$ ex post, but ex ante, gains and losses cancel in expectation so that the marginal utility of money is $1 < 1 + \tilde{\eta}$. $\lambda$ in our model captures the overall impact of loss aversion *and* reference dependence, and therefore roughly corresponds to $\tilde{\eta}(\tilde{\lambda} - 1)$ in the usual formulation. For a more detailed comparison we refer to Section D of the web appendix.

**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 $\phi \in \Theta$,*

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

## 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, higher types demand strictly higher quantities. For higher unit prices, there are multiple personal equilibria, all of which involve bunches in the demand schedule. Since we are interested in flat rates, we will not analyze these "bunching" personal equilibria.

In Assumption 1, we assumed that 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.[20]

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 $\theta$, he therefore feels losses compared to all types below a cutoff type $\alpha(q)$. $\alpha(q)$ is given by $\hat{q}(\alpha(q), p) = q$.[21] With this notation, the utility of a consumer with demand type $\theta$ who consumes $q$, can be expressed as

$$U(q|\theta, \langle \hat{q}(\phi, p)\rangle) = u(q, \theta) - pq - L - \lambda p \int_{\underline{\theta}}^{\alpha(q)} [q - \hat{q}(\phi, p)] f(\phi)\, d\phi \ . \tag{3}$$

The utility function (3) 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 $\theta$ feels losses compared to all types below $\theta$, i.e. $\alpha(q) = \theta$, we obtain the following condition which characterizes the demand function of a strictly increasing personal equilibrium:

$$\partial_q u(\tilde{q}(\theta, p), \theta) = p[1 + \lambda F(\theta)]. \tag{4}$$

[20] 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 weakly monotone.

[21] Lemma 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.

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 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. 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 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 ($\lambda$ is small), and there is sufficient heterogeneity of preferences measured by $\partial_{q\theta} u(q, \theta)/f(\theta)$.[22]

Define $\bar{p} := \min_\theta \{\kappa/[\lambda f(\theta)]\}$, where $\kappa > 0$ is the lower bound of $\partial_{q\theta} u(q, \theta)$ (cf. Assumption 1). A sufficient condition for (C1) is $p < \bar{p}$. With the quadratic utility function $\hat{u}(q, \theta) = \theta q - (1/2)q^2$ and uniformly distributed types $\theta \sim U[\underline{\theta}, \underline{\theta} + 1]$ we have $\bar{p} = 1/\lambda$. If loss utility is less important than intrinsic utility ($\lambda < 1$), then $\bar{p} > 1$ in this example.

Condition 1 is independent of the marginal cost. A flat rate, however, can only be optimal if marginal costs are not too high. As we will show later, if marginal costs are sufficiently low, then the gains from trade for prices above $\bar{p}$ are lower than the gains from trade generated by a flat rate. Hence, for low marginal costs it is never optimal for the monopolist to set unit prices for which bunching occurs. Since we are interested in situations where it is optimal for firms to offer flat-rate tariffs, we will restrict attention to unit prices for which there is a unique and strictly increasing personal equilibrium.[23]

[22] 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.

[23] It is straight forward to show that there is no strictly increasing personal equilibrium if Condition 1 is violated. Furthermore, a personal equilibrium with bunching consists only of strictly increasing parts that coincide with $\tilde{q}$, and flat parts for which the end points coincide with $\tilde{q}$ (cf. Lemma 4 in the appendix).

*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,

$$\mathbb{E}_\theta[U(\hat{q}(\theta,p)|\theta,\langle\hat{q}(\theta,p)\rangle)] = \int_{\underline{\theta}}^{\bar{\theta}} [u(\hat{q}(\theta,p),\theta) - p\hat{q}(\theta,p)]f(\theta)\, d\theta - L \\ - \lambda p \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,p) - \hat{q}(\phi,p)] f(\phi) f(\theta)\, d\phi d\theta \,. \quad (5)$$

The first term of this expression represents expected intrinsic utility. The second term is the ex ante expected loss, which is weighted by $\lambda$. The expected loss vanishes if the 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 consumers prefer 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 that differ in their degrees of loss aversion and in the distributions of their demand types. The optimal menu of two-part tariffs includes a measured tariff and a flat rate, and we show that the loss-averse consumers exhibit the flat-rate bias. They prefer the flat rate even though the basic charge is greater than the expected payment of satiated consumption under the measured tariff. Hence, when choosing the flat rate, the consumers pay a premium in order to be insured against variations in payments.

## 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.

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

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 loss utility. The optimal unit price $p^*$ maximizes the joint surplus of the two contracting parties which we denote by $S(p)$:

$$S(p) = \int_{\underline{\theta}}^{\bar{\theta}} \left\{ u(\hat{q}(\theta,p),\theta) - c\hat{q}(\theta,p) - p\lambda \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,p) - \hat{q}(\phi,p)] f(\phi) d\phi \right\} f(\theta) \, d\theta. \tag{6}$$

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.

If Condition 1 is violated, the personal equilibrium is not unique and $S(p)$ depends on the selected equilibrium. The next lemma shows that we can focus on cases in which the personal equilibrium is unique, as long as the marginal cost is not too high.

**Lemma 1.** *Suppose the marginal cost $c > 0$ is sufficiently low. Then the joint surplus $S(p)$ is maximized for a unit price $p \in [0, \bar{p})$.*

For low marginal costs, a high unit price $p > \bar{p}$ leads to severe underconsumption compared to the first-best quantities. Furthermore, the joint surplus is decreased by losses that the consumers may feel. A flat-rate tariff, on the other hand, eliminates losses, and the efficiency loss due to overconsumption is also small, if the marginal cost is small. Therefore, prices $p > \bar{p}$ are dominated by a flat rate if marginal costs are not too high.

Lemma 1 does not provide an explicit bound on $c$. The range of marginal cost levels for which the personal equilibrium is unique, however, is increasing in $\bar{p}$. Hence, uniqueness is guaranteed for a large range of marginal cost levels, if preference heterogeneity is high.

In what follows, we assume that $c$ is such that $p < \bar{p}$ is optimal. Hence, we can focus on the case that the personal equilibrium $\langle \hat{q}(\theta,p)\rangle_{\theta\in\Theta}$ is characterized by Proposition 1.

Using integration by parts and equation (4) allows us to write the derivative of the joint surplus with respect to the marginal price $p$ as

$$S'(p) = \int_{\underline{\theta}}^{\bar{\theta}} \left\{ \left[p(1+\lambda(1-F(\theta))-c\right]\partial_p \hat{q}(\theta,p) - \lambda \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,p) - \hat{q}(\phi,p)] f(\phi) d\phi \right\} f(\theta)\, d\theta. \tag{7}$$

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 $p > c/[1+\lambda(1-F(\theta))]$ the joint surplus is strictly decreasing in $p$. Thus, $p^* \in [0, c/[1+\lambda(1-F(\theta))]]$. Without further assumptions, however, the joint surplus is not necessarily quasi concave. The following assumption guarantees strict concavity for $p \in [0, c/[1+\lambda(1-F(\theta))]]$.[24]

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

In order to ensure concavity, 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 expected losses. On the one hand, it increases expected losses due to increased variations in payments for 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.

Assumption 2 is by far not necessary for flat-rate tariffs to be optimal. If $S(p)$ is concave, however, $S'(p) \leq 0$ is a necessary and sufficient condition for optimality of a flat rate. This condition is easy to interpret and allows us to make statements about comparative statics. To cut back on our lengthy formulas we define

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

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

**Proposition 3.** *Suppose 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 on the side of 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

[24] 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 $2-1$ loss aversion.

consumption which is costly to the firm. 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 production costs due to overconsumption, and thus a flat-rate tariff is optimal.

Moreover, 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.[25] 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, strong variation in satiation points increases $\Sigma(\lambda)$ and makes a flat rate optimal for a wider range of marginal cost levels.[26]

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 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 $\sigma$).[27]

For $p < \bar{p} = 2\sigma/\lambda$, the personal equilibrium is unique and demand is strictly increasing in the demand type. Solving (4) yields

$$\hat{q}(\theta, p) = \theta - p\left[1 + \lambda \frac{1}{2\sigma}(\theta - \mu + \sigma)\right] = \left[1 - \frac{p}{\bar{p}}\right]\theta - p\left[1 + \frac{\sigma - \mu}{\bar{p}}\right]. \tag{8}$$

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 $(\mu + \sigma) - p(1 + \lambda) \leq \bar{q} \leq (\mu - \sigma) - p$.[28] At $p = \bar{p}$ the two bounds coincide. Moreover, evaluated at $p = \bar{p}$, we obtain that $\hat{q}(\bar{p}, \theta)$ equals the unique value $\bar{q}$ for all $\theta$.

In this example, the joint surplus is a quasi-concave function for all unit prices where

[25] Note, however, that $\Sigma(\lambda)$ is not linear in $\lambda$. For very high cost levels, there may be no $\lambda$ such that a flat rate is optimal. The example at the end of this section illustrates this point.

[26] Equation (1) shows that the difference between two satiation points is increasing in the cross derivative of the intrinsic utility function.

[27] Keeping $\sigma$ constant and varying the cross derivative of $\hat{u}$ would have the same qualitative effects.

[28] This follows from Lemma 4 in the appendix and the fact that $\tilde{q}$ is decreasing for $p \geq \bar{p}$. See also Footnote 23.

demand is positive, including $p > \bar{p}$.[29] Thus, a flat-rate tariff is optimal if and only if $S'(p)|_{p=0} \leq 0$, which is equivalent to

$$c \leq \frac{2}{3} \frac{\lambda\sigma}{2+\lambda} = \Sigma(\lambda). \tag{9}$$

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

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

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

[29]For $p \geq \bar{p}$ a personal equilibrium has to be selected. Kőszegi and Rabin (2006) define the *preferred personal equilibrium* as the plan among all credible plans—personal equilibria—that maximizes the decision maker's ex ante expected utility. If we select the preferred personal equilibrium demand is given by $q^{ppe}(p) = \max\{(\mu - \sigma) - p, 0\}$.

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

The monopolist's problem is then given by:

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

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]

$$P'(\theta) = \frac{\partial_q u(q(\theta),\theta)}{1+\lambda F(\theta)} q'(\theta). \tag{10}$$

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

$$\partial_q u(q(\theta),\theta) = \frac{(1+\lambda F(\theta))^2}{\lambda+1}c - F(\theta)(1-F(\theta))\frac{\lambda(1+\lambda F(\theta))}{\lambda+1}\frac{\partial_{q\theta} u(q(\theta),\theta)}{f(\theta)}. \tag{11}$$

By Assumption 3, this equation has a solution—there exists a $\tilde{q}(\theta) < q^S(\theta)$ that solves (11)—if and only if

$$c - \frac{\lambda F(\theta)(1-F(\theta))}{1+\lambda F(\theta)}\frac{\partial_{q\theta}\hat{u}(q^S(\theta),\theta)}{f(\theta)} > 0. \tag{12}$$

Assumption 3 also implies uniqueness of the solution if it exists. If (12) is violated, the optimal solution is given by $q^S(\theta)$. We define

$$q^*(\theta) = \begin{cases} \tilde{q}(\theta), & \text{if (12) is fulfilled for } \theta, \\ q^S(\theta), & \text{otherwise.} \end{cases}$$

Moreover, we define $P^*(\theta)$ as the payment rule that satisfies the (IR) constraint with equality and equation (10). 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.*

[31] In the web appendix, we show that the optimal mechanism is absolutely continuous and that the revenue equivalence formula is sufficient for (PE).

Equations (11) and (12) 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 (12), 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 $\theta \in \{\underline{\theta}, \bar{\theta}\}$, (12) is never violated and we have

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

We see that 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 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. The intuition is straight forward: a flatter quantity schedule decreases the variation in payments (cf. equation (10)) which decreases ex-ante expected losses.

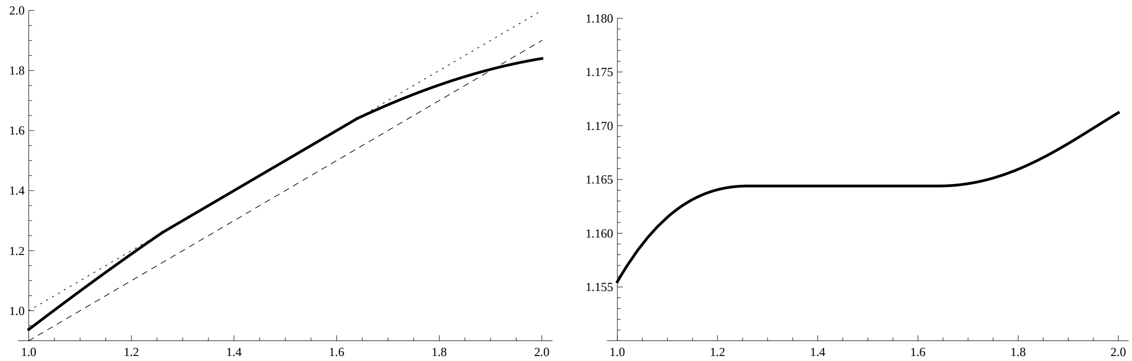
The second term in the first-order condition leads to an increase in the quantity consumed by intermediate types if $\lambda > 0$. If (12) is violated, this increase is so large that the optimal quantity coincides with the satiation point. In this case, the optimal tariff becomes flat. A flat part is 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 (10)). Increasing the quantity consumed by a type $\theta$ to the satiation point, where the marginal utility is zero, therefore makes the payment schedule flat around $\theta$. Compared to a situation where $P(\theta)$ increases steeply around $\theta$, this reduces the losses that types above $\theta$ feel compared to types below $\theta$. 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 $\theta$ in the loss function $(F(\theta))$. Hence, a flat part is most likely to be observed in the middle of the type-space where $F(\theta)(1 - F(\theta))$ is large. We conclude this section by a parametric example that illustrates the main insights.

**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

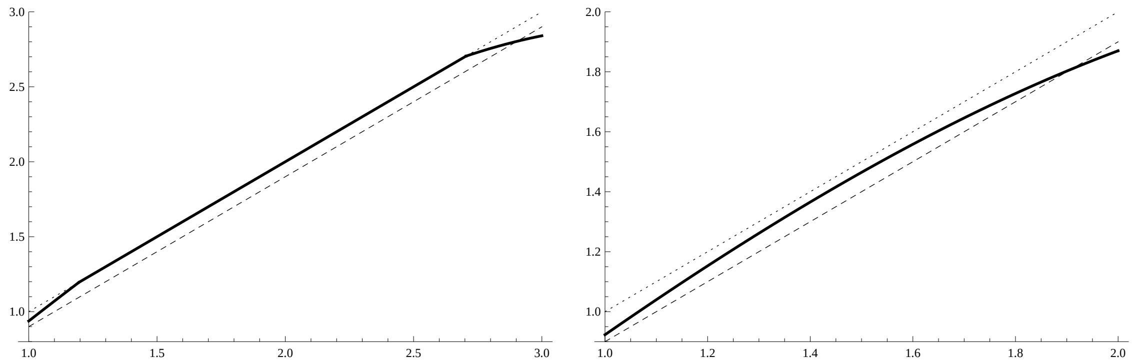
$$q^*(\theta) = \min\left\{\theta\,,\, \theta - \frac{(1 + \lambda d(\theta - \underline{\theta}))^2}{\lambda + 1} c + \frac{\lambda}{\lambda + 1}(1 + \lambda d(\theta - \underline{\theta}))(1 - d(\theta - \underline{\theta}))(\theta - \underline{\theta})\right\},$$

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 $\lambda \in \{0.3, 0.6\}$ and preference heterogeneity $d \in \{0.5, 1\}$. Moreover, we set $c = 0.1$ and $\underline{\theta} = 1$. In panel (a), loss aversion is high ($\lambda = 0.6$) and preference




(a) quantity $q^*(\theta)$ ($\lambda = 0.6$, $\Theta = [1, 2]$) (b) payment $P^*(\theta)$ ($\lambda = 0.6$, $\Theta = [1, 2]$)




(c) quantity $q^*(\theta)$ ($\lambda = 0.6$, $\Theta = [1, 3]$) (d) quantity $q^*(\theta)$ ($\lambda = 0.3$, $\Theta = [1, 2]$)

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

heterogeneity is low ($d = 1$). 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 $\theta$ (panel (c)), this fraction increases to more than 3/4. Conversely, if we keep preference heterogeneity fixed and decrease loss aversion to $\lambda = 0.3$, the flat part in the tariff vanishes and no demand type is satiated.

## 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 $\Sigma(\lambda_2) \geq c$. In these cases, the monopolist can screen differently loss-averse consumers at no cost. A characterization of the optimal menu of two-part tariffs for the case $\Sigma(\lambda_2) < c$ is outside the scope this paper.

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 this is also the case if $T_1$ is a measured tariff.

**Lemma 2.** *Consider a two-part tariff $(p, L)$ with $p < \bar{p}$ for $\lambda_2$. 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 it 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 degree 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 $(0, L^F)$ are characterized by: $S'(p_1^*|\lambda_1) = 0$, $L_1^* = S(p_1^*|\lambda_1) + (c - p_1^*)\int_{\underline{\theta}}^{\bar{\theta}} \hat{q}(\theta, p_1^*) f(\theta) d\theta$, 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.[32] 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 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 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 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 $(\mu_1, \sigma_1) = (1.3, 0.1)$ and $(\mu_2, \sigma_2) = (1.17, 1)$. Moreover, we assume that loss aversion is low in group one, $\lambda_1 = 0.3$, and high in group two, $\lambda_2 = 0.6$. The marginal cost of the monopolist is $c = 0.1$. 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 $(L_1, p_1) \approx (0.744, 0.081)$. For group two, it is optimal to offer a flat rate with basic charge $L_2 \approx 0.851$. 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 0.839 which is lower than

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

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.[33] 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. For instance, in Heidhues and Kőszegi (2008), there exists a focal price equilibrium, i.e., differentiated firms with different marginal costs charge the same focal price, only if competition is soft.[34] 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, say the rental car, which increases $\bar{u}$. A higher degree of competition corresponds to a higher $\bar{u}$. Notice, that the outside option is only a positive constant in the participation constraint of the firm's optimization problem. Thus, the optimal marginal 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 $L^* = c\int_{\underline{\theta}}^{\bar{\theta}} q^S(\theta) f(\theta) d\theta$.*

### *7.3. Loss Aversion in Both Dimensions*

In this part, 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.

$$\forall\, \theta \in \Theta \qquad u(q^S(\theta), \theta) \equiv \bar{u}. \tag{13}$$

This implies that higher demand types achieve lower utility levels, i.e., $\partial_\theta \hat{u}(q, \theta) < 0$. 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.

[33] A formal model of imperfect competition is analyzed in the web appendix.

[34] Similarly, in Heidhues and Kőszegi (2010a), a firm optimally sets variable sales prices next to sticky regular prices only if it has sufficient market power.

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 $\chi$ is chosen such that $\mathbb{E}_\theta[u(0,\theta)] = 0$ and $q^S(\theta) = \theta$, is of the form described above.

We posited that a consumer's well-being when consuming his ideal quantity is independent of the demand type. Intuitively, 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 $\theta$ 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.

If a consumer is loss averse also in the good dimension and the utility of not consuming the good at all is uncertain, the value of the outside option depends on the degree of loss aversion. When rejecting the contract, the consumer's expected utility is given by

$$E_\theta[U(0|\theta,\langle 0\rangle)] = \int_{\underline{\theta}}^{\bar{\theta}} \left\{ u(0,\theta) - \lambda \int_{\underline{\theta}}^{\theta} [u(0,\phi) - u(0,\theta)] f(\phi)\, d\phi \right\} f(\theta)\, d\theta \equiv K(\lambda), \quad (14)$$

with $K'(\lambda) < 0$.

Suppose that the consumer accepted a two-part tariff $(p, L)$ and that his expected demand is $\hat{q}(\cdot, p)$. If demand type $\theta$ 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 $\theta$. If demand type $\theta$ is realized and the consumer demands quantity $q$, then his utility is

$$\begin{aligned} U(q|\theta, \langle \hat{q}(\theta,p)\rangle) = u(q,\theta) - pq - L - \lambda \int_{Y(q,\theta)} \big[u(\hat{q}(z,p), z) - u(q,\theta)\big] f(z)\, dz \\ - \lambda \int_{X(q)} p\big[q - \hat{q}(z,p)\big] f(z)\, dz. \quad (15) \end{aligned}$$

The analysis of losses in both dimensions is more challenging because the set $Y$ does not only depend on $q$ but also on the demand type $\theta$. It is intricate to narrow down the set of potential personal equilibria or to show uniqueness.[35] Since we are mainly interested in the robustness of our findings, we construct only one personal equilibrium demand function with reasonable properties. The demand function should be continuous in $\theta$, since the utility function, the type space, and the two-part tariff are continuous. If the consumer expects a continuous demand function $\langle \hat{q}(\theta,p)\rangle$, choosing a slightly higher or lower quantity should have only a second-order effect on the set of types compared to which the consumer feels a loss in the good or the money dimension. Higher types have a higher intrinsic preference of the good. This also implies that for high types, increased consumption

[35] 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.

decreases the losses in the good dimension to a greater extent than for low types. Thus, higher types should demand (weakly) more in a personal equilibrium. Moreover, it seems reasonable that higher types do not consume so much more that they achieve higher intrinsic utility levels than lower types. Thus, we posit that on the equilibrium path the consumer feels losses in the good dimension as well as in the money dimension compared to lower types. Given that higher types consume strictly more and are strictly worse off compared to lower types, the personal equilibrium is characterized by

$$\partial_q u(\hat{q}(\theta, p), \theta) = p. \tag{16}$$

The demand function characterized by (16) is increasing in the demand type. Whether the second hypothesis—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 (16), 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* (16)*. 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 $\lambda$. Hence, if the consumer's degree of loss aversion is sufficiently large, then a flat-rate tariff is always optimal. This is not the case with loss aversion only in the money dimension as was illustrated in the example of Section 5.

**Example 4.** In order to obtain clear cut comparative static results regarding the optimality of flat-rate tariffs, we consider the following example. 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 (16) is

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

It can easily be verified that the demand function (17) indeed constitutes a personal equilibrium. In this personal equilibrium, the intrinsic utility in the good dimension equals $u(\hat{q}(\theta, p), \theta) = -(1/2)p^2$ 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

$$c \leq (1/3)\lambda\sigma = \Sigma(\lambda) . \tag{18}$$

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, the optimal pricing strategy of a monopolistic supplier when consumers are loss averse is analyzed. The optimal two-part tariff is a flat-rate contract if marginal costs are low and if consumers value sufficiently the insurance provided by the flat-rate contract. A flat-rate contract insures a loss-averse consumer against fluctuations in his billing amounts and this insurance is particularly valuable when 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 contains a large flat part for intermediate quantities if marginal costs are low and demand of the loss-averse consumer is highly uncertain.

## A. Appendix

### *A.1. Proofs of Section 4*

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:

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

We need two lemmas to prove Proposition 1.

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

*Proof.* The utility function $U(q|\theta, \langle \hat{q}(\phi, p)\rangle)$ is continuous in $(q, \theta)$. 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 $\theta$ by the theorem of the maximum.[36] □

[36] We thank an anonymous referee for suggesting this proof.

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

**Lemma 4.** *Let $\langle \hat{q}(\theta, p) \rangle$ be a personal equilibrium. Suppose $\hat{q}(\theta, p) = \bar{q}$ for all $\theta$ 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 $U(q|\theta, \langle \hat{q}(\phi, p) \rangle)$ is concave it has left and right derivatives at $\bar{q}$ which are given by

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

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, $\partial_q^- U(\bar{q}|\theta, \langle \hat{q}(\phi, p) \rangle) \approx \partial_q u(\bar{q}, a) - (1 + \lambda F(a))$ for $\theta$ close to $a$. Hence, all types $\theta$ 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. $\square$

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

*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

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

The consumer prefers the flat-rate tariff ex ante if

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

with

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

Equation (A.2) can be rearranged to

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

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

## *A.2. Proofs of Section 5*

*Proof of Lemma 1.* First, note that any personal equilibrium is bounded from above by $q^{\mathrm{MAX}}(p)$, which is implicitly defined by $\partial_q u(q^{\mathrm{MAX}}, \bar{\theta}) = p$. Let $q^{\mathrm{FB}}(\theta)$ denote the first-best quantities, i.e., $\partial_q u(q^{\mathrm{FB}}(\theta), \theta) = c$. We now show that for $p \geq \bar{p}$ the joint surplus $S(p)$ is bounded from above and that this bound is lower than $S(0)$. We define $\check{q}(\theta) := \min\{q^{\mathrm{FB}}(\theta), q^{\mathrm{MAX}}(\bar{p})\}$. There is a positive mass of types for which $\check{q}(\theta) = q^{\mathrm{MAX}}(\bar{p})$ if $\bar{p} > c$. The joint surplus generated with a unit price $p \geq \bar{p}$ is less than or equal

$$\check{S} = \int_{\underline{\theta}}^{\bar{\theta}} \big[u(\check{q}(\theta), \theta) - c\check{q}(\theta)\big] f(\theta) d\theta \;, \tag{A.5}$$

because with a positive unit price the consumer may incur some losses. If $S(0) \geq \check{S}$, then $S(p)$ is maximized by a unit price $p \in [0, \bar{p})$. $S(0) \geq \check{S}$ is equivalent to

$$\int_{\underline{\theta}}^{\bar{\theta}} \Big\{ u(q^S(\theta), \theta) - u(\check{q}(\theta), \theta) - c[q^S(\theta) - \check{q}(\theta)] \Big\} f(\theta) d\theta \geq 0 \;. \tag{A.6}$$

This condition is satisfied for $c$ sufficiently small, which completes the proof. □

*Proof of Proposition 3.* As argued in the main text, $S'(p) < 0$ for $p \geq c/(1+\lambda(1-F(\theta)))$. Next we show that $S''(p) \leq 0$ for $p \leq c/(1+\lambda(1-F(\theta)))$. $S''(\cdot)$ is given by

$$S''(p) = \int_{\underline{\theta}}^{\bar{\theta}} \{[p(1+\lambda(1-F(\theta)) - c]\partial_{pp}\hat{q}(\theta, p) + [1 + 2\lambda - 3\lambda F(\theta)]\partial_p \hat{q}(\theta, p)\}\, f(\theta)\; d\theta. \tag{A.7}$$

(A.7) is negative because $\partial_p \hat{q}(\theta, p) < 0$, $\lambda \leq 1$, and by Assumption 2,

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

Next, we show that $S'(0) \leq 0$ is equivalent to $\Sigma(\lambda) \geq c$. By evaluating (7) at $p = 0$, it is obvious that $S'(0) \leq 0$ if and only if

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

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

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

$$Z(\lambda) \equiv \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta, 0) - \hat{q}(\phi, 0)] f(\phi) f(\theta) d\phi d\theta, \tag{A.8}$$

$$\text{and} \qquad N(\lambda) \equiv - \int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta, 0) f(\theta) d\theta \, . \tag{A.9}$$

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

$$\partial_p \hat{q}(\theta, p) = \frac{1 + \lambda F(\theta)}{\partial_{qq} u(\hat{q}(\theta, p), \theta)} < 0 \tag{A.10}$$

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

$$\Sigma'(\lambda) = \frac{Z(\lambda)}{N(\lambda)} + \lambda \frac{Z'(\lambda) N(\lambda) - N'(\lambda) Z(\lambda)}{[N(\lambda)]^2}. \tag{A.11}$$

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.10) with respect to $\lambda$, and using $\partial_\lambda \hat{q}(\theta, 0) = 0$, yields

$$\partial_{\lambda p} \hat{q}(\theta, 0) = \frac{F(\theta)}{\partial_{qq} u(\hat{q}(\theta, 0), \theta)} (< 0) \, .$$

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.11) simplifies to

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

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 $q \leq q^S(\theta)$. $\square$

## *A.3. Proofs of Section 7.1*

*Proof of Lemma 2.* Define $V(\lambda, \theta)$ as the consumer's surplus for a given demand type on the personal equilibrium path. Formally,

$$V(\lambda, \theta) = u(\hat{q}(\theta, p), \theta) - p \hat{q}(\theta, p) - L - \lambda p \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta, p) - \hat{q}(\phi, p)] f(\phi) d\phi \tag{A.12}$$

Taking the derivative of $V(\cdot,\theta)$ with respect to $\lambda$ yields

$$V'(\lambda,\theta) = \partial_\lambda \hat{q}(\theta,p)\underbrace{[\partial_q u(\hat{q}(\theta,p),\theta) - p[1+\lambda F(\theta)]]}_{=0} - pF(\theta)\hat{q}(\theta,p) + p\int_{\underline{\theta}}^{\theta} \hat{q}(\phi,p)f(\phi)d\phi + \lambda p \int_{\underline{\theta}}^{\theta} \partial_\lambda \hat{q}(\phi,p)f(\phi)d\phi. \quad \text{(A.13)}$$

Using that $\partial_\lambda \hat{q}(\theta,p) = pF(\theta)/[\partial_{qq}u(\hat{q}(\theta,p))] \leq 0$, we have that $V'(\lambda,\theta) \leq 0$ if

$$\hat{q}(\theta,p)F(\theta) - \int_{\underline{\theta}}^{\theta} \hat{q}(\phi,p)f(\phi)d\phi \geq 0. \quad \text{(A.14)}$$

This condition is satisfied, because $\hat{q}(\theta,p)$ is non-decreasing in $\theta$.

The consumer's expected utility is given by $\mathbb{E}_\theta[V(\lambda,\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} V(\lambda,\theta)f(\theta)d\theta$. Hence, the change in expected utility due to an increase in the consumer's degree of loss aversion is given by $\frac{d}{d\lambda}\mathbb{E}_\theta[V(\lambda,\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} V'(\lambda,\theta)f(\theta)d\theta \leq 0$. □

## References


J. Abeler, A. Falk, L. Götte, and D. Huffman. Reference points and effort provision. *American Economic Review*, forthcoming.

L. Berseghyan, F. Molinari, T. O'Donoghue, and J. C. Teitelbaum. The nature of risk preferences: Evidence from insurance choices. working paper, Cornell University, 2010.

C. F. Camerer. Prospect theory in the wild: Evidence from the field. In C. F. Camerer, G. Loewenstein, and M. Rabin, editors, *Advances in Behavioural Economics*, chapter 5, pages 148–161. Princeton University Press, Princeton, New Jersey, 2004.

S. Choi, R. Fisman, D. Gale, and S. Kariv. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, 97(5):1921–1938, 2007.

K. Clay, D. S. Sibley, and P. Srinagesh. Ex post vs. ex ante pricing: Optimal calling plans and tapered tariffs. *Journal of Regulatory Economics*, 4:115–138, 1992.

P. Courty and H. Li. Sequential screening. *Review of Economic Studies*, 67:697–717, 2000.

V. Crawford and J. Meng. New york city cabdrivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review*, forthcoming.

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

S. DellaVigna and U. Malmendier. Paying not to go to the gym. *American Economic Review*, 96(3):694–719, 2006.

K. Eliaz and R. Spiegler. Contracting with diversely naive agents. *Review of Economic Studies*, 73:689–714, 2006.

K. Eliaz and R. Spiegler. Consumer optimism and price discrimination. *Theoretical Economics*, 3(4):459–497, 2008.

K. M. M. Ericson and A. Fuster. Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. working paper, Harvard University, 2010.

S. Esteban and E. Miyagawa. Temptation, self-control, and competitive nonlinear pricing. *Economics Letters*, 90:348–355, 2006.

S. Esteban, E. Miyagawa, and M. Shum. Nonlinear pricing with self-control preferences. *Journal of Economic Theory*, 135:306–338, 2007.

G. R. Faulhaber and C. Hogendorn. The market structure of broadband telecommunications. *Journal of Industrial Economics*, 48(3):305–329, 2000.

D. Gill and V. Prowse. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, forthcoming.

M. Grubb. Selling to overconfident consumers. *American Economic Review*, 99:1770–1807, 2009.

F. Gul and W. Pesendorfer. Temptation and self control. *Econometrica*, 60:1293–1330, 2001.

J.-H. Hahn, J. Kim, S.-H. Kim, and J. Lee. Screening loss averse consumers. working paper, 2010. URL `http://ssrn.com/abstract=1678825`.

P. Heidhues and B. Kőszegi. Competition and price variation when consumers are loss averse. *American Economic Review*, 98:1245–1268, 2008.

P. Heidhues and B. Kőszegi. Regular prices and sales. working paper, University of California, Berkeley, 2010a.

P. Heidhues and B. Kőszegi. Exploiting naivete about self-control in the credit market. *American Economic Review*, 100(5):2279–2303, 2010b.

F. Herweg, D. Müller, and P. Weinschenk. Binary payment schemes: Moral hazard and loss aversion. *American Economic Review*, 100(5):2451–2477, 12 2010.

M. Hobson and R. H. Spady. The demand for local telephone service under optimal local measured service. Bellcore Economics Discussion Paper, No. 50, 1988.

H. Karle and M. Peitz. Pricing and information disclosure in markets with loss-averse consumers. CEPR Discussion Paper, No. 7785, 2010a.

H. Karle and M. Peitz. Consumer loss aversion and the intensity of competition. Working Paper, Université Libre de Bruxelles., 2010b.

B. Kőszegi and M. Rabin. A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121:1133–1165, 2006.

B. Kőszegi and M. Rabin. Reference-dependent risk preferences. *American Economic Review*, 97:1047–1073, 2007.

J. P. Kling and S. S. van der Ploeg. Estimating local elasticities with a model of stochastic class of service and usage. In A. de Fontenay, M. H. Shugard, and D. S. Sibley, editors, *Telecommunications Demand Modeling: An Integrated View*, pages 119–136. North Holland, Amsterdam, 1990.

D. J. Kridel, D. E. Lehman, and D. L. Weisman. Option value, telecommunication demand, and policy. *Information Economics and Policy*, 5(2):125–144, 1993.

A. Lambrecht and B. Skiera. Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases. *Journal of Marketing Research*, 63(3): 212–223, 2006.

A. Lambrecht, K. Seim, and B. Skiera. Does uncertainty matter? Consumer behavior under three-part tariffs. *Marketing Science*, 26(5):698–710, 2007.

R. Macera. Intertemporal incentives under loss aversion. working paper, University of California, Berkeley, 2009.

E. Miravete. Estimating demand for local telephone service with asymmetric information and optional calling plans. *Review of Economic Studies*, 69:943–971, 2002.

E. J. Miravete. Choosing the wrong calling plan? Ignorance and learning. *American Economic Review*, 93(1):297–310, 2003.

B. M. Mitchell and I. Vogelsang. *Telecommunications Pricing: Theory and Practice*. Cambridge University Press, Cambridge, 1991.

Y.-K. Ng and M. Weissner. Optimal pricing with a budget constraint: The case of two-part tarifs. *Review of Economic Studies*, 41:337–345, 1974.

J. C. Nunes. A cognitive model of people's usage estimations. *Journal of Marketing Research*, 37:397–409, 2000.

T. Post, M. J. Van den Assem, D. Baltussen, and R. H. Thaler. Deal or no deal? Decision making under risk in a large-payoff game show. *American Economic Review*, 98(1): 38–71, 2008.

D. Prelec and G. Loewenstein. The red and the black: Mental accounting of savings and dept. *Marketing Science*, 17:4–28, 1998.

R. Spiegler. Monopoly pricing when consumers are antagonized by unexpected price increases: A "cover version" of the Heidhues-Kőszegi-Rabin model. working paper, London School of Economics, 2010.

A. Sundararajan. Nonlinear pricing of information goods. *Management Science*, 50(12): 1660–1673, 2004.

R. Thaler. Mental accounting matters. *Journal of Behavioral Decision Making*, 12:183–206, 1999.

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

K. E. Train, D. L. McFadden, and M. Ben-Akiva. 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, 1987.

K. E. Train, M. Ben-Akiva, and T. Atherton. Consumption patterns and self-selecting tariffs. *Review of Economics and Statistics*, 1:62–73, 1989.

A. Uthemann. Competitive screening of customers with non-common priors. working paper, University College London, 2005.

## Web Appendix

### A. Supplementary Material to Section 6: General Nonlinear Tariffs

The consumer's expected utility from accepting a direct mechanism $\langle q(\theta), P(\theta)\rangle$ for which truth-telling is a personal equilibrium and $q(\theta)$ is increasing, is given by

$$\begin{aligned}\mathbb{E}_\theta[V(\theta,\theta)] &= \int_{\underline{\theta}}^{\bar{\theta}} \left\{ u(q(\theta),\theta) - P(\theta) - \lambda \int_{\underline{\theta}}^{\theta} [P(\theta) - P(z)] f(z)\, dz \right\} f(\theta) d\theta. \\ &= \int_{\underline{\theta}}^{\bar{\theta}} [u(q(\theta),\theta) - P(\theta)] f(\theta) d\theta - \lambda \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} [P(z) - P(\theta)] f(z) f(\theta) dz d\theta\end{aligned}$$

Using integration by parts, the last term—the expected loss—can be simplified further.

$$\begin{aligned}& \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} [P(z) - P(\theta)] f(z) f(\theta) dz d\theta \\ = \; & \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} P(z) f(z) dz\, f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} P(\theta) \int_{\theta}^{\bar{\theta}} f(z) dz\, f(\theta) d\theta \\ = \; & \int_{\theta}^{\bar{\theta}} P(z) f(z) dz\, F(\theta) \Big|_{\underline{\theta}}^{\bar{\theta}} - \int_{\underline{\theta}}^{\bar{\theta}} -P(\theta) f(\theta) F(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} P(\theta)[1 - F(\theta)] f(\theta) d\theta \\ = \; & \int_{\underline{\theta}}^{\bar{\theta}} P(\theta)[2F(\theta) - 1] f(\theta) d\theta.\end{aligned}$$

Thus,

$$\mathbb{E}_\theta[V(\theta,\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} \left\{ u(q(\theta),\theta) - P(\theta)\xi(\theta) \right\} f(\theta) d\theta \tag{A.1}$$

where $\xi(\theta) = 1 + \lambda(2F(\theta) - 1)$.

We will formulate the monopolist's problem as an optimal control problem. In order to incorporate the participation constraint, we define

$$X(\theta) = \int_{\underline{\theta}}^{\theta} \left\{ u(q(z),z) - P(z)\xi(z) \right\} f(z) dz,$$

and impose $X(\bar{\theta}) \geq 0$. Note that $X(\bar{\theta}) = \mathbb{E}_\theta[V(\theta,\theta)]$, $X(\underline{\theta}) = 0$ and

$$X'(\theta) = \{u(q(\theta),\theta) - P(\theta)\xi(\theta)\} f(\theta).$$

In order to solve the monopolist's problem, we want to express the payment rule as a function of the allocation rule. The integral version of the local PE constraint can be derived from the envelope theorem (see Milgrom, 2004):

$$V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \partial_\theta u(q(s),s) ds = u(q(\theta),\theta) - P(\theta) - \lambda \int_{\underline{\theta}}^{\theta} (P(\theta) - P(z)) f(z) dz. \tag{A.2}$$

If $P$ is absolutely continuous, we can equivalently work with the differential version of this constraint

$$P'(\theta) = \frac{\partial_q u(q(\theta),\theta)}{\rho(\theta)} q'(\theta). \tag{A.3}$$

where $\rho(\theta) = 1 + \lambda F(\theta)$.

**Lemma 5.** *Let* $(q, P)$ *be a mechanism such that* $q$ *and* $P$ *are absolutely continuous and satisfy* (A.3)*. Then* $(q, P)$ *satisfies* (PE)*.*

*Proof.* Differentiating $V(\theta, \phi)$ with respect to the report $\theta$, we get

$$\begin{aligned}\partial_\theta V(\theta, \phi) &= \partial_q u(q(\theta), \phi) q'(\theta) - P'(\theta)\rho(\theta) \\ &= \left[\partial_q u(q(\theta), \phi) - \partial_q u(q(\theta), \theta)\right] q'(\theta) \begin{cases} \leq 0, & \text{if } \theta > \phi, \\ \geq 0, & \text{if } \theta < \phi. \end{cases}\end{aligned}$$

Hence, truth-telling $(\theta = \phi)$ maximizes $V(\theta, \phi)$. □

We cannot assume a priori, however, that $P$ is absolutely continuous. Instead, we first solve the monopolist's problem assuming that $q$ is globally Lipschitz continuous (which implies absolute continuity). The following Lemma shows that this also implies Lipschitz continuity of $P$ and hence absolute continuity.

**Lemma 6.** *Let* $\langle q(\theta), P(\theta)\rangle_{\theta \in \Theta}$ *be a direct mechanism that satisfies (PE). If* $q$ *is globally Lipschitz continuous and Assumption 1 is fulfilled, then* $P$ *is also globally Lipschitz continuous.*

*Proof.* For $\theta_1 < \theta_2$, the personal equilibrium constraint implies

$$\begin{aligned}P(\theta_2) - P(\theta_1) &\leq u(q(\theta_2), \theta_2) - u(q(\theta_1), \theta_2) \\ &\qquad - \lambda \int_{\underline{\theta}}^{\theta_2} \left[P(\theta_2) - P(z)\right] f(z) dz + \lambda \int_{\underline{\theta}}^{\theta_1} \left[P(\theta_1) - P(z)\right] f(z) dz \\ &= u(q(\theta_2), \theta_2) - u(q(\theta_1), \theta_2) \\ &\qquad - \lambda \int_{\theta_1}^{\theta_2} \left[P(\theta_2) - P(z)\right] f(z) dz + \lambda \left[P(\theta_1) - P(\theta_2)\right] F(\theta_1) \\ &\leq u(q(\theta_2), \theta_2) - u(q(\theta_1), \theta_2) \\ &\leq K\left(\theta_2 - \theta_1\right)\end{aligned}$$

for some $K < \infty$, where the last line follows from Lipschitz continuity of $q$ and $u$ (Assumption 1). □

Assuming Lipschitz continuity of $q$, we can restate the problem of the monopolistic seller as the following control problem.

Control Problem:

$$\max_{(q,P,X,v)} \int_{\underline{\theta}}^{\bar{\theta}} [P(\theta) - cq(\theta)] f(\theta)\, d\theta$$

$$\begin{aligned} \text{subject to} \qquad & \text{(i)} \quad P'(\theta) = \frac{\partial_q u(q,\theta)}{\rho(\theta)} v \\ & \text{(ii)} \quad q'(\theta) = v \\ & \text{(iii)} \quad X'(\theta) = (u(q,\theta) - P(\theta)\xi(\theta)) f(\theta) \\ & \text{(iv)} \quad v(\theta) \in [0, K] \\ & \text{(v)} \quad q(\underline{\theta}) \geq 0 \\ & \text{(vi)} \quad X(\underline{\theta}) = X(\bar{\theta}) = 0 \\ & \text{(vii)} \quad q(\theta) \leq q^*(\theta),\ \forall\theta. \end{aligned}$$

$v$ is the control variable and $P$, $q$, and $X$ are state variables. The state variable $X$ and constraints (iii) and (vi) are introduced to capture the participation constraint. The constant $K > 0$ in constraint (iv) is the Lipschitz-constant.

In the control problem, we impose the additional assumption that $q$ is Lipschitz continuous with constant $K$. We will show that for $K$ sufficiently large, the constraint $v(\theta) \leq K$ is not binding in the solution to the control problem. The next Lemma shows that an optimal solution for which $v(\theta) < K$, is also an optimal solution to the seller's problem without the Lipschitz constraint.

**Lemma 7.** *Let $(P, q, X, v)$ be an optimal solution to the control problem for some $K$, that satisfies*

$$v(\theta) < K, \qquad \textit{for all } \theta \in \Theta.$$

*Then $(q, P)$ is an optimal solution to the monopolist's problem without the assumption of Lipschitz continuity.*

*Proof.* Suppose $\langle q^*(\theta), P^*(\theta)\rangle_{\theta\in\Theta}$ is an optimal mechanism for the monopolist ($q^*$ need not be Lipschitz continuous). If we extend $q^*$ to the real line by setting $q^*(\theta) = q^*(\underline{\theta})$ for $\theta < \underline{\theta}$, and $q^*(\theta) = q^*(\bar{\theta})$ for $\theta > \bar{\theta}$, we can approximate $q^*$ by

$$q_k(\theta) := k \int_{\theta - \frac{1}{2k}}^{\theta + \frac{1}{2k}} q^*(s) ds.$$

This yields a sequence of functions $(q_k)_{k\in\mathbb{N}}$ such that $q_k$ is Lipschitz continuous with constant $k$, and $q_k(\theta) \to q^*(\theta)$ and $q_k'(\theta) \to q^{*\prime}(\theta)$ as $k \to \infty$ for almost every $\theta$. Using the local PE constraint (i) and the participation constraint (vi), we define the corresponding payment rules $(P_k)_k$.

If we ignore the participation constraint and arbitrarily set the payment of the lowest type to zero, we get a sequence of payments rules $(\tilde{P}_k)_k$. ($P_k$ differs from $\tilde{P}_k$ by the constant $P_k(\underline{\theta})$, which is determined by the participation constraint.) By Helly's theorem, there exists a sub-sequence $(\tilde{P}_{k_n})_n$ and an increasing function $\tilde{P}$, such that $\tilde{P}_{k_n}(\theta) \to \tilde{P}(\theta)$ for almost every $\theta$. Furthermore (possibly after taking sub-sequences again), $P_{k_n}(\underline{\theta})$ converges to some value $\hat{P}(\underline{\theta})$ if $n \to \infty$. For $\theta > \underline{\theta}$ we define $\hat{P}(\theta) = \hat{P}(\underline{\theta}) + \tilde{P}(\theta)$. By definition, each mechanism in the sequence $(\langle q_{k_n}(\theta), P_{k_n}(\theta)\rangle_{\theta\in\Theta})_n$ fulfills the local PE constraint and the participation constraint. Furthermore the sequence converges to the mechanism $\langle q^*(\theta), \hat{P}(\theta)\rangle_{\theta\in\Theta}$. Since the sequence of mechanisms is bounded and converges almost everywhere, the limit also fulfills participation and PE constraints.

Next we show that $\hat{P}(\theta) = P^*(\theta)$. Denote the utility of type $\theta$ in the mechanism $\langle q^*(\theta), \hat{P}(\theta)\rangle_{\theta\in\Theta}$ by $\hat{V}(\theta)$ and in the optimal mechanism by $V^*(\theta)$. Define $D(\theta) = P^*(\theta) - \hat{P}(\theta)$. Since both mechanisms fulfill (A.2), we have $D(\underline{\theta}) = \hat{V}(\underline{\theta}) - V^*(\underline{\theta})$. Subtracting (A.2) for the two mechanism yields

$$D(\theta)\rho(\theta) = \hat{V}(\underline{\theta}) - V^*(\underline{\theta}) + \lambda \int_{\underline{\theta}}^{\theta} D(z)f(z)\,dz$$

Hence, $D$ is absolutely continuous. Differentiating and rearranging we get

$$D'(\theta) = 0.$$

Therefore, $D(\underline{\theta}) \neq 0$ is not possible because both mechanisms fulfill the individual rationality constraint with equality. We have shown that $\hat{P} = P^*$.

Now let $(q, P, X, v)$ be an optimal solution to the control problem such that $v(\theta) < K$ for some $K$. Since the Lipschitz constraint for $K$ is not binding for $(q, P, X, v)$, the expected revenue from $(q, P)$ is at least as high as the revenue from $(q_{k_n}, P_{k_n})$ for all $k_n > K$. As $q_{k_n}(\theta)$ and $P_{k_n}(\theta)$ converge to $q^*(\theta)$ and $P^*(\theta)$ almost everywhere, the revenue from $(q, P)$ is also weakly greater than the revenue from $(q^*, P^*)$. Hence, $(q, P)$ is an optimal mechanism. □

The Hamiltonian corresponding to the above problem is given by

$$\mathcal{H}(\theta, q, P, X, p_q, p_P, p_X, v) = (T - cq)f(\theta) + p_q v + \frac{p_P}{\rho(\theta)}\partial_q u(q,\theta)v + p_X[u(q,\theta) - T\xi(\theta)]f(\theta).$$

Applying the Pontryagin Maximum Principle (cf. Clarke, 1983) we obtain the following necessary conditions for an optimal control. ($p_q$, $p_P$ and $p_X$ are absolutely continuous functions.)

(i) adjoint equations: for almost every $\theta \in \Theta$,

$$p_q'(\theta) \begin{cases} = cf(\theta) - \frac{p_P(\theta)}{\rho(\theta)}\partial_{qq}u(q(\theta),\theta)v(\theta) - p_X(\theta)\partial_q u(q(\theta),\theta)f(\theta), & \text{if } q(\theta) < q^S(\theta), \\ \in \left[cf(\theta) - \frac{p_P(\theta)}{\rho(\theta)}\partial_{qq}\hat{u}(q^S(\theta),\theta)q^{s\prime}(\theta),\ cf(\theta)\right], & \text{if } q(\theta) = q^S(\theta), \\ = cf(\theta), & \text{if } q(\theta) > q^S(\theta). \end{cases} \tag{A.4}$$

$$p_P'(\theta) = -f(\theta) + p_X(\theta)\xi(\theta)f(\theta), \tag{A.5}$$

$$p_X'(\theta) = 0 \quad (\Rightarrow \ p_X'(\theta) = p_X). \tag{A.6}$$

(ii) optimality of control: for almost every $\theta \in \Theta$,

$$v(\theta) \begin{cases} = K, & \text{if } p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) > 0, \\ \in [0, K], & \text{if } p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) = 0, \\ = 0, & \text{if } p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) < 0. \end{cases} \tag{A.7}$$

(iii) transversality conditions:

$$p_q(\bar{\theta}) = p_P(\underline{\theta}) = p_P(\bar{\theta}) = 0, \tag{A.8}$$

$$p_q(\underline{\theta}) \leq 0 \ \ (\text{“} = \text{”} \text{ if } q(\underline{\theta}) > 0), \tag{A.9}$$

$$\text{and} \qquad p_X(\underline{\theta}),\ p_X(\bar{\theta}) \text{ free} \tag{A.10}$$

**Lemma 8.** $p_X = 1$ *and* $p_P(\theta) = \lambda((F(\theta)^2 - F(\theta))(< 0)$.

*Proof.* The adjoint equation for $p'_P$ and the transversality condition imply

$$p_P(\theta) = \int_{\underline{\theta}}^{\theta} -f(s) + p_X f(s)\xi(\theta) ds. \tag{A.11}$$

Evaluating equation (A.11) at $\theta = \bar{\theta}$ and using the transversality condition for $p_P(\bar{\theta})$ we obtain

$$\begin{aligned} \int_{\underline{\theta}}^{\bar{\theta}} -f(s) + p_X f(s)[1 + \lambda(2F(s) - 1)] ds &= 0 \\ \Leftrightarrow -1 + p_X[1 + \lambda \int_{\underline{\theta}}^{\bar{\theta}} (2F(s) - 1) f(s) ds] &= 0 \\ \Leftrightarrow p_X &= 1. \end{aligned}$$

Inserting $p_X = 1$ into (A.5) and (A.11) we obtain

$$\begin{aligned} p'_P(\theta) &= \lambda(2F(\theta) - 1) f(\theta) \\ p_P(\theta) &= \lambda[(F(\theta))^2 - F(\theta)]. \end{aligned}$$

□

Denote the Lipschitz constants of $\tilde{q}$ (from the main text) and $q^S$ by $\tilde{K}$ and $K^S$, respectively. Our assumptions on the utility function guarantee that $\max\{\tilde{K}, K^S\} < \infty$.

**Lemma 9.** *If* $K > \max\{\tilde{K}, K^S\}$, *then* $q(\theta) \leq q^S(\theta)$ *for all* $\theta \in \Theta$.

*Proof.* Suppose by contradiction $q(\theta) > q^S(\theta)$ for all $\theta$ in a maximal interval $(a, b)$ with $a < b$. Equation (A.4) then implies that $p'_q(\theta) = cf(\theta)$ for $\theta \in (a, b)$. If $p_q(\theta) \geq 0$, for some $\theta < b$ then we must have $v(t) = K$ for all $t \in (\theta, b)$ and $b = \bar{\theta}$. This implies $p_q(\bar{\theta}) > 0$ in contradiction to the transversality condition. Hence, we have $p_q(\theta) < 0$ for all $\theta < a$. This implies $v(\theta) = 0$ for all $\theta \in (a, b)$ and since $q^S$ is increasing we must have $a = \underline{\theta}$. But $q(a) > q^S(a) > 0$ implies $p_q(a) = 0$ by the transversality condition. This is a contradiction. □

**Lemma 10.** *Suppose that $K > \max\{\tilde{K}, K^S\}$. If $q(\theta) = q^S(\theta)$ for all $\theta \in (a,b)$, $a < b$, then for all $\theta \in (a,b)$*

$$cf(\theta) - \frac{p_P(\theta)}{\rho(\theta)} \partial_{qq}\hat{u}(q^S(\theta), \theta) q^{S\prime}(\theta) \leq 0. \tag{A.12}$$

*Proof.* Since $\partial_q u(q^S(\theta), \theta) = 0$ and $q^{S\prime}(\theta) \in (0, K)$, (A.7) implies that $p_q(\theta) = 0$ for all $\theta \in [a,b]$. Hence, $p_q'(\theta) = 0$ and from (A.4) we obtain (A.12). $\square$

Next we derive properties of the optimal solution if $q(\theta) < q^S(\theta)$. Integrating (A.4), yields

$$\begin{aligned}
p_q(t) &= p_q(s) + \int_s^t cf(r) - \frac{p_P(r)}{\rho(r)} \partial_{qq} u(q(r), r) v(r) - \partial_q u(q(r), r) f(r) dr \\
&= p_q(s) + \int_s^t cf(r) - \frac{p_P(r)}{\rho(r)} \left[ \frac{d}{dr} \partial_q u(q(r), r) - \partial_{q\theta} u(q(r), r) \right] - \partial_q u(q(r), r) f(r) dr \\
&= p_q(s) + \frac{p_P(s)}{\rho(s)} \partial_q u(q(s), s) - \frac{p_P(t)}{\rho(t)} \partial_q u(q(t), t) \\
&\quad + \int_s^t cf(r) + \left[ \frac{d}{dr} \frac{p_P(r)}{\rho(r)} \right] \partial_q u(q(r), r) - \partial_q u(q(r), r) f(r) + \frac{p_P(r)}{\rho(r)} \partial_{q\theta} u(q(r), r) dr
\end{aligned}$$

Using

$$\left[ \frac{d}{d\theta} \frac{p_P(\theta)}{\rho(\theta)} - f(\theta) \right] = -(\lambda + 1) \frac{f(\theta)}{(\rho(\theta))^2}$$

we get

$$\begin{aligned}
p_q(t) = p_q(s) &+ \frac{p_P(s)}{\rho(s)} \partial_q u(q(s), s) - \frac{p_P(t)}{\rho(t)} \partial_q u(q(t), t) \\
&+ \int_s^t cf(r) - (\lambda + 1) \frac{f(r)}{(\rho(r))^2} \partial_q u(q(r), r) + \frac{p_P(r)}{\rho(r)} \partial_{q\theta} u(q(r), r) dr
\end{aligned} \tag{A.13}$$

**Lemma 11.** *If $K > \max\{\tilde{K}, K^S\}$, then $v(\theta) < K$ for all $\theta \in \Theta$.*

*Proof.* Suppose by contradiction, that for all $\theta$ in a maximal interval $(a,b)$, $a < b$, the control variable is $v(\theta) = K$. Then $p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) \geq 0$ with equality for $\theta \in \{a, b\}$. If the endpoints are $a = \underline{\theta}$ or $b = \bar{\theta}$, respectively, then equality follows from the transversality conditions. Otherwise, it follows because $(a,b)$ is chosen maximally and the left-hand side of the inequality is continuous in $\theta$. A strict inequality at an endpoint would imply that the interval where $v(\theta) = K$, extends beyond the endpoint.

We show that $q(a) \geq q^*(a)$ and $q(b) \leq q^*(b)$. First, suppose by contradiction that $q(a) < q^*(a)$. Using (A.13) for $\theta > a$ close to $a$ we get:

$$\begin{aligned}
&p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) \\
&= \int_a^\theta cf(r) - (\lambda + 1) \frac{f(r)}{(\rho(r))^2} \partial_q u(q(r), r) + \frac{p_P(r)}{\rho(r)} \partial_{q\theta} u(q(r), r) dr \\
&< \int_a^\theta cf(r) - (\lambda + 1) \frac{f(r)}{(\rho(r))^2} \partial_q u(q^*(r), r) + \frac{p_P(r)}{\rho(r)} \partial_{q\theta} u(q^*(r), r) dr \leq 0
\end{aligned}$$

The strict inequality follows from $q(a) < q^*(a)$ because for $r$ close to $a$, $q(\theta) < q^*(\theta)$ and hence $\partial_q u(q(r), r) > \partial_q u(q^*(r), r)$ by concavity of $u$ and $\partial_{q\theta} u(q(\theta), \theta) \geq \partial_{q\theta} u(q^*(\theta), \theta)$ by Assumption 3. The weak inequality follows from the definition of $q^*$. But this contradicts $u(\theta) = K$. Hence $q(a) \geq q^*(a)$. Similarly, it can be shown that $q(b) \leq q^*(b)$. But $q(a) \geq q^*(a)$, $q(b) \leq q^*(b)$ and $v(\theta) = K$ for $\theta \in (a, b)$ cannot be fulfilled simultaneously if $K > \max\{\tilde{K}, K^S\}$. □

**Lemma 12.** *If $q^*$ is strictly increasing, then every optimal solution to the control problem is strictly increasing.*

*Proof.* Suppose by contradiction, that the control is zero ($v(\theta) = 0$) on a maximal interval $(a, b)$. Then $q(\theta) < q^S(\theta)$ for $\theta \in (a, b)$ and $p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) \leq 0$, with equality for $\theta = b$ and for $\theta = a$ unless $a = \underline{\theta}$ and $q(a) = 0$. We first show that $q(a) = q^*(a)$. Suppose by contradiction that $q(a) < q^*(a)$. Using (A.13) for $\theta > a$ close to $a$ we get

$$\begin{aligned}
&p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) \\
&< \int_a^\theta cf(r) - (\lambda + 1) \frac{f(r)}{(\rho(r))^2} \partial_q u(q^*(r), r) + \frac{p_P(r)}{\rho(r)} \partial_{q\theta} u(q^*(r), r)\, dr \leq 0
\end{aligned}$$

This implies $p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) < 0$ and $v(\theta) = 0$ for all $\theta \in (a, \bar{\theta})$ if $q^*$ is strictly increasing. Since $p_P(\bar{\theta}) = 0$, this implies $p_q(\bar{\theta}) < 0$ in contradiction to the transversality condition. Hence $q(a) \geq q^*(a)$. If $q(a) > q^*(a)$, we have $q^*(a) = \tilde{q}(a)$ and $q(a) > \tilde{q}(a)$. Hence, using (A.13) we get $p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) > 0$ for $\theta$ close to $a$ which contradicts $v(\theta) = 0$. Similarly, we can show that $q(b) \geq q^*(b)$ which yields the desired contradiction if $q^*(\theta)$ is strictly increasing. □

*Proof of Proposition 4.* Consider an optimal solution $(q, P, X, v)$ to the control problem for $K > \max\{\tilde{K}, K^S\}$. By Lemmas 11 and 12, we have $v(\theta) \in (0, K)$. If $q(\theta) < q^S(\theta)$ for all $\theta \in (a, b)$ this implies $p_q(\theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_q u(q(\theta), \theta) = 0$ for all $\theta \in [a, b]$. Inserting into (A.13) we get

$$\int_a^\theta cf(r) - (\lambda + 1) \frac{f(r)}{(\rho(r))^2} \partial_q u(q(r), r) + \frac{p_P(r)}{\rho(r)} \partial_{q\theta} u(q(r), r) dr = 0.$$

Differentiating this with respect to $\theta$ yields for almost every $\theta \in (a, b)$:

$$cf(\theta) - (\lambda + 1) \frac{f(\theta)}{(\rho(\theta))^2} \partial_q u(q(\theta), \theta) + \frac{p_P(\theta)}{\rho(\theta)} \partial_\theta \partial_q u(q(\theta), \theta) = 0$$

Rearranging, we get

$$\begin{aligned}
\partial_q u(q(\theta), \theta) &= \frac{(\rho(\theta))^2}{\lambda + 1} c + \frac{p_P(\theta)\rho(\theta)}{(\lambda + 1)} \frac{\partial_{\theta q} u(q(\theta), \theta)}{f(\theta)} \qquad \text{(A.14)} \\
&= \frac{(1 + \lambda F(\theta))^2}{\lambda + 1} c - F(\theta)(1 - F(\theta)) \frac{\lambda(1 + \lambda F(\theta))}{\lambda + 1} \frac{\partial_{q\theta} u(q(\theta), \theta)}{f(\theta)}.
\end{aligned}$$

This is the first-order condition (11) from the main paper. Inserting $v^S(\theta) = -\frac{\partial_{q\theta}\hat{u}(q^S(\theta),\theta)}{\partial_{qq}\hat{u}(q^S(\theta),\theta)}$ into (A.12) yields

$$f(\theta)c + \frac{p_P(\theta)}{\rho(\theta)}\partial_{q\theta}\hat{u}(q^S(\theta),\theta) \leq 0,$$

or equivalently,

$$\frac{(\rho(\theta))^2}{\lambda+1}c + \frac{p_P(\theta)\rho(\theta)}{(\lambda+1)}\frac{\partial_{q\theta}\hat{u}(q^S(\theta),\theta)}{f(\theta)} \leq 0.$$

This is the opposite of condition (12) in the main paper. By Assumption 1 (A.14) has a solution $q(\theta) < q^S(\theta)$ if and only if (A.12) is violated. Assumption 1 also guarantees uniqueness of the solution. Hence, $q(\theta) = \tilde{q}(\theta)$ if (A.12) is violated. If (A.12) is fulfilled

$$0 \in \left[cf(\theta) - \frac{p_P(\theta)}{\rho(\theta)}\partial_{qq}\hat{u}(q^S(\theta),\theta)q^{S\prime}(\theta),\ cf(\theta)\right],$$

and hence $q(\theta) = q^S(\theta)$ fulfills the necessary condition from the maximum principle. We have shown that $q^*(\theta)$ fulfills the necessary conditions for optimality and because $q(\theta) \leq q^S(\theta)$ it is the unique solution. Existence of a solution can be shown by standard techniques. Therefore, we have have constructed an optimal solution to the control problem. By Lemma 7 it is also a solution to the general problem. □

## B. Supplementary Material to Section 7.2: Duopolistic Competition

### *B.1. Market Framework*

In this part of the web appendix, a formal model of imperfect competition is considered. Moreover, we allow for heterogeneity among consumers with respect to the degree of loss aversion. Consider a market for one good or service where two firms, $A$ and $B$, are active. Moreover, there is a continuum of ex ante heterogeneous consumers whose measure is normalized to one.

*Players & Timing.*—The consumers can be partitioned into two groups that differ in their degrees of loss aversion. Let the two groups be denoted by $j = 1, 2$ with $0 \leq \lambda_1 < \lambda_2$. The distribution of demand types is identical for both groups of loss-averse consumers. As before, the demand type is unknown to consumers and firms at the point of contracting.

The two symmetric firms, $A$ and $B$, produce at constant marginal cost $c > 0$ and without fixed cost. Each firm $i = A, B$ offers a two-part tariff to each group of consumers $j = 1, 2$. The tariff is given by $T^i_j(q) = L^i_j + p^i_j q$, where $q \geq 0$ is the quantity, and $L^i_j$ and $p^i_j$ denote the fixed fee and the unit price, respectively, charged by firm $i$ from consumers of type $j$. We will analyze the symmetric information case in which firms can observe $\lambda$, as well as the asymmetric information case in which $\lambda$ is private information of the consumer.

The timing is as follows: (1) Firms simultaneously and independently offer a menu of two-part tariffs $\{(L^i_j, p^i_j)\}_{j=1,2}$ to consumers. (2) Each consumer either signs exactly one contract or none. (3) Each consumer privately observes his demand type. Thereafter, each consumer who accepted a contract chooses a quantity. (4) Finally, payments are made according to the demanded quantities and the concluded contracts.

*Discrete Choice Framework.*—The products of the two firms are symmetrically differentiated. We assume that, next to $\lambda$, consumers are ex ante heterogeneous with respect to their brand preferences. Each consumer has idiosyncratic preferences for differing brands of the product (firms), which are parameterized by $\zeta = (\zeta^0, \zeta^A, \zeta^B)$. A consumer with brand preferences $\zeta$ has net utility $v^i + \zeta^i$ if he buys from firm $i$, and net utility $\zeta^0$ if no contract is signed, where $v^i = \mathbb{E}_\theta[U(\cdot)]$. The brand preferences $\zeta = (\zeta^0, \zeta^A, \zeta^B)$ are independently and identically distributed according to a known distribution among the two groups of consumers.

To solve for the tariffs that are offered in the pure-strategy Nash equilibrium by the two firms, we follow the approach of Armstrong and Vickers (2001) and model firms as offering utility directly to consumers. Each two-part tariff can be considered as a deal of a certain expected value that is offered by a firm to its consumers. Thus, firms compete over customers by trying to offer them better deals, i.e., a two-part tariff that yields higher utility (including loss utility). Put differently, we decompose a firm's problem into two parts. First, we solve for the two-part tariff that maximizes profits subject to the constraint that the consumer receives a certain utility level. Thereafter, we solve for the utility levels $(v_1^i, v_2^i)$ a firm $i$ offers to its customers. It is important to note that when $\lambda$ is unobservable, the two-part tariffs have to be designed such that each group of consumers prefers the offer that is intended for them. Suppose that the utility offered to consumers of group $j$ by firm $A$ and firm $B$ is $v_j^A$ and $v_j^B$, respectively. Furthermore, assume that the incentive constraints are satisfied. Then, the market share of firm $A$ in the sub market $j$ is $m_j(v_j^A, v_j^B)$ and the market share of firm $B$ is $m_j(v_j^B, v_j^A)$, with $m_j(v_j^A, v_j^B) + m_j(v_j^B, v_j^A) \leq 1$. The market share function $m_j(\cdot)$ is increasing in the first argument and decreasing in the second. Since the brand preferences are identically distributed among the two groups, the market share functions are identical for the two sub markets, i.e., $m_1(\cdot) = m_2(\cdot) = m(\cdot)$. Following Armstrong and Vickers, we impose some regularity conditions in order to guarantee existence of equilibrium. First, we assume that

$$\frac{\partial_{v^A} m(v^A, v^B)}{m(v^A, v^B)} \text{ is non-decreasing in } v^B.$$

Second, we assume that for each sub market the collusive utility level $\tilde{v}_j$ exists which maximizes (symmetric) joint profits.[37]

*B.2. Firm's Subproblem: Joint Surplus Maximization*

For this part, suppose firms can observe consumers' types $\lambda \in \{\lambda_1, \lambda_2\}$. With consumers' loss aversion types being observable, the two market segments of types $\lambda_1$ and $\lambda_2$ can

[37] For a detailed description of the competition-in-utility-space framework and the needed assumptions see Armstrong and Vickers (2001). A similar approach is used by Rochet and Stole (2002).

be viewed as distinct markets. Thus, for the analysis we can focus on one market where consumers are homogeneous with respect to their degree of loss aversion, which is denoted by $\lambda$.

Suppose firm $i \in \{A, B\}$ offers consumers a "deal" using a two-part tariff $(L^i, p^i)$ that gives them utility $v^i$. Then, if a consumer with brand preferences $\zeta = (\zeta^0, \zeta^A, \zeta^B)$ purchases from firm $i$ his net utility is $v^i + \zeta^i$. Let $\pi_j(v^i)$ be firm $i$'s maximum profit per customer of type $j$ when offering them a deal that yields utility $v^i$. The per-consumer profit function is the same for both firms, but in general it depends on the consumer's degree of loss aversion $\lambda$. For now we focus on one market segment and therefore the subscript indicating the loss-aversion type can be omitted without confusion. Since $\pi(\cdot)$ is the same for both firms, we will omit firm's superscript in the following. With this notation, $\pi(v)$ is given by the solution to the problem:

$$\pi(v) \ = \ \max_{L,p\geq 0} \ : \ \left\{ L + (p-c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}(\theta,p) f(\theta) d\theta \ \middle| \ \mathbb{E}_\theta[U(\hat{q}(\theta,p)|\theta, \langle q(\phi,p)\rangle)] = v \right\}. \quad \text{(B.1)}$$

First, we study the firm's subproblem, that is, we derive the optimal two-part tariff that solves the above problem. Thereafter, we solve for the utility levels and the corresponding tariffs which are offered by the two firms in equilibrium. Put differently, the task is to maximize a firm's profit over the choice variables $p$ and $L$ subject to the constraint that the consumer's expected utility from the offered deal is $v$. The firm's tariff choice problem can be restated as a problem of choosing only the unit price $p$. The firm chooses $p$ to maximize $S(p) - v$, i.e., the firm chooses the marginal price $p$ such that the joint surplus of the two contracting parties, the consumer and the firm, is maximized. The optimal marginal price $\hat{p}$ is independent of the utility level $v$, that the firm offers to the consumer. This immediately implies that $\pi'(v) = -1$. More importantly, the optimal marginal price is characterized by the same conditions as in the case of a monopolistic firm.

In the following we focus on the profit maximization problem of firm $A$. We assume that Assumption 2 holds for both types of loss-averse consumers, i.e., for $\lambda \in \{\lambda_1, \lambda_2\}$. Moreover, it is assumed that $\Sigma(\lambda_2) \geq c$.

*B.3. Symmetric Information Case*

Consider market segment $j \in \{1, 2\}$. For a given utility level $v_j^B$ offered by firm $B$, the profit maximization problem of firm $A$ is given by

$$\max_{v_j^A} \ m(v_j^A, v_j^B)\pi_j(v_j^A) \ . \quad \text{(B.2)}$$

The necessary first-order condition for profit maximization amounts to

$$\partial_{v_j^A} m(v_j^A, v_j^B) \, \pi_j(v_j^A) + m(v_j^A, v_j^B) \, \pi_j'(v_j^A) = 0 \ . \quad \text{(B.3)}$$

Remember that $\pi_j'(v^A) = -1$. The optimal marginal price is unaffected by the choice of $v_j^A$. If firm $A$ offers one unit utility more to consumers, then this is optimally achieved by lowering the fixed fee by one unit. The fixed fee is a one-to-one transfer from the consumer to the firm. Define

$$\Phi(v) \equiv \frac{m(v,v)}{\partial_{v^A} m(v,v)} \; .$$

Applying Proposition 1 of Armstrong and Vickers (2001), the firm's per customer profit in sub market $j$ in the symmetric equilibrium is given by

$$\pi_j(\hat{v}_j) = \Phi(\hat{v}_j) \; ,$$

where $\hat{v}_j$ denotes the utility offered to the consumers of type $\lambda_j$ by both firms in equilibrium. As is shown by Armstrong and Vickers, there are no asymmetric equilibria. Moreover, the equilibrium often is unique.[38] The following proposition summarizes the tariffs offered by the two firms in equilibrium.

**Proposition 7** (Full Information)**.** *Suppose that Assumption 2 holds for consumers of both groups. Then, in equilibrium,*

*(i) if $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$ both firms offer the tariff $(\hat{p}, \hat{L})$ with a positive unit price to consumers of type $\lambda_1$, and a flat-rate tariff $(0, L^F)$ to consumers of type $\lambda_2$.*

*(ii) if $c \leq \Sigma(\lambda_1) < \Sigma(\lambda_2)$, then both firms offer the flat-rate tariff $(0, L^F)$ to both types of loss-averse consumers.*

*The tariffs $(\hat{p}, \hat{L})$ and $(0, L^F)$ are characterized by: $S_1'(\hat{p}) = 0$,*

$$\begin{aligned} \hat{L} &= \Phi(\hat{v}_1) - (\hat{p} - c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_1(\theta, \hat{p}) f(\theta) d\theta \\ \textit{and} \qquad L^F &= \Phi(\hat{v}_2) + c \int_{\underline{\theta}}^{\bar{\theta}} q^S(\theta) f(\theta) d\theta \; , \end{aligned}$$

*respectively, with $\hat{p} \in (0, c]$.*

*Proof.* In order to apply Proposition 1 of Armstrong and Vickers (2001), the following three properties have to be satisfied: (i) $\partial_{v^A} m(v^A, v^B)/m(v^A, v^B)$ is non-decreasing in $v^B$, (ii) there exists $\tilde{v}_j > -\infty$ that maximizes $m(v,v)\pi_j(v)$ for $j = 1, 2$, and (iii) for $j = 1, 2$ there exists $\bar{v}_j$ defined by $\pi_j(\bar{v}_j) = 0$, $\pi_j(v) < 0$ if $v > \bar{v}_j$. Since we explicitly assumed (i) and (ii) these properties are satisfied. To see that (iii) is also satisfied note that $\bar{v}_j = \max_p\{S_j(p)\}$. Obviously, $\pi_j(\bar{v}_j) = 0$ and $\pi_j(v) < 0$ if $v > \bar{v}_j$. Hence, we can apply Proposition 1 of Armstrong and Vickers. According to this proposition, there are no asymmetric equilibria and the equilibrium utility level $\hat{v}_j$ satisfies $\hat{v}_j \in (\tilde{v}_j, \bar{v}_j)$. Since $m(v^A, v^B)\pi_j(v^A)$ is continuously differentiable, the equilibrium utility level satisfies the first-order condition of profit maximization. Thus, $\pi_j(\hat{v}_j) = \Phi(\hat{v}_j)$.

From Proposition 3 it follows that the optimal marginal price $\hat{p}_j$ is greater than zero if and only if $\Sigma(\lambda_j) < c$. If this is the case, then $\hat{p}_j$ is characterized by $S_j'(\hat{p}_j) = 0$, as was

[38] See Armstrong and Vickers (2001) for sufficient conditions for a unique equilibrium.

shown in the proof of Proposition 3. The per-customer profit of a firm is given by

$$\pi_j = L + (p - c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_j(\theta, p) f(\theta) d\theta \ . \tag{B.4}$$

Since, in equilibrium, $\pi_j = \Phi(\hat{v}_j)$ the equilibrium fixed fee is given by

$$L_j = \Phi(\hat{v}_j) - (p_j - c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_j(\theta, p_j) f(\theta) d\theta \ . \tag{B.5}$$

Replacing $p_j$ by $\hat{p}$ and 0, leads to the fixed fees $\hat{L}$ and $L^F$, respectively. □

If the degree of loss aversion of the less loss-averse consumers is below the threshold given by $\sigma(\lambda) = c$, then firms offer a measured tariff to these consumers. Next to the measured tariff, firms offer a flat-rate tariff to the more loss-averse consumers. If the degree of loss aversion of both types is above the threshold, then firms offer only a single tariff, which is a flat-rate tariff.

### *B.4. Asymmetric Information Case*

In this subsection, we investigate the tariffs offered by the two firms when facing a screening problem, i.e., when the degree of loss aversion is private information. We show that the firms can screen consumers with respect to their degree of loss aversion without costs, if $\Phi'(v) \geq 0$.[39] The main challenge is to show that consumers self-select into the right tariff if the firms offer a flat rate next to a measured tariff.

To fix ideas, suppose that $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$, so that such a menu would be optimal in the symmetric information case (see Proposition 7). Furthermore, suppose that the firms offer the same tariffs as in the symmetric information case. $\Phi'(v) \geq 0$ implies that the additional surplus generated for the less loss-averse consumers of group one, is shared between the two contracting parties. In other words, it implies $\hat{v}_1 > \hat{v}_2$ and that in equilibrium, the profit that a firm earns from a consumer of group one who subscribes to the measured tariff, is greater than the profit from a consumer from group two, who subscribes to the flat rate.

Remember that the expected utility from a flat rate is independent of the degree of loss aversion. Therefore, $\hat{v}_1 > \hat{v}_2$ immediately implies that the less loss-averse consumers of group one do not have an incentive to choose the flat rate. Conversely, we have to show that the more loss-averse consumers of group one do not have an incentive to deviate to the measured tariff. Since $\hat{v}_1 > \hat{v}_2$, we cannot simply use Lemma 2 in order to conclude that such a deviation lowers their utility. By inspecting the profit of a firm from the measured tariff, however, we observe that it decreases with demand because the unit price is below marginal cost. Since demand is decreasing in the degree of loss aversion, the profit from a (deviating) consumer of group two who subscribes to the measured tariff is higher than

[39] For instance, this condition is satisfied for the standard Hotelling model and for the logit demand model, see Section B.5 of this appendix.

the profit from a consumer from group one. Furthermore, $\Phi'(v) \geq 0$ implies that the latter profit is greater than the profit from the flat rate. Hence, a firm's profit is increased by a deviation of a consumer of group two. On the other hand, the joint surplus is decreased by the deviation—since $c \leq \Sigma(\lambda_2)$, the flat rate maximizes the joint surplus for consumers of group two. Therefore, the expected utility must decrease if a consumer from group two deviates.

**Proposition 8** (Asymmetric Information)**.** *Suppose that Assumption 2 holds for consumers of both groups and that $\Phi'(v) \geq 0$. Then,*

*(i) if $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$ both firms offering tariff $(\hat{p}, \hat{L})$ with a positive unit price to consumers of type $\lambda_1$, and flat-rate tariff $(0, L^F)$ to consumers of type $\lambda_2$ is an equilibrium.*

*(ii) If $c \leq \Sigma(\lambda_1) < \Sigma(\lambda_2)$, then in equilibrium both firms offer the flat-rate tariff $(0, L^F)$ to both types of loss-averse consumers.*

*The tariffs, $(\hat{p}, \hat{L})$ and $(0, L^F)$, are characterized in Proposition 7.*

*Proof.* Irrespective of the rival's tariff offer, if the sorting constraint is satisfied it is optimal for a firm to choose $p_j$ such that $S_j(p_j)$ is maximized. Put differently, the firm will choose the method of generating $v_j$ that maximizes its (per-customer) profits. Thus, if no type $\lambda \in \{\lambda_1, \lambda_2\}$ has an incentive to mimic the other type, it is an equilibrium that the firms offer the same tariffs as in the full information case. Obviously, in case (ii) where $c \leq \Sigma(\lambda_1) < \Sigma(\lambda_2)$, both firms offer a flat-rate tariff to consumers. In this case, a flat-rate tariff maximizes $S_1(p)$ as well as $S_2(p)$. Moreover, the generated joint surplus is the same for both types of loss-averse consumers. Since the brand preferences are i.i.d. across the $\lambda_1$ and $\lambda_2$ types, in any equilibrium each firm offers a single flat-rate tariff to consumers.

In the remaining part of the proof we show that in the case where $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$, neither type $\lambda_1$ has an incentive to choose the tariff $(0, L^F)$ nor does type $\lambda_2$ have an incentive to choose the tariff $(\hat{p}, \hat{L})$.

**Claim 1.** $\hat{v}_1 \geq \hat{v}_2$.

*Proof.* Let $S_j^* \equiv \max_p\{S_j(p)\}$. Note that $S_1(0) = S_2(0) = S_2^*$. The firm's per customer profit from type $j = 1, 2$ when offering utility $v$ is

$$\pi_j(v) = S_j^* - v \ . \tag{B.6}$$

Thus, for any $v$ it holds that $\pi_1(v) \geq \pi_2(v)$, since $S_1^* - v \geq S_2^* - v$. The equilibrium utilities are characterized by $\pi_j(\hat{v}_j) = \Phi(\hat{v}_j)$. Hence, we obtain the following relations:

$$\Phi(\hat{v}_1) = \quad \pi_1(\hat{v}_1) \geq \pi_2(\hat{v}_1) \tag{B.7}$$

$$\pi_1(\hat{v}_2) \geq \pi_2(\hat{v}_2) \quad = \Phi(\hat{v}_2) \ . \tag{B.8}$$

Suppose, in contradiction, $\hat{v}_1 < \hat{v}_2$. This immediately implies that $\pi_j(\hat{v}_1) > \pi_j(\hat{v}_2)$. Hence,

$$\Phi(\hat{v}_1) = \pi_1(\hat{v}_1) > \pi_1(\hat{v}_2) \geq \pi_2(\hat{v}_2) = \Phi(\hat{v}_2) \ . \tag{B.9}$$

Since $\Phi'(v) \geq 0$ the above formula holds only if $\hat{v}_1 > \hat{v}_2$, a contradiction. $\square$

With $\hat{v}_1 \geq \hat{v}_2$ and the expected utility from a flat-rate tariff being independent of $\lambda$, one can conclude that a consumer of type $\lambda_1$ has no incentive to choose the tariff $(0, L^F)$ that is designed for consumers of type $\lambda_2$. Finally, we show that type $\lambda_2$ has no incentive to mimic type $\lambda_1$. Let $v_2^{\text{DEV}}$ denote the expected utility of a consumer of type $\lambda_2$ who accepts the tariff $(\hat{p}, \hat{L})$ designed for type $\lambda_1$.

**Claim 2.** *$v_2^{DEV} < \hat{v}_2$.*

*Proof.* The expected utility of type $\lambda_2$ from the tariff $(\hat{p}, \hat{L})$ equals the generated joint surplus minus the profits of the firm he purchases from. Thus,

$$v_2^{\text{DEV}} = S_2(\hat{p}) - \hat{L} - (\hat{p} - c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_2(\theta, \hat{p}) f(\theta) d\theta, \tag{B.10}$$

where $\hat{q}_2(\theta, p)$ denotes the demand of type $\lambda_2$ in the personal equilibrium. Inserting the explicit formula of $\hat{L}$ into (B.10) yields

$$v_2^{\text{DEV}} = S_2(\hat{p}) - \Phi(\hat{v}_1) - (c - \hat{p}) \int_{\underline{\theta}}^{\bar{\theta}} [\hat{q}_1(\theta, \hat{p}) - \hat{q}_2(\theta, \hat{p})] f(\theta) d\theta. \tag{B.11}$$

Note that $\hat{q}_1(\theta, \hat{p}) > \hat{q}_2(\theta, \hat{p})$ for all $\theta \in \Theta$, since $\partial_\lambda \hat{q} < 0$ if $p > 0$. By Proposition 7, $c \geq \hat{p}$, and hence

$$v_2^{\text{DEV}} < S_2(\hat{p}) - \Phi(\hat{v}_1). \tag{B.12}$$

The expected utility of a consumer of type $\lambda_2$ when choosing the tariff that is intended for him can be expressed as follows,

$$\hat{v}_2 = S_2^* - \Phi(\hat{v}_2). \tag{B.13}$$

Hence, a deviation is not utility improving if

$$S_2^* - \Phi(\hat{v}_2) \geq S_2(\hat{p}) - \Phi(\hat{v}_1) \tag{B.14}$$

$$\iff [S_2^* - S_2(\hat{p})] + [\Phi(\hat{v}_1) - \Phi(\hat{v}_2)] \geq 0\ . \tag{B.15}$$

The above inequality is satisfied since $\Phi'(\cdot) \geq 0$ and $\hat{v}_1 \geq \hat{v}_2$. $\square$

Thus, if the firms offer the optimal tariffs of the full information case, each type of loss-averse consumer selects the tariff that is designed for him, which completes the proof. $\square$

As in the symmetric information case, if $\lambda_1$ is below and $\lambda_2$ is above the threshold given by $\sigma(\lambda) = c$, then firms offer a measured tariff to the less loss-averse types and a flat-rate tariff to the more loss-averse consumers. The fixed fee of the flat-rate tariff is higher than the fixed fee of the measured tariff. In this case, we do not make any claims about the uniqueness of this equilibrium.[40] If the degree of loss aversion of both types exceeds the threshold, then we obtain a pooling equilibrium: each firm offers only a single tariff that is accepted by both types of consumers.

### *B.5. Examples of Discrete Choice Models*

*Hotelling Model with Linear Transport Cost.*—Suppose consumers' ideal brands are uniformly distributed on the unit interval $[0, 1]$. The brands of the two firms, $A$ and $B$, are located at the two extreme points, brand $A$ at zero and brand $B$ at one. A consumer with ideal brand $x \in [0, 1]$ has brand preferences $\zeta = (0, -tx, -t(1 - x))$. The parameter $t > 0$ is a consumer's "transport cost" per unit distance between his ideal brand and the brand he purchases from. For the Hotelling specification, the market share function takes the following form,

$$m(v^A, v^B) = \min\left\{\frac{1}{2t}(t + v^A - v^B), \frac{v^A}{t}\right\}. \tag{B.16}$$

The market share function has to be modified if $v^A$ and $v^B$ differ by so much that $m(\cdot) \notin [0, 1]$ (this never happens in equilibrium). Moreover, the Hotelling model has the well-known drawback that market shares are kinked. If, however, the transport cost is sufficiently low, then one can focus on the case where the market share function is given by the first term of the above expression and thus well behaved. Formally, for $t \leq (2/3)S_2^*$ it suffices to analyze firms' profit maximization problem for[41]

$$m(v^A, v^B) = \frac{1}{2t}(t + v^A - v^B). \tag{B.17}$$

Hence, $\partial_{v^A} m(v^A, v^B) = (2t)^{-1}$ which immediately implies that

$$\Phi(v) \equiv \frac{m(v, v)}{\partial_{v^A} m(v, v)} = t. \tag{B.18}$$

Obviously, $\Phi(\cdot)$ is non-decreasing. Note that

$$\frac{\partial_{v^A} m(v^A, v^B)}{m(v^A, v^B)} = \frac{1}{t + v^A - v^B}. \tag{B.19}$$

It can easily be seen that the above fraction is increasing in $v^B$. Thus, the Hotelling model satisfies all imposed assumptions if the transport cost is sufficiently low. One can check that the collusive utility level exists. To calculate the collusive utility level one has to use the market share function given in (B.16).

*Logit Demand Model.*—An obvious drawback of the Hotelling specification is that a firm does not compete with the rival and the outside option at the same time. A model

[40] To analyze all equilibria we cannot apply the competition in utility space framework, since we have to take the sorting constraints explicitly into account.

[41] See Lemma 1 of Armstrong and Vickers (2001).

that accounts for this simultaneous competition on two fronts is the logit demand model. Here, a consumer's brand preferences $\zeta^i$ for $i = 0, A, B$ are i.i.d. according to the double exponential distribution with mean zero and variance $\mu^2\pi^2/6$, where $\pi$ (here) denotes the circular constant. Thus, the cumulative distribution function is

$$G(\zeta^i) = \exp\big\{-\exp[-(\gamma + \zeta^i/\mu)]\big\}, \tag{B.20}$$

where $\gamma$ is the Euler–Mascheroni constant and $\mu$ is a positive constant. With this specification, the market share of firm $A$ is given by (see Anderson et al., 1992)

$$m(v^A, v^B) = \frac{\exp[v^A/\mu]}{\exp[v^A/\mu] + \exp[v^B/\mu] + 1}. \tag{B.21}$$

The parameter $\mu$ captures the degree of heterogeneity among consumers with respect to their brand preferences. Put differently, $\mu$ measures the degree of product differentiation. A lower value of $\mu$ corresponds to a more competitive market. For $\mu \to \infty$ the firms are local monopolists. Taking the partial derivative of (B.21) with respect to $v^A$ yields

$$\partial_{v^A} m(v^A, v^B) = \frac{\exp[v^A/\mu]\{\exp[v^B/\mu] + 1\}}{\mu\{\exp[v^A/\mu] + \exp[v^B/\mu] + 1\}^2}. \tag{B.22}$$

Thus,

$$\frac{m(v^A, v^B)}{\partial_{v^A} m(v^A, v^B)} = \frac{\mu\{\exp[v^A/\mu] + \exp[v^B/\mu] + 1\}}{\exp[v^B/\mu] + 1}. \tag{B.23}$$

Evaluating the above expression at $v^A = v^B = v$ leads to

$$\Phi(v) = \mu \frac{2\exp[v/\mu] + 1}{\exp[v/\mu] + 1}. \tag{B.24}$$

Taking the derivative of $\Phi(\cdot)$ with respect to $v$ yields

$$\Phi'(v) = \frac{\exp[v/\mu]}{(\exp[v/\mu] + 1)^2} > 0. \tag{B.25}$$

Moreover, the derivative of $\partial_{v^A} m(v^A, v^B)/m(v^A, v^B)$ with respect to $v^B$ amounts to

$$\frac{d}{dv^B}\left[\frac{\partial_{v^A} m(v^A, v^B)}{m(v^A, v^B)}\right] = \frac{1}{\mu^2}\,\frac{\exp[v^B/\mu]\{\exp[v^B/\mu] + 1\}}{\mu\{\exp[v^B/\mu] + \exp[v^B/\mu] + 1\}^2} > 0. \tag{B.26}$$

The collusive utility level $\tilde{v}$ maximizes $m(v, v)\pi(v)$. Note that $m(v, v) \to 0$ for $v \to -\infty$ and $\pi(v) \leq 0$ if $v \geq \max_p\{S(p)\}$. Thus, the collusive utility exists, since $m(v, v)\pi(v)$ is continuously differentiable.

## C. Supplementary Material to Section 7.3: Loss Aversion in Both Dimensions

*Proof of Proposition 6.* The monopolist maximizes the expected joint surplus by choosing the unit price $p$. Given that the consumer plays the personal equilibrium $\langle \hat{q}(\theta, p)\rangle$

characterized by (16), the joint surplus is given by

$$S(p) = \int_{\underline{\theta}}^{\bar{\theta}} \Big\{ u(\hat{q}(\theta,p),\theta) - c\hat{q}(\theta,p) - \lambda \int_{\underline{\theta}}^{\theta} \big[u(\hat{q}(z,p),z) - u(\hat{q}(\theta,p),\theta)\big] f(z)\, dz$$
$$- \lambda \int_{\underline{\theta}}^{\theta} p\big[\hat{q}(\theta,p) - \hat{q}(z,p)\big] f(z)\, dz \Big\} f(\theta)\, d\theta. \quad \text{(C.1)}$$

Taking the derivative of $S(p)$ with respect to $p$ and using the personal equilibrium condition (16), we obtain

$$S'(p) = (p-c)\int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(p,\theta) f(\theta) d\theta - \lambda \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [\hat{q}(p,\theta) - \hat{q}(p,z)] f(z) f(\theta) dz d\theta \quad \text{(C.2)}$$

The optimal unit price fulfills $p^* \in [0,c)$, since $S(p)$ is decreasing for prices above the marginal cost. If the consumer is not loss averse, i.e., $\lambda = 0$, then $p^* = c$. Given that $\partial_{qqq} u(q,\theta) \geq 0$ and $\lambda \leq 1$, the joint surplus is a strictly concave function for $p < c$. Formally,

$$S''(p) = -(p-c)\int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial_{qqq} u(\hat{q}(\theta,p),\theta)}{[\partial_{qq} u(\hat{q}(\theta,p),\theta)]^3} f(\theta) d\theta + \int_{\underline{\theta}}^{\bar{\theta}} \frac{1-\lambda[2F(\theta)-1]}{\partial_{qq} u(\hat{q}(\theta,p),\theta)} f(\theta) d\theta < 0, \quad \text{(C.3)}$$

for $p < c$. With the joint surplus being strictly concave, a flat-rate tariff is optimal when $S'(p)|_{p=0} \leq 0$, which is equivalent to

$$c \leq \lambda \frac{\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [q^S(\theta) - q^S(\phi)] f(\phi) f(\theta) d\phi d\theta}{\int_{\underline{\theta}}^{\bar{\theta}} \partial_p \hat{q}(\theta,p)\; f(\theta) d\theta} \equiv \Sigma(\lambda). \quad \text{(C.4)}$$

□

**Construction of a Personal Equilibrium** Since a higher demand type is associated with a stronger need for the good, we posit that demand is increasing in the type. For a given quantity, higher types are worse off compared to lower types. We posit that this is still the case in the personal equilibrium. Put differently, the increase in intrinsic utility due to a higher consumption of a higher type does not outweigh the direct negative effect on intrinsic utility of being a higher type. Formally, the second hypothesis with respect to the personal equilibrium requires that the following inequality is satisfied:

$$\partial_q u(\hat{q}(\theta,p),\theta) \times \partial_\theta \hat{q}(\theta,p) + \partial_\theta u(\hat{q}(\theta,p),\theta) \leq 0. \quad \text{(C.5)}$$

Given these hypotheses, the consumer's utility can be written as

$$U(q|\theta,\cdot) = u(q,\theta) - pq - L - \lambda \int_{\underline{\theta}}^{\beta(q)} \big[u(\hat{q}(z,p),z) - u(q,\theta)\big] f(z)\, dz$$
$$- \lambda \int_{\underline{\theta}}^{\alpha(q)} p\big[q - \hat{q}(z,p)\big] f(z)\, dz, \quad \text{(C.6)}$$

where $\alpha(q)$ and $\beta(q)$ are implicitly defined by[42]

$$\hat{q}(\alpha(q), p) \equiv q \quad \text{and} \quad u(\hat{q}(\beta(q)), \beta(q)) \equiv u(q, \theta),$$

respectively. Under the hypotheses, $\alpha'(q) > 0$ and $\beta'(q) < 0$. Differentiating (C.6) with respect to $q$ yields

$$U'(q|\theta) = \partial_q u(q, \theta)[1 + \lambda F(\beta(q))] - p[1 + \lambda F(\alpha(q))]. \tag{C.7}$$

The utility function is strictly concave and thus the first-order condition is necessary and sufficient for optimality. Moreover, in equilibrium it has to hold that $\alpha(q) = \beta(q) = \theta$. Hence, the personal equilibrium is characterized by $\partial_q u(\hat{q}(\theta, p), \theta) = p$. Obviously, the demand function characterized by (16) is increasing in the demand type. The second hypothesis (C.5)—higher types achieve lower utility levels in equilibrium—is also satisfied for relatively low marginal prices, i.e., for

$$p \leq \partial_{qq} u(q, \theta) \frac{\partial_\theta u(q, \theta)}{\partial_{q\theta} u(q, \theta)}. \tag{C.8}$$

**(Personal) Equilibrium Selection** A final comment to the personal equilibrium selection is in order. There may exist multiple personal equilibria for the case analyzed above. So far, we constructed only one personal equilibrium. As outlined in the text it is reasonable to assume that higher types demand more. Higher types have a higher marginal utility which implies that a higher $q$ increases the intrinsic utility in the good dimension and reduces the loss in the good dimension more for a higher than for a lower type. Moreover, it is reasonable to assume that higher types do not consume that much more such that they achieve a higher intrinsic utility than lower types. Given a personal equilibrium has to satisfy these features what would be the ex ante optimal plan—the choice acclimating personal equilibrium (CPE)? Using integration by parts, the consumer's ex ante expected utility can be written as

$$EU = \int_{\underline{\theta}}^{\bar{\theta}} \Big\{ u(q(\theta), \theta)[1 + \lambda(2F(\theta) - 1)] - p[1 + \lambda(2F(\theta) - 1)] \Big\} f(\theta)\, d\theta. \tag{C.9}$$

As it is well-known in the literature, with CPE a decision maker is highly risk averse and may prefer stochastically dominated options. This behavior can be ruled out by assuming that loss utility is less important than intrinsic utility, i.e., $\lambda \leq 1$. For $\lambda \leq 1$, the expected utility is strictly concave in $q(\theta)$ for all $\theta$. The first-order condition obtained from point wise maximization is

$$\partial_q u(q(\theta), \theta) = p. \tag{C.10}$$

Thus, at least for $\lambda \leq 1$ the demand function (16) is the ex ante preferred plan among all plans where higher types consume more but still achieve a lower intrinsic utility level.

[42]Strictly speaking, $\beta(q) = \beta(q, \theta)$

## D. Comparison to the loss function of Kőszegi and Rabin (2006, 2007)

For applications, Kőszegi and Rabin (2006, 2007) propose the following gain-loss function:

$$\tilde{\mu}(x) = \begin{cases} \tilde{\eta}x, & \text{if } x \geq 0, \\ \tilde{\eta}\tilde{\lambda}x, & \text{if } x < 0. \end{cases}$$

The parameter $\tilde{\eta} \geq 0$ is the degree of reference dependence and $\tilde{\lambda} \geq 1$ is the degree of loss aversion. ($\tilde{\lambda} = 1$ means that the consumer is not loss averse.)

With loss utility only in the money dimension and reference-dependent preferences ($\tilde{\eta} > 0$), this formulation has the drawback that at the contracting stage, the marginal rate of substitution (MRS) between money and consumption differs from the MRS at the consumption stage. To see this, we rewrite $\tilde{\mu}$ as

$$\tilde{\mu}(x) = \tilde{\eta}x + \begin{cases} 0, & \text{if } x \geq 0, \\ \tilde{\eta}(\tilde{\lambda} - 1)x, & \text{if } x < 0. \end{cases} \tag{D.1}$$

Consider a consumer who is not loss averse ($\tilde{\lambda} = 1$) but has reference-dependent preferences ($\tilde{\eta} > 0$). Ex-post, his marginal utility of money is $1+\tilde{\eta}$. Ex ante, however, gains and losses cancel in expectation because the first part of the gain-loss function in (D.1) is linear. Therefore, ex ante, the marginal utility of money is 1. This time-inconsistency arises because we restrict reference dependent utility to the money dimension. If preferences were also reference dependent in the good dimension, utility from consumption would also be multiplied by $1 + \tilde{\eta}$ ex-post, so that the marginal rate of substitution between money and the good would remain unchanged and equal to the ex-ante MRS.

To avoid the shift in the MRS we use the following modified loss function:[43]

$$\mu(x) = \begin{cases} 0, & \text{if } x \geq 0, \\ \lambda x, & \text{if } x < 0. \end{cases}$$

If we set $\lambda = \tilde{\eta}(\tilde{\lambda} - 1)$, this corresponds to the second part of equation (D.1). This formulation eliminates reference dependence for consumers that are not loss averse because the first part of equation (D.1) was dropped. The ex ante expected utility of a consumer is unchanged because it only depends on $\tilde{\eta}(\tilde{\lambda} - 1)$ (Compare equation (5) in this paper with the same equation in an older working paper version, Herweg (2010)). The same argument applies to the ex ante expected joint surplus. Therefore, the condition for the optimality of a flat rate remains qualitatively unchanged between the different formulations.[44]

Since we do not want to model time-inconsistency, the new formulation which holds the MRS constant, is the natural choice. Also, this formulation is closer to the original

[43] Since only losses matter, we use $-\mu$ instead of $\mu$ in the main text. This is ignored here to facilitate the comparison with the formulation of Kőszegi and Rabin (2006, 2007).

[44] Quantitatively, the conditions differ since optimal demand at the consumption stage is depressed in the $\tilde{\eta}$-$\tilde{\lambda}$-formulation compared to the $\lambda$-formulation, because of the different MRS.

formulation of Kőszegi and Rabin (2006, 2007) with loss aversion in both dimensions, because this formulation also has a stable MRS.

## References


S. Anderson, A. dePalma, and J. F. Thisse. *Discrete Choice Theory of Product Differentiation.* MIT Press, Cambridge, 1992.

M. Armstrong and J. Vickers. Competitive price discrimination. *The RAND Journal of Economics*, 32(4):579–605, 2001.

F. H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, New York, 1983.

F. Herweg. Uncertain demand, consumer loss aversion and flat-rate tariffs. SFB TR/15 Discussion Paper Nr. 330, 2010.

B. Kőszegi and M. Rabin. A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121:1133–1165, 2006.

B. Kőszegi and M. Rabin. Reference-dependent risk preferences. *American Economic Review*, 97:1047–1073, 2007.

P. Milgrom. *Putting Auction Theory to Work.* Cambridge University Press, 2004.

J.-C. Rochet and L. A. Stole. Non-linear pricing with random participation. *Review of Economic Studies*, 69(1):277–311, 01 2002.