Shrouded Transaction Costs*

Hélène Bourguignon[†] **Renato Gomes**[‡] **Jean Tirole**[§]

September 2014

Abstract: The proliferation of new payment methods on the Internet rekindles the old and unsettled debate about merchants' incentive and ability to differentiate price according to payment choice. This paper develops an imperfect-information framework for the analysis of platform and social regulation of card surcharging and cash discounting. It makes three main contributions. First, it identifies the conditions under which concerns about missed sales induce merchants to perceive that they must take the card. Second, it derives a set of predictions about cash discounts, card surcharges and platform fees that match, and shed light on existing evidence. Finally, it studies regulation when surcharging is allowed, showing that (i) capping merchant fees reduces welfare if surcharges are unconstrained, (ii) recent reforms to cap surcharges at or above merchant cost ignore the merchant's benefit from card payments and are accordingly too lenient; indeed surcharges should be banned when the merchant fee is regulated optimally.

Keywords: Payment cards, cash discounts, card surcharges, hold-ups in two-sided markets, missed sales.

JEL numbers: D83, L10, L41.

*The authors are grateful to Jean-Charles Rochet and Julian Wright, and to seminar participants at EEA-ESEM 2014 (Toulouse) and EARIE 2014 (Milan), for very helpful comments. Renato Gomes and Jean Tirole are members of IDEI, which receives financial support from La Banque Postale. Hélène Bourguignon was employed by Banque Postale during most of the period over which this research was performed. Jean Tirole also gratefully acknowledges financial support from the ERC grant FP7/2007-2013 No. 249429 "Cognition and Decision-Making: Laws, Norms and Contracts".

[†]MAPP, hbourguignon@mapp-economics.com

[‡]Toulouse School of Economics and CEPR, renato.gomes@tse-fr.eu.

[§]Toulouse School of Economics and IAST, jean.tirole@tse-fr.eu.

1 Introduction

The payment industry is rapidly evolving. The proliferation of new payment methods on the Internet, such as PayPal, Google Checkout, Amazon Payments and Bitcoin, challenges the business and regulatory models of traditional card payment systems, such as Visa, MasterCard and American Express. An important dimension of this reconsideration concerns the extent to which merchants should be allowed to price discriminate according to payment method. Absent public regulation on the matter, payment systems have prohibited merchants from surcharging for payments methods other than cash, so as to favor consumers at the expense of merchants, who often face high fees.[1]

Merchants have long complained that their inability to pass through high merchant fees to consumers interfered with the proper working of a free market. Over the last decade, most regulators[2] sided with the merchants and struck down the uniform-pricing[3] mandate imposed by card systems, only to discover a few years later that substantial surcharges may end up being levied on unfortunate consumers. In response to abusive surcharging, the European Union, the UK and Australia all have proposed variations on the idea that surcharges should be limited to some variant of the notion of "cost of acceptance", which includes the merchant fee plus possible various other costs.[4]

Meanwhile, most regulators have capped the fees that card systems can levy on merchants. For example, the European Commission adopted as its benchmark for the regulation of merchant fees charged by Visa and MasterCard members the "tourist or avoided cost test", according to which the merchant fee should not exceed the merchant's convenience benefit of a card payment.[5] The tourist test induces the cardholder to internalize the merchant's welfare

[1]For instance, PayPal, Google Checkout and Amazon Payments charge merchant fees of 2.9% of the retail price plus $0.3 per transaction for merchants with low volume on the payment platform. Discounts are granted to merchants who actively use the payment platform.

[2]Australia pioneered this regulation in 2003. A key provision of a long-standing litigation brought by a class of US retailers in 2005 was to force Visa to allow merchants to surcharge; since January 2013, merchants in the United States are permitted to impose a surcharge on consumers when they use a credit card, except in states with laws prohibiting surcharging, such as California, New York and Massachusetts.

[3]Technically, all platforms allow cash discounts. As we will demonstrate in the paper, though, this degree of freedom is unlikely to generate departures from uniform pricing.

[4]For example, in the UK surcharge regulation, the attributable costs can include direct costs beyond the merchant service charge, such as point of sale devices, risk management, charges for reversing or refunding a payment, or payments for services from intermediaries who provide equipment, fraud detection and processing services for card payments.

[5]That is, conditional on the customer being at the point of sale and willing to pay regardless of the payment method, the merchant is happy to have the customer pay by card. The reference to a tourist captures the idea that there is no repeat purchase, and so accepting the card does not bring about any benefit from the customer's being more willing to return to the store because he now knows that cards are accepted. The test was developed in Rochet and Tirole (2011). Note also that in the case of Visa and MasterCard, the merchant fee is regulated only indirectly through the so-called "interchange fee" that the merchant's bank pays to the cardholder's bank. The interchange fee is set by the platform (e.g., Visa or MasterCard). For extensive overviews of the economics of interchange fees, see Evans and Schmalensee (2005) and Evans (2011).

when choosing the payment method and is thus the incarnation of the Pigovian precept in the payment context. Because the merchants' surcharging behavior ultimately depends on the fees they pay on card payments, the merchant fee and surcharging regulations should not be designed separately, unlike under current practice.[6]

The purpose of this paper is to clarify the surcharging debate and determine how the regulation of surcharges should be related to public policy toward merchant fees. To this purpose, we develop a micro-founded framework for the study of payment-method-based price discrimination. In contrast with the existing literature, we assume that consumers are imperfectly informed about the merchant's "payment policy" (card acceptance, card surcharge, cash discount). Accordingly, the transaction costs per sale are shrouded from consumers' sight (hence the title of the paper).

We obtain three main insights:

1) We show that concerns about missed sales induce merchants to perceive that they must take the card. By assuming perfect consumer information about merchant payment policy, the economics literature has ignored the possibility that card refusal or a high surcharge may cause the merchant to lose business at the point of sale, a widespread concern. Besides introducing missed sales in the literature, the paper also shows that, in the absence of surcharging, merchants will accept cards with inefficiently high merchant fees, and that payment systems will indeed want to levy such fees. The "must-take-card" argument has been the centerpiece for many antitrust lawsuits against Visa and MasterCard. We here provide a different and complementary rationale for the already existing must-take card argument according to which, under complete consumer information, the payment platform can charge the merchant for the increased attractiveness stemming from card acceptance.[7]

2) Second, we derive a set of predictions that match, and shed light on existing evidence on merchant comparative resistance to merchant fees for debit and credit cards, on the pattern of surcharging when allowed (credit/debit, Internet/brick and mortar, business/personal cards), and on how platforms adjust their fee structures in response to laws allowing merchants to practice cash discounts/card surcharges. The theoretical predictions and the available evidence are compared in the conclusion.

3) Finally, we show that the optimal regulation of surcharging is related to public policy toward merchant fees and substantially differs from current practice. First, we note that prohibiting surcharging increases welfare if the merchant fee lies above some threshold that exceeds the tourist test level, and decreases it otherwise; in particular allowing surcharging is very likely to reduce welfare if merchant fees are otherwise capped. We then show that the optimal regulation consists in allowing surcharges but capping them at the merchant fee minus

[6] An exception is the European Commission's July 2013 proposal to allow surcharging for cards which fee structure is currently *not* subject to regulation, and to ban it for the others.

[7] See Rochet and Tirole (2002), Vickers (2005), Rochet and Tirole (2011) and Wright (2012).

the merchant's convenience cost of card payment; thus the cap should be equal to zero if the merchant fee is regulated according to the tourist test.

Organization of the paper

Section 2 sets up the model. The payment platform charges a merchant fee and a cardholder fee per card transaction. The merchant's listed price is just a commitment to sell the good at that price through some payment method (cash or card). Customers differ in their relative cost of a cash payment, and choose a payment method as a function of the platform's cardholder fee and the merchant's payment policy. A crucial ingredient of our analysis is that consumers are imperfectly informed about the merchant's cash/card policy (and, in equilibrium, form rational expectations about it). This assumption is natural in industries characterized by one-time (or infrequent) shopping, such as travel (e.g., airlines and gas stations), tourism, and various retail/service sectors (e.g., durable goods). More broadly, this assumption is likely to hold whenever consumers have bounded memory or are rationally inattentive regarding past purchasing experiences.

Section 3 considers the benchmark of uniform pricing, where neither cash discounts nor card surcharges are permitted, and develops the notions of missed sales and must-take cards in our imperfect information setting. A missed sale occurs when the customer is in the shop and eager to buy, but has a high inconvenience cost of paying by cash, and is discouraged by either a high card surcharge or an outright rejection of the card.

Section 3 then relates the missed sales concern, and the concomitant low merchant resistance to high merchant fees, to the nature of the demand curve. An inelastic demand for the good calls for a high markup, and therefore makes the merchant particularly wary of missed sales. We characterize the maximum merchant fee that the merchant is willing to accept. Importantly, this *uniform pricing threshold* violates the tourist test and under weak conditions is selected by the platform.

Section 4 studies the merchant's optimal payment-method-based price discrimination (cash discount and card surcharge). Because consumers imperfectly observe the merchant's payment policy, a cash discount is a giveaway to consumers who already are in the shop. In contrast, a card surcharge holds up the consumer, who has made the specific investment to come to the store and inspect wares. While card acceptance is a no-brainer for a merchant whose surcharging behavior is unconstrained, it is not so when only cash discounts are permitted. In this case, steering customers away from using the card is costly for the merchant, and, as a result, merchants may refuse cards if merchant fees are sufficiently high. Yet, because cash discounts improve the merchant's ability to avoid card fees, the card acceptance threshold is higher than under uniform pricing.

The merchant's optimal cash discount (card surcharge) balances the marginal revenue from steering customers to pay by cash and the marginal cost (gain) directly generated by the discount (surcharge). A cash discount (card surcharge) is optimal for the merchant if and only if the merchant fee exceeds a second, *price discrimination threshold*. The price discrimination threshold is lower for a card surcharge, which brings in additional revenue to the seller, than for a cash discount, which benefits customers. As a result, card surcharges occur even when the tourist test is met, while cash discounts only occur if the merchant fee is sufficiently above the tourist test level.

Price discrimination then leads us to define *effective* merchant and cardholder fees, which are the fees that merchants and consumers effectively pay once the merchants' discount or surcharge are factored in. Because merchants use a card surcharge as a price discrimination device (screening consumers with different costs of cash payments), the merchant always "overshoots" in the surcharge, leading to an effective merchant fee that strictly satisfies the tourist test. In contrast, because a cash discount is a costly gift to consumers, the merchant always "undershoots" in the discount, leading to an effective merchant fee that strictly violates the tourist test. Furthermore, if given the choice, the merchant *always* prefers a card surcharge to a cash discount.

Section 4 also shows that, *fixing* the platform's fee structure, allowing cash discounts raises welfare. By contrast, the welfare impact of card surcharges depends on the level of the merchant fee; if the latter satisfies the tourist test or even lies reasonably above the merchant's convenience benefit, allowing card surcharge reduces welfare by excessively reducing card usage. For very high merchant fees by contrast, allowing card surcharges improves welfare.

Section 5 then studies how the platform optimally adjusts its fee structure in response to laws allowing merchants to practice cash discounts/card surcharges. We show that the *possibility* of cash discounts/card surcharges reduces the effective merchant fee and increases the cardholder fee (relative to uniform pricing). The magnitude of these effects is strictly higher in the case of surcharging, as a result of the "abusive" surcharges practiced by merchants. In particular, when surcharging is allowed, the effective merchant fee in equilibrium is below the tourist test level, and card usage is inefficiently low.

Interestingly, we show that the platform is able to achieve maximal card usage by selecting *nominal* merchant and cardholder fees that generate no card surcharges on the equilibrium path. Arguably, such fees should be the ones selected by the platform whenever surcharges involve any extra convenience cost for merchants. This implication is consistent with ample evidence documenting that public regulations authorizing card surcharging do not generate much actual surcharging.

Section 6 derives the optimal regulation of surcharging and shows that a surcharge cap set at or above the merchant fee is too lax. It derives the optimal cap. Section 6 then studies the impact of a recent proposal for mandated transparency by the merchant. According to this regulation, the merchant would be required to post the surcharge(s) together with the price. Mandated transparency regulation eliminates hold-ups for attentive customers; however, it (i) may not be feasible (the consumer learns price through a national advertising campaign or a price comparison engine), (ii) may involve transaction costs for the merchant, (iii) does not address the existence of inattentive consumers, and (iv) does not prevent inefficient surcharging if the consumer's willingness to pay is correlated with his desire to use the card (as we argue is likely to be the case). Thus transparency is not a perfect regulatory response to the inefficiencies attached to surcharging, and the two regulations may well be complementary.

Section 7 offers two extensions/robustness checks. Section 7.1 assumes that consumers are naïve about the existence of surcharges; the case for the regulation of surcharging is then even stronger than with rational consumers. Section 7.2 endogenizes the consumer's cash holdings. Intuitively, an increase in the anticipated card surcharge makes cash holdings more attractive; in turn, higher cash holdings limit the merchant's incentive to surcharge.

Section 8 summarizes the main insights and policy implications and concludes with evidence consistent with various predictions of the theory. Proofs that are missing in the text can be found in the technical appendix.

Related Literature

Our study of surcharging is related to the literatures on add-ons (Shapiro 1995, Ellison 2005) and on shrouded attributes and incomplete contracting (Gabaix and Laibson 2006, Tirole 2009, Heidhues et al 2012, 2014, Grubb 2014). In these one-sided market contributions, competing sellers may conceal the prices of add-ons when consumers are naïve or unaware of the existence of the add-on, that is, either do not think about the possibility of ex-post hold up, or are inattentive in that they fail to take measures (such as tracking their bank balance to avoid overdrafts) that protect them from seller hold ups.

A large literature looks at add-ons or tied aftermarket sales in environments mixing sophisticated and naïve customers. The emphasis in this literature, reviewed by Armstrong and Vickers (2012), is on externalities between the two groups of consumers. Sometimes, sophisticated users benefit naïve ones; they may for instance exert downward pressure on prices. Conversely, they may be cross-subsidized by naïve users, as when competitive merchants charge low base fees and break even through high ex-post mark-ups paid by the unsophisticated. Our paper allows for various levels of rationality/attention levels (see Section 7), but does not investigate redistributive consequences under a mixed population of users. Nor does it analyze the consumers' incentives to become informed, or forms of regulation that aim at educating naïve consumers.[8]

While the impact of add-ons in one-sided markets is now well-understood, little is known regarding add-ons in two-sided markets. In a two-sided market, the platform intermediating transactions between the buyer and the seller can affect the seller's behavior, indirectly through the fees charged for using the platform's services, or directly by regulating the pricing of the

[8] See, e.g., Kostfeld and Schüwer (2011) and Armstrong et al (2009 a,b) for theoretical analyses of the effects of regulatory intervention via educating naïve consumers. For an empirical assessment, see Agarwal et al (2014).

add-on. This yields a richer and more complex environment for the study of add-ons, leading to business strategy and regulatory insights that substantially differ from those gleaned from one-sided market theory. Another difference with the earlier literature is that in the latter missed sales concerns do not arise in the following sense: The seller first sells the basic good and then may or may not sell the add-on. In our paper, by contrast, the price of the add-on (the card transaction) is learned before the sale of the basic good is completed. Hence the basic good itself may not be traded because of excessive add-on prices.

The relevance of merchant fees under cash discounts or card surcharges in our paper contrasts with the neutrality result of Carlton and Frankel (1995), Rochet and Tirole (2002) and, in more generality, Gans and King (2003). These papers have shown that, under perfect consumer information, the choice of merchant fee has no impact on the real allocation. Imperfect information generates imperfect passthrough, which explains the difference in results.

Another contribution to the payments literature is to bring life to common policy issues that have eluded previous analysis. A case in point is missed sales, which is a prominent concern for merchants. Because perfectly informed consumers would never come to the shop and not purchase the item, missed sales have not been accounted for in the economics literature.[9] Similarly, we endogenize cash holdings and thereby the distribution of inconvenience costs of cash payments. Finally, the current literature provides limited guidance to assess the conventional wisdom according to which allowing surcharging is an antidote to abusive merchant fees.[10]

The recurrent lawsuits against card associations are predicated on the "must-take-card" argument, which merchants cannot turn down the card even when the merchant fee exceeds the merchant's card convenience benefit (Vickers 2005). The standard modeling of the must-take-card argument is based on the idea that card acceptance makes the merchant more attractive to consumers, so that payment platforms can charge a merchant fee beyond the socially efficient level (Rochet and Tirole 2002, Tirole 2011 and Wright 2012). The attractiveness channel hinges on the consumers being perfectly informed about the merchant's cash/card policy. By con-

[9] Chakravorti and To (2007) study a related issue; namely, that credit cards enable liquidity-constrained consumers to make purchases for which they cannot pay with their current income. In their perfect-information framework, missed sales do not occur, as consumers would never visit merchants who do not take cards. Relatedly, Rochet and Wright (2010) assume that some customers need to use credit to purchase, which can be provided by the merchant (in the form of store credit). Because consumers are perfectly informed about the merchant cash/card policy, there cannot be missed sales; rather, missed sales concerns motivate the offering of store credit by the merchant. The must-take card channel in Rochet-Wright is the attractiveness channel rather than the missed-sales channel emphasized in our paper.

[10] In a consumer-full-information environment, Proposition 6 in Rochet and Tirole (2002) shows that, besides making the interchange fee irrelevant, surcharging leads to an underuse of cards and may increase or decrease welfare; this reflects a horse race between two competitive distortions: the overuse of cards induced by the platform's exploitation of the must-take nature of the card, and the underuse of cards due to issuer market power; the former distortion can be eliminated by allowing surcharging; the second can no longer be corrected through the interchange fee under surcharging as the interchange fee is then made neutral. One driver of the Rochet-Tirole result – market power on the issuer side – plays no role in this paper; while the other driver, their must-take card argument (merchant attractiveness under perfect consumer information), is complementary to ours (missed sales under imperfect consumer information).

sidering the other polar case of uninformed consumers, this paper completely shuts down the attractiveness channel.[11] It therefore contributes to the literature by identifying a novel (and independent) channel for must-take cards: the merchant's concern about missed sales.[12]

Other papers have investigated the impact of surcharging card transactions. Assuming perfect consumer information about the merchant's payment policy and a perfectly inelastic demand for the good, Wright (2003) concludes that lifting the no-surcharge rule decreases social welfare, and may completely shut down the card payment network. By contrast, among other differences, we assume that the demand for the good is elastic, and that consumers are imperfectly informed about the merchants' payment policy (what generates the missed sales concerns described above). This leads to different results: We show that allowing for card surcharging is welfare-enhancing if and only if the merchant fees are high. We also show that lifting the no-surcharge rule, while properly regulating the merchants' behavior through surcharging caps, can lead to substantive welfare improvements relative to uniform pricing.

Schwartz and Vincent (2006) consider a model with fixed populations of cash and card users. Their analysis emphasizes the role of the no-surcharge rule in preventing cross-subsidization from cash to card users. Lifting the no-surcharge rule is shown to increase social welfare provided the fraction of cash users is sufficiently small. By assuming that the choice of payment method is exogenous, their model ignores the cash/card steering effects that are central to our analysis. More recently, Edelman and Wright (2014) consider the problem of an intermediary (such as a booking website, a financial brokerage firm, or a card payment platform) that may invest to provide a non-pecuniary benefit to consumers. They show that, under uniform pricing, the intermediary over-invests in the provision of the non-pecuniary benefit, and can actually decrease social welfare. By assuming perfect consumer information, these works abstract from missed sales and the hold-up issues that lie at the core of the present article.[13]

[11] As we point out in Section 3, in the absence of missed sales, merchants refuse cards if the merchant fee exceeds the merchant's card benefit (i.e., cards are no longer "must-take-cards"). In this case, the merchants' card acceptance behavior is as in Baxter (1983) – see also Wright (2004).

[12] In Bedre-Defolie and Calvano (2013), consumers make membership and usage decisions, while merchants only make membership decisions. The asymmetry in decision-making between consumers and merchants generates inefficiently high merchant fees. Yet, in their model, cards are not must-take cards, as merchants accept cards if and only if the average benefit of card usage exceeds the merchant fee.

[13] The same applies to the contribution of Economides and Henriques (2011), who investigate the impact of the no-surcharge rule in a model with competing payment platforms. Taking a reduced-form approach to consumer demand for cards (à la Schmalensee 2002), they conclude that the no-surcharge rule increases social welfare if the network externality exerted by merchants on cardholders is sufficiently weak.

2 Model and preliminaries

2.1 Description

Consumers contemplate purchasing a good, which they may pay for either in cash or by card. Consumers differ in their valuation v for the good; v is distributed according to cumulative distribution G with support $[0, \infty)$. The density of G is g and the hazard rate $\frac{g}{1-G}$ is assumed to be weakly increasing in v.

The good sought by consumers is sold by a (brick-and-mortar or internet) merchant at a *retail price* p. Together with the retail price, the merchant chooses his *cash/card policy*, which consists of the decision to accept/refuse cards, and, when allowed, the card surcharge/cash discount.

A *merchant* has convenience cost normalized at 0 for a card payment and a known $b_S > 0$ for a cash payment. The merchant pays a fee m to his acquirer bank for each card transaction. We normalize production costs to zero, in which case the (endogenous) retail price p coincides with the merchant's mark-up.

The *tourist test* is said to be satisfied if and only if the merchant fee does not exceed the merchant's convenience benefit of a card payment, also called the avoided cost:

$$b_S \geq m.$$

That is, conditional on the customer being at the point of sale and willing to pay regardless of the payment method, the merchant is happy to have the customer pay by card. When satisfied with an equality, the tourist test implies that the merchant is indifferent as to the payment method and so cardholders do not exert an externality on the merchant when choosing whether to pay by cash or card.

Under "uniform pricing", the merchant is required to set a single retail price applying regardless of the payment method. Under "discriminatory pricing", a merchant who accepts cards may practice a cash discount $\sigma \geq 0$ or a card surcharge $\tau \geq 0$, depending on what is allowed. The notation $(\sigma; \tau)$ reflects the fact that a discount (surcharge) is a subsidy granted (a tax levied) by the merchant on card payments.

Once in the shop, each consumer learns his cost b_B of paying by cash. The parameter b_B is distributed according to H with support $[\underline{b}_B, +\infty)$. The density of H is h, and the hazard rate $\frac{h}{1-H}$ and the reverse hazard rate $\frac{h}{H}$ are assumed to be weakly monotone (increasing and decreasing, respectively) in b_B. As is standard in the literature, we assume that the distributions G and H are independent. We allow the lower limit $\underline{b}_B$ to be strictly negative, in which case the support of H contains positive and negative realizations. The cost b_B is positive when paying by cash is a nuisance for the consumer (due, e.g., (i) to the inconvenience of going to an ATM prior to the purchase, or just after the purchase to prepare for the next expenses, (ii) to the convenience of paying by card for recording purposes). The cost b_B is negative when the consumer wants to get rid of cash (e.g., a foreign tourist in the last days of vacation), or prefers paying by cash for tax evasion purposes.[14] Another interesting interpretation of b_B is specific to credit cards (relative to direct debit, checks or cash). The ability to use credit allows the consumer to smooth consumption. The cost of "cash" (here including direct debit) is then the sacrifices incurred on current consumption.

Under our convention, the consumer's net surplus from buying the good by cash under a cash discount σ is[15]

$$v - (p - \sigma + b_B).$$

When using the card, the consumer pays a per-transaction fee f to his issuer bank.[16] Following our convention, the consumer's net surplus from buying the good by card under a card surcharge $\tau \geq 0$ is[17]

$$v - (p + f + \tau).$$

We assume that the consumer learns the retail price p set by the merchant *before* observing his cash/card policy; for instance, retail prices can be known from manufacturer or retailer advertising campaigns or price comparison websites. Knowing the retail price, the consumer chooses whether to incur the *shopping cost* s. The shopping cost covers finding and going to the shop, inspecting or trying the good, thinking about potential usage, size and model, loss aversion, etc. For instance, in the case of a brick-and-mortar merchant, the shopping cost may capture, for instance, transportation costs and/or the time spent in the store before getting to the cashier (where card acceptance and cash discounts/card surcharges are usually revealed). In the case of online merchants, the shopping cost captures the nuisance of completing the usually numerous steps before card surcharges are revealed. For brevity, we shall say that a consumer "visits the store" when he incurs the shopping cost s.

Upon learning the realization of b_B, the consumer decides whether to buy and, if he does, through which payment method. *The payment card system* incurs per-transaction total cost c; this cost decomposes between the cost on the merchant side (the "acquiring cost") c_A and that on the buyer side (the "issuing cost") c_I: $c = c_A + c_I$. The platform chooses fees (m, f).

---

[14] Of course, the social welfare function need not respect end-user preferences in this case.

[15] To rule out pathological cases, we will assume that $\underline{b}_B$ is sufficiently high so that $p - \sigma + \underline{b}_B > 0$ prevails in equilibrium. This is a very mild condition, as the consumer could throw away an amount of cash equal to the net purchase price $p - \sigma$, rather than buy the item and not use it.

[16] Note that f can in principle have either sign; indeed, in practice, it is often equal to zero or is negative (discount at the point of sale, cash-back bonuses, frequent flyer miles, free insurance against cancellation or damage, etc).

[17] To rule out pathological cases, we shall make the weak assumption that $p + f > 0$ in equilibrium (otherwise, absent card surcharges, consumers would benefit from buying the good even if their valuation for it is zero).

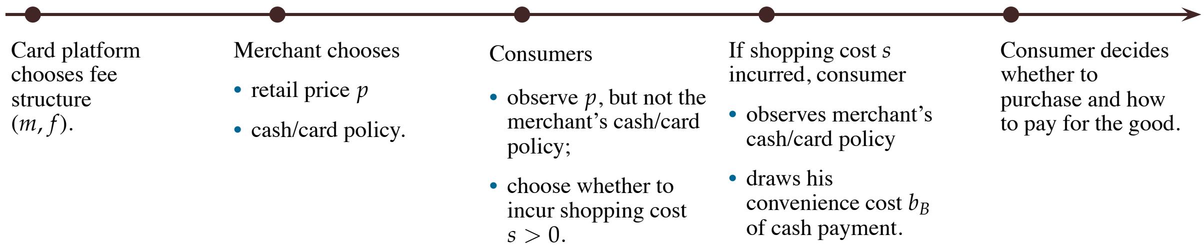

Figure 1: Timing

Open system: If the system is a payment card association (such as Visa or MasterCard) with interchange fee a (the payment from the merchant's bank to the cardholder's bank), and the issuing and acquiring sides are populated by competitive banks, then $m + f = c$, as $m = c_A + a$, and $f = c_I - a$; more generally, c stands for total cost, including markups levied by acquirers or issuers with market power. Like in the rest of the literature (to which we refer for a foundation of this assumption), we will assume that an open system's members select (m, f) so as to maximize the volume of transactions.

Closed system: By contrast, closed systems like Amex or PayPal face no such constraint. They only need to meet the break-even condition: $m + f \geq c$. We will assume that a platform operating as a closed system selects (m, f) so as to maximize profit. We will point out which results are specific to an open system.

The timing and information structure of our model is summarized as follows:

1. The card platform publicly chooses a fee structure (m, f).

2. The merchant chooses retail price and cash/card policy (i.e., whether or not to accept cards, and, if allowed, the card surcharge/cash discount).

3. Consumers decide whether to visit the store, having observed the retail price, but not the merchant's cash/card policy. When taking this decision, consumers correctly anticipate the merchant's cash/card policy.

4. If a consumer visits the store, he observes his convenience cost of cash payment, b_B, and learns the merchant's cash/card policy. The consumer then decides whether to purchase, and how to pay for the good.

Our solution concept is Perfect Bayesian equilibrium. This concept in particular assumes that consumers rationally anticipate the card surcharge or cash discount selected by the merchant. This assumption therefore presumes non-negligible rationality, but allows us to study the least favorable environment for the regulation of surcharges. Section 7.1 studies the case of naive consumers who fail to comprehend the merchant's incentives and do not anticipate surcharges.

2.2 Consumer behavior and missed sales

We will study equilibria of the game above under four alternative scenarios. In the first scenario, the merchant accepts cards, but private or social regulations prevent him from using cash discounts or card surcharges, in which case *uniform pricing* prevails: $\sigma = \tau = 0$. In the second scenario, the merchant is allowed to practice *cash discounts*, but not card surcharges: $\sigma \geq 0$ and $\tau = 0$. In the third scenario, *card surcharges* are allowed, but not cash discounts: $\tau \geq 0$ and $\sigma = 0$. There is no loss of generality involved in disregarding the case where *both* cash discounts and card surcharges are allowed, as the latter will be always preferred to the former by merchants (and so equilibrium is as if only card surcharges were permitted). In the fourth scenario, the merchant does not accept cards (*cash only*). Because consumers do not observe the merchant's cash/card policy before visiting the store, this is equivalent to the merchant accepting cards but setting $\tau = +\infty$.

Consumer behavior

To describe the consumers' behavior in a unified manner, consider a merchant who practices a cash discount $\sigma \geq 0$ and a card surcharge $\tau \geq 0$. Once in the store, and facing a cost b_B of cash payments, the consumer buys the good if and only if his willingness to pay exceeds the total price with the cheapest payment option:

$$v \geq p + \min\{f + \tau, b_B - \sigma\}.$$

Note that the shopping cost s is sunk at this stage and therefore does not influence the consumer purchasing or payment decision. In turn, the consumer's decision to go to the store is taken ex ante, i.e., before the realization of the cost b_B is known. To analyze this decision, it is convenient to define the function

$$T_B(v - p, f + \tau, \sigma) \equiv \mathbb{E}\left[\min\{v - p, f + \tau, b_B - \sigma\}\right], \tag{1}$$

which is the *expected transaction cost of a consumer with valuation v, facing a price p, a card fee f, a cash discount σ and a card surcharge τ* (expectations are taken with respect to b_B). To understand the expression above, note that (i) if the consumer pays by card, his transaction cost is $f + \tau$, (ii) if the consumer pays by cash, his transaction cost is $b_B - \sigma$, and (iii) if the consumer decides not to buy the good (in which case *missed sales* occur), his transaction cost is the gross surplus from the foregone purchase, $v - p$. We will clarify below under what conditions missed sales *do not* occur. In this case, the expected transaction cost of consumers does not depend on the gross surplus $v - p$, and we denote it by $T_B(\infty, f + \tau, \sigma)$.

In equilibrium, ex-ante optimality implies that only consumers with valuation greater than a threshold $V(p, f + \tau^a, \sigma^a)$ go to the store, where τ^a and σ^a are the anticipated surcharge and discount (in equilibrium equal to the values τ and σ chosen by the merchant). This threshold solves the following equation in v:

$$v = s + p + T_B\left(v - p, f + \tau^a, \sigma^a\right). \tag{2}$$

In words, the only consumers to make the trip to the store are those whose valuations exceed the sum of the shopping cost, the advertised price and the expected transaction costs at the equilibrium cash discount $\sigma^a = \sigma$ or card surcharge $\tau^a = \tau$.[18]

Missed sales

Missed sales occur when some consumers who have incurred the shopping cost s decide that not buying the good is the best course of action. The following lemma provides a necessary and sufficient condition on the shopping cost for missed sales not to occur in equilibrium.

Lemma 1 ***(missed sales)*** *Missed sales do not happen in equilibrium if and only if the equilibrium values of σ and τ satisfy:*

$$s + \mathbb{E}\left[\min\left\{b_B - f - \sigma - \tau, 0\right\}\right] \geq 0. \tag{3}$$

Proof. Notice first that, because the support of $H(b_B)$ contains arbitrarily large numbers, missed sales do not occur if and only if the lowest-valuation consumer to go to the store finds it worthwhile to pay by card:

$$V(p, f + \tau, \sigma) \geq p + f + \tau. \tag{4}$$

To establish the "only if" part of the lemma, assume that (4) holds. It then follows from (2) that

$$s + T_B\left(\infty, 0, \sigma + f + \tau\right) \geq 0,$$

which is equivalent to (3).

For the "if" part of the lemma, assume that condition (4) is violated. Because

$$v - T_B\left(v - p, f + \tau, \sigma\right)$$

is strictly increasing in v (as implied by straightforward differentiation), we obtain that

$$v - s - p - T_B\left(v - p, f + \tau, \sigma\right) > 0$$

[18] An alternative, and perhaps more familiar, derivation of the threshold $V(p, f + \tau^a, \sigma^a)$ is the following. Let us define $V(p, f + \tau^a, \sigma^a)$ as the unique solution of the following equation in v:

$$\int_{\underline{b}_B}^{v-p+\sigma^a} \left(v - p - \min\left\{f + \tau^a, b_B - \sigma^a\right\}\right) dH(b_B) + \left(1 - H\left(v - p + \sigma^a\right)\right) \cdot \max\left\{v - p - f - \tau^a, 0\right\} = s.$$

Accordingly, $V(p, f + \tau^a, \sigma^a)$ is the smallest valuation such that the expected benefit of visiting the store equals the shopping cost. Straightforward algebra shows that the equation above is equivalent to (2).

at $v = p + f + \tau$. This is equivalent to

$$s + \mathbb{E}\left[\min\left\{b_B - \sigma - f - \tau, 0\right\}\right] < 0,$$

establishing the claim. ■

Missed sales do not occur provided that shopping costs are high relative to the equilibrium card fee and merchant's cash discount/card surcharge. While stated in terms of endogenous quantities, this result will be useful to determine when missed sales occur on the equilibrium path.

An immediate consequence of the lemma above is that missed sales *always* occur if the merchant does not accept cards (in which case $\tau = +\infty$). As we explain in the next section, the fear of missed sales underlies the novel "must-take-card" argument developed in this paper.

Before proceeding, we shall impose a weak regularity condition.

Assumption 1 ***(convenience of card payments)*** *The following inequality holds:*

$$s + \mathbb{E}\left[\min\left\{b_B + b_S - \tilde{c}, 0\right\}\right] > 0,$$

where $\tilde{c} \equiv m + f$ is aggregate fee charged by the platform.

Intuitively, this assumption means that, relative to card payments, payments by cash are on average sufficiently inconvenient to consumers and merchants. This is a very mild assumption, as shopping costs are usually significant, and paying by cash is more often costly ($b_B > 0$) than beneficial ($b_B < 0$) to consumers. In the case of an open system (where $\tilde{c} = c$), this inequality is trivially satisfied whenever $\underline{b}_B + b_S \geq c$. In the case of a closed system, this assumption requires that the platform's profit margin per transaction is small relative to the inconvenience cost of cash payments jointly experienced by consumers and merchants.

In light of Lemma 1, Assumption 1 implies that missed sales do not occur provided that $f + \sigma + \tau \leq b_S - \tilde{c}$. As shown in the next subsection, this guarantees that any fee structure (m, f) and cash/card policy (σ, τ) conductive to efficient card usage generates no missed sales.

2.3 Aggregate transaction costs and the first best

In the analysis that follows, we will employ the welfare-maximizing outcome as a benchmark to identify the relevant distortions in the scenarios of uniform pricing, cash discount and card surcharge. It is convenient to disregard the possibility of missed sales, and verify ex post that missed sales indeed do not happen in the welfare-maximizing allocation.

Merchant and aggregate transaction costs

Let

$$T_S(m, f, \sigma, \tau) \equiv H(\sigma + \tau + f) \cdot [\sigma + b_S] + [1 - H(\sigma + \tau + f)] \cdot [m - \tau] \quad (5)$$

denote the *merchant's expected transaction costs under the fee structure* (m, f) *and the cash/card policy* (σ, τ). The expression above is the sum of transaction costs per sale (which are $\sigma + b_S$ in the case of cash, and $m - \tau$ in the case of card payments) weighted by the fractions of consumers who use either cash or card.

The *aggregate transaction costs* (borne by consumers and the merchant) are then equal to

$$T(m, f, \sigma, \tau) \equiv \int_{\underline{b}_B}^{\sigma+\tau+f} (b_B + b_S)\, dH(b_B) + [1 - H(\sigma + \tau + f)] \cdot [m + f], \quad (6)$$

which is the sum of (1) and (5). The maximization of total welfare, as pursued by social regulation, is closely related to the minimization of aggregate transaction costs (when missed sales are ruled out). The next lemma will play a key role in the analysis that follows.

Lemma 2 ***(minimizing aggregate transaction costs)*** *For any fee structure* (m, f) *and cash/card policy* (σ, τ)*, let us define the* <u>*effective merchant fee*</u> *and the* <u>*effective cardholder fee*</u> *as*

$$\hat{m} \equiv m - \tau - \sigma \quad \text{and} \quad \hat{f} \equiv f + \tau + \sigma.$$

Holding constant the platform's aggregate fee $\tilde{c} = m + f$*, the quadruple* (m, f, σ, τ) *minimizes aggregate transaction costs if and only if*

$$\hat{m} = b_S \quad \text{and} \quad \hat{f} = \tilde{c} - b_S. \quad (7)$$

Moreover, along the locus $\hat{m} + \hat{f} = \tilde{c}$*, the aggregate transaction cost is a U-shaped function of* $\hat{m}$ *with strict minimum at* $\hat{m} = b_S$*.*

Intuitively, the minimization of aggregate transaction costs requires that the *effective* merchant fee be equal to the merchant's convenience cost of cash payments. That is, the tourist test has to be satisfied with equality, after one replaces the merchant fee by its effective counterpart, which accounts for the merchant's cash discounts or card surcharges. The effective cardholder fee $\hat{f}$ is then determined by the accounting identity $\hat{m} + \hat{f} = m + f = \tilde{c}$.

First best

In line with the objectives of most regulatory authorities, our social welfare measure accounts for the surplus jointly obtained by consumers *and* merchants, but does not account for those obtained by the card platform, issuer and acquirer banks.[19]

In order to define social welfare, let us denote by

$$p^*(m, f, \sigma, \tau) \equiv \arg\max_p \left(1 - G\left(V(p, f + \tau, \sigma)\right)\right) \cdot \left(p - T_S(m, f, \sigma, \tau)\right) \tag{8}$$

the merchant's profit-maximizing retail price under a fee structure (m, f) and the cash/card policy (σ, τ). Our measure of social welfare is then the sum of the valuations of all consumers who visit the store, net of the shopping cost and the aggregate transaction costs incurred in each purchase:[20]

$$W(m, f, \sigma, \tau) \equiv \int_{V(p^*(m,f,\sigma,\tau), f+\tau, \sigma)}^{+\infty} \left(v - s - T\,(m, f, \sigma, \tau)\right) dG(v). \tag{9}$$

The next lemma characterizes the fee structure and cash/card policies that maximize social welfare.

Lemma 3 ***(welfare)***. *For any fee structure (m, f) and cash/card policy (σ, τ) satisfying the no-missed-sales condition (3),*

1. *the threshold valuation $V(p^*(m, f, \sigma, \tau), f + \tau, \sigma)$ is a univariate and strictly increasing function of the aggregate transaction costs,*

2. *social welfare $W(m, f, \sigma, \tau)$ is a univariate and strictly decreasing function of aggregate transaction costs.*

Therefore, holding constant the platform's aggregate fee $\tilde{c}$, any fee structure (m, f) and cash/card policy (σ, τ) satisfying condition (7) is welfare-maximizing. Moreover, social welfare strictly decreases with the platform's aggregate fee $\tilde{c}$.

Note that Assumption 1 guarantees that missed sales do not occur at any fee structure (m, f) and cash/card policy (σ, τ) conducive to efficient card usage (where condition (7) holds), or conductive to inefficiently high card usage (where $\hat{f} < b_S - \tilde{c}$).

[19]This qualification is obviously mute if the platform operates as an open-system in a market with perfectly competitive issuer and acquirer sectors. See Tirole (2011) for a discussion of whether issuer and acquirer profits should be included into the measure of social welfare.

[20]As we shall see in Section 4, the equilibrium cash/card policy (σ, τ) is a function of the fee structure (m, f), and so one could write social welfare as a function of (m, f) only. It is however convenient to explicitly indicate the dependence of social welfare on (σ, τ), since the merchant's cash/card policy might be subject to direct regulation (as discussed in Section 6). Because retail prices are usually out of the scope of regulatory authorities, we directly plug the *equilibrium* retail prices in defining social welfare.

3 Uniform pricing: Missed sales and must-take cards

This section studies equilibria when private or social regulations prevent the merchant from practicing cash discounts or card surcharges. It thereby characterizes the uniform pricing benchmark with which discriminatory pricing should be compared.

3.1 Markups and merchant resistance

When facing a high merchant fee and a prohibition to surcharge, the merchant may refuse the card, leading the customer to forgo buying if he finds it too costly or does not have time to go and fetch cash. As our analysis reveals, concerns about missed sales are an important new rationale for the "must-take-card" argument. In particular, the possibility of missed sales has been ignored in the card payment literature; it is therefore of independent interest to study its implications.

Note that, if the tourist test is satisfied, i.e., $m \leq b_S$, it is a dominant strategy for merchants to accept payments by card. The more interesting case happens when the tourist test is violated, i.e. $m > b_S$, introducing a genuine trade-off between economizing on transaction costs and losing sales. In a first step, we take the price p as fixed and investigate equilibria with *card acceptance* and *card refusal*. We will then look at the choice of price.

Card acceptance equilibrium. For a given retail price, an equilibrium is characterized by the merchant's strategy to accept or not card payments, and the consumers' beliefs about the merchant's policy (which determines their decision to go to the store). We shall proceed by assuming that missed sales do not happen if the merchant accepts cards. After completing our characterization, we will then confirm that this is indeed true in equilibrium.

In an equilibrium in which the merchant accepts cards, all consumers with valuation $v \geq V(p, f, 0)$ go the store, where the threshold $V(p, f, 0)$ is given by (2). Moreover, merchants have to find it profitable to accept card transactions given the consumers' expectations of card acceptance. Equivalently, the profit differential between card acceptance and refusal should be positive:

$$\Delta\pi(p, m, f) \equiv \int_{V(p,f,0)}^{\infty} \{p - T_S(m, f, 0, 0) - H(v - p) \cdot (p - b_S)\} \, dG(v) \geq 0,$$

where $T_S(m, f, 0, 0)$ is the merchant's expected transaction cost per sale under uniform pricing (as defined in (5)). Note that the last term of the integrand above captures the effect of missed sales on profits when cards are not accepted.

Card refusal equilibrium. For a given retail price, a card refusal equilibrium is characterized by the threshold $V(p, \infty, 0)$, defined in (2), such that only consumers with valuation above this threshold visit the store. Moreover, merchants must find it profitable to refuse card transactions

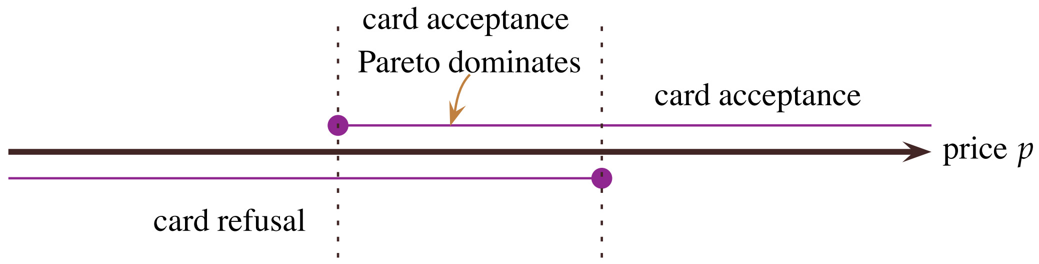

Figure 2: Equilibrium multiplicity under uniform pricing

given the consumers' expectations of card refusal.

The next proposition shows that a card acceptance equilibrium exists if and only if the markup p is sufficiently high, and that a card refusal equilibrium equilibrium exists if and only if the markup p is sufficiently low.

Proposition 1 ***(markups and merchant resistance under uniform pricing)*** *Assume that the tourist test is violated (i.e., $m > b_S$), and consider the case of uniform pricing.*

1. *There exists a threshold $P^u(m, f)$ such that a card acceptance equilibrium exists if and only if $p \geq P^u(m, f)$. This threshold is continuous with $P^u(b_S, f) = b_S$ for all f. It is also strictly increasing in m.*

2. *There exists a threshold $Q^u(m, f)$ such that a card refusal equilibrium exists if and only if $p \leq Q^u(m, f)$. This threshold is continuous with $Q^u(b_S, f) = b_S$ for all f. It is also strictly increasing in m.*

3. *For any (m, f), the thresholds identified above satisfy $Q^u(m, f) \geq P^u(m, f)$, and there exist multiple equilibria when $p \in [P^u(m, f), Q^u(m, f)]$. In this case, the card acceptance equilibrium Pareto-dominates the card refusal equilibrium.*

Proposition 1 unveils an important driver of merchant resistance, namely the merchant's markup. A high markup makes it costly to the merchant to lose sales by turning down the card. Thus the card acceptance equilibrium exists only for a sufficiently high markup.

The multiplicity of equilibria, illustrated in Figure 2, stems from complementarities between the merchant's card acceptance decision and the consumers' expectations about card acceptance (which determine who visit the store). For intermediate retail price levels, if consumers expect the merchant to accept cards, the volume of missed sales that would result from card refusal is high enough to induce the merchant to indeed accept card transactions. In turn, if consumers expect merchants to refuse cards, the reduction in missed sales that would result from card acceptance is not sufficient to compensate for the expenses associated with card fees.

Obviously, consumers are better off at the card acceptance equilibrium, as they enjoy an extra payment option. The same is true for the merchant, as in this equilibrium, more consumers go to the store and the increase in transaction costs is more than compensated by the extra sales made possible by card acceptance.

The next corollary shows how the thresholds $P^u(m,f)$ and $Q^u(m,f)$ change as the inconvenience costs of cash payments increase (in the appropriate probabilistic sense). As intuition suggests, the range of prices under which a card acceptance (refusal) equilibrium exists expands (contracts) as cash payments become more inconvenient.

Corollary 1 ***(merchant resistance and cash inconvenience).*** *Consider two distributions of the consumers' cost of cash payments H and $\hat{H}$, and denote their respective thresholds of card acceptance and refusal equilibria by $(P^u(m,f), Q^u(m,f))$ and $(\hat{P}^u(m,f), \hat{Q}^u(m,f))$. Then, if H is dominated by $\hat{H}$ according to first-order-stochastic dominance in the range $(f,+\infty)$,*[21]

$$\hat{P}^u(m,f) \leq P^u(m,f) \quad \textit{and} \quad \hat{Q}^u(m,f) \leq Q^u(m,f).$$

Remark 1 (credit vs. debit cards). Corollary 1 also reveals that merchant resistance to card acceptance goes down as the inconvenience cost of cash payments faced by consumers increases. This result allows us to make an interesting comparison between merchant resistance to credit vis-a-vis debit cards. Because credit cards allow consumers to defer payments (and smooth consumption), cash payments are more inconvenient than card payments when the card is a credit rather than a debit card. Therefore, *for a fixed fee structure*, the corollary above implies that merchant resistance is lower to credit than to debit cards. Of course, the platform takes that into account when it chooses its fee structure, as discussed in Subsection 3.3.

3.2 Merchant resistance with optimal pricing

We will now study card acceptance by merchants when retail prices are endogenous. More formally, this corresponds to studying perfect equilibria of the game defined from stage 2 onwards (i.e., taking the fee structure (m,f) as given), where merchants simultaneously choose retail prices and the cash/card policy.

An equilibrium with endogenous pricing is then characterized by the merchant's decision to accept/reject card payments, the optimal retail price $p^*(m,f,0,0)$ defined in (8), and the belief of consumers about card acceptance *for each possible retail price p.*

We construct out-of-equilibrium beliefs by selecting the Pareto-dominant equilibrium in the subgame that follows each price announcement. Namely, following any announcement

[21]That is, $\hat{H}(b_B) = H(b_B)$ for all $b_B \in [\underline{b}_B, f]$, and $\hat{H}(b_B) \leq H(b_B)$ for all $b_B \in (f,+\infty)$. So, relative to H, the distribution $\hat{H}$ corresponds to a higher cost of using cash and generates more missed sales in the case of card refusal.

$p \geq P^u(m,f)$, we assume that the card acceptance equilibrium is played in the continuation game that follows the announcement. When the merchant sets price $p < P^u(m,f)$, card refusal is the unique equilibrium and so no selection is required.

The next proposition shows that an equilibrium with card acceptance exists provided that the merchant fee is smaller than a threshold $\mathbb{M}^u(f)$. Most importantly, the threshold $\mathbb{M}^u(f)$ violates the tourist test, i.e. $\mathbb{M}^u(f) > b_S$. Accordingly, missed sales induce merchants to accept inefficiently high merchant fees, establishing a new foundation for the "must-take-card" argument.

Proposition 2 ***(equilibrium merchant resistance under uniform pricing)*** *When the retail price p is endogenized, the merchant accepts cards if and only if the merchant fee is no larger than a threshold $\mathbb{M}^u(f)$, where $\mathbb{M}^u(f)$ violates the tourist test: $\mathbb{M}^u(f) > b_S$ for all f.*

The intuition for Proposition 2 is simple. Suppose for a moment that $m = b_S$, in which case the tourist test is satisfied with equality. In this case, the merchant faces the same transaction cost per sale as if only cash transactions were accepted. The merchant however does not incur any missed sales. Therefore, the merchant has a strict gain from card acceptance at $m = b_S$. As $m > b_S$, the gains from avoiding missed sales are mitigated by higher transactions costs per sale. The rise in transaction costs impacts retail prices and leads to a lower volume of consumers visiting the store (as implied by Lemma 3). As established by Proposition 2, the gains from avoiding missed sales are larger than the losses due to higher transaction costs provided that the merchant fee is lower than some threshold $\mathbb{M}^u(f) > b_S$.

3.3 Equilibrium merchant fee

We now study the equilibrium fee structure set by the platform in the case of uniform pricing. We again assume that following the platform's choice of (m,f), the Pareto-superior Nash equilibrium identified in Proposition 2 is played from stage 2 onwards.

In the case of an open system, we assume that the card association maximizes the volume of card transactions. Accordingly, the fee structure is chosen to solve

$$\max_{\{m,f\}} \; \left(1 - G\left(V(p^*(m,f,0,0),f,0)\right)\right) \cdot \left(1 - H(f)\right), \tag{10}$$

subject to $c - f = m \leq \mathbb{M}^u(f)$. That the merchant fee m has to be lower than the threshold $\mathbb{M}^u(f)$ reflects the fact that the platform is constrained to choose fee structures for which the card is accepted by the merchant. We refer to this requirement as the *card acceptance constraint*.

In the case of a closed system, we assume that the card platform maximizes profit. Accordingly, the fee structure is chosen to solve

$$\max_{\{m,f\}} \ (1 - G\left(V(p^*(m, f, 0, 0), f, 0)\right)) \cdot (1 - H(f)) \cdot (m + f), \tag{11}$$

subject to $c - f \leq m \leq \mathbb{M}^u(f)$. To simplify the technical discussion, we will assume from now on that programs (10) and (11) are quasi-concave.[22]

As the programs above reveal, the fee structure (m, f) affects the platform's objective by (i) impacting the retail prices practiced by the merchant, (ii) changing the fraction of consumers who pay by card and, in the case of a closed system, (iii) determining the platform's profit *per* card transaction. The next proposition establishes that the card platform has an incentive to set an inefficiently high merchant fee, so as to steer consumers towards using the card. As established in Proposition 2, the fear of missed sales makes the merchant accept merchant fees that exceed the benefit of card usage.

Proposition 3 ***(equilibrium fee structure under uniform pricing)*** *Under uniform pricing, for either an open or closed system, the card platform sets a fee structure (m^u, f^u) such that*

$$\underbrace{\frac{h(f^u)}{1 - H(f^u)}}_{\substack{\textit{card usage} \\ \textit{semi-elasticity}}} \quad - \quad \underbrace{\left(\frac{\partial p^*}{\partial m} - (1 - H(f^u))\right)}_{\textit{demand sensitivity}} \cdot \underbrace{\frac{g\left(V^u\right)}{1 - G\left(V^u\right)}}_{\substack{\textit{demand} \\ \textit{semi-elasticity}}} \quad \geq \quad 0, \tag{12}$$

where $V^u \equiv V(p^*(m^u, f^u, 0, 0), f^u, 0)$. This condition holds with equality if and only if the card acceptance constraint is slack. Moreover, the merchant fee m^u strictly violates the tourist test: $m^u > b_S$.

Proof (sketch). Consider first the case of an open system, in which case $m + f = c$. As a first step, note from Lemmas 2 and 3 that the merchant fee that maximizes the volume of consumers visiting the store (as opposed to the volume of card transactions) satisfies the tourist test with equality: $m = b_S$. Consider now the platform's incentive to increase the merchant fee (and decrease the cardholder fee) at the fee structure $(m, f) = (b_S, c - b_S)$. Optimality implies that a marginal increase in m generates a *second-order loss* at the volume of consumers visiting the store. It however entails a *first-order gain* in the share of card transactions (as $1 - H(f)$ is strictly decreasing in f), as opposed to cash. As a consequence, a volume-maximizing platform chooses a merchant fee that violates the tourist test. In particular, because $f^u < c - b_S$, Assumption 1 guarantees that missed sales do not happen on the equilibrium path.

The first-order condition (12) follows from straightforward differentiation of the platform's objective in (10). The proof in the appendix shows that the case of a closed system can be

[22] This assumption is made only to avoid a tedious discussion of second-order conditions. This requirement holds for most distributions of interest, for example when G and H are either exponential or lognormal.

treated similarly to that of an open system: It suffices to replace the transaction cost c with a generalized cost $\tilde{c}$ that incorporates the platform's profit margin over each card transaction. ■

The first-order condition (12) illustrates the trade-off faced by the platform in choosing merchant fees. On the one hand, high merchant fees serve to "cross-subsidize" card usage by consumers (by means of a lower card fee f). On the other hand, high merchant fees increase aggregate transaction costs, resulting in higher retail prices and a lower volume of consumers visiting the store (i.e., demand for the good). The fee structure set by the platform balances the impacts of a marginal increase in m on the share of card transactions *per sale*, and on the volume of consumers visiting the store. The effect on the share of card transactions is captured by the semi-elasticity of card usage evaluated at the optimal card fee, as it appears on the first term of (12). In turn, the effect on volume is captured by the second term of (12): It equals the semi-elasticity of demand (evaluated at the threshold V^u, which determines which consumers visit the store) adjusted by the sensitivity of demand to the merchant fee. In particular, the demand sensitivity term is always positive: a higher merchant fee increases the threshold valuation $V(p^*(m, f, 0, 0), f, 0)$, therefore decreasing the volume of consumers visiting the store.[23]

It should be noted that the first-order condition (12) is satisfied *with equality* at the optimum *only if* the card acceptance constraint is slack: $m^u < \mathbb{M}^u(f)$. If this constraint is binding, the platform will be obliged to set $m^u = \mathbb{M}^u(f)$, as it cannot further raise the merchant fee without losing the merchant's membership. The next remark argues that this corner solution will most likely prevail at the optimum.

Remark 2 (card acceptance constraint). How likely is it that the card acceptance constraint binds, in which case $m^u = \mathbb{M}^u(f^u)$? Letting η_C denote the elasticity of card usage to cardholder fee and η_D the price elasticity of demand, the card acceptance constraint binds at the optimum when

$$\frac{|f^u|}{p} < \left|\frac{\eta_C}{\eta_D}\right| \cdot \left(\frac{\partial p^*}{\partial m} - (1 - H(f^u))\right)^{-1}.$$

In practice f^u is negative and typically smaller than 1% of the retail price. As result, for reasonable values of elasticities η_C and η_D and demand sensitivity, the merchant fees observed in reality should equal the acceptance threshold $\mathbb{M}^u(f^u)$.

Remark 3 (credit vs. debit cards). It is straightforward to infer from Proposition 3 that the equilibrium merchant fee m^u increases as the distribution of consumers' inconvenience costs of cash benefits increases in the sense of hazard-rate dominance (which implies first-order stochastic dominance).[24] Therefore, in the spirit of Remark 1, the merchant fees for debit cards

[23]The demand sensitivity term equals the total derivative of the threshold valuation with respect to m along the locus $m + f = m^u + f^u$, evaluated at (m^u, f^u). This derivative is positive by virtue of Lemmas 2 and 3.

[24]The distribution $\hat{H}$ dominates H according to hazard-rate dominance if $\frac{\hat{h}(b_B)}{1-\hat{H}(b_B)} \leq \frac{h(b_B)}{1-H(b_B)}$ for all b_B.

should be lower than those for credit cards. This prediction is in line with the fee structure of most card payment platforms.

Importantly, Proposition 3 reveals that, absent regulation, the economy exhibits an *inefficiently high* use of cards. As established in Lemma 2, aggregate transaction costs are strictly higher under the equilibrium fee structure (m^u, f^u) than under the efficient fee structure $(b_S, c-b_S)$. Lemma 3 then implies that the social welfare when platform fees are left unregulated is strictly below first best.

4 Cash discounts and card surcharges

When consumers are perfectly informed about the merchant's card policy, the difference between cash discounts and card surcharges is semantic. After all, the merchant can always redefine retail prices and interchangeably label the difference between cash and card payments as either a discount or a surcharge. This observation comes at odds with the fact that public authorities and business executives do not see discounts and surcharges as equivalent, as the recent and intense antitrust litigation on this matter attests.

In this section we derive the merchant's optimal cash discount and card surcharge. As our analysis reveals, imperfect consumer information about the merchant cash/card policy breaks the aforementioned equivalence between discounts and surcharges. More importantly, in equilibrium, merchants shall always prefer surcharges to discounts.

We start the analysis with the merchant's card acceptance decision, and then study optimal cash discounts and card surcharges.

4.1 Card acceptance

As one should expect, merchant resistance towards card acceptance is lower when discriminatory pricing is allowed. In the card surcharge scenario, the merchant accepts cards *regardless* of the merchant fee, as he has the ability to pass through to consumers the merchant fee. In the case of cash discounts, the merchant is willing to accept cards provided that the merchant fee is smaller than a threshold $\mathbb{M}^d(f)$. Because the merchant can use discounts to steer consumers to use cash, merchant resistance to card transactions is smaller than in the case of uniform pricing, in the sense that $\mathbb{M}^d(f) \geq \mathbb{M}^u(f)$. These results are collected in the next lemma.

Lemma 4 ***(equilibrium merchant resistance under discriminatory pricing)***

1. *When cash discounts are allowed, the merchant accept cards if and only if the merchant fee is no larger than a threshold $\mathbb{M}^d(f)$, where $\mathbb{M}^d(f) \geq \mathbb{M}^u(f) > b_S$.*

2. *When card surcharges are allowed, the merchant accept cards regardless of the merchant fee $(\mathbb{M}^s(f) = +\infty)$.*

The proof in the appendix proceeds as in the scenario of uniform pricing. It first constructs card acceptance and refusal equilibria taking retail prices as exogenous. It then incorporates optimal retail prices, and relates the merchant's card acceptance decision to the merchant fee set by the platform. As before, we construct out-of-equilibrium beliefs by selecting the Pareto-dominant equilibrium in the continuation game that follows each price announcement.[25]

4.2 Cash discounts

In this subsection, which can be skipped in a first reading, we analyze the merchant's optimal choice of cash discounts. Given the fee structure announced by the platform at stage 1, the merchant's problem at stage 2 is that of choosing a retail price p and a cash discount σ to maximize

$$(1 - G\left(V(p, f, \sigma^a)\right)) \cdot (p - T_S(m, f, \sigma, 0)),$$

where σ^a is the cash discount *anticipated* by consumers. That the threshold $V(p, f, \sigma^a)$ is evaluated at σ^a reflects the fact that consumers do not observe the merchant's card policy before visiting the store. Of course, in equilibrium, the cash discount anticipated by consumers will coincide with the actual cash discount practiced by the merchant.

The equilibrium cash discount then has to minimize the expected transaction costs faced by the merchant:

$$\sigma^*(m, f) \;\in\; \arg\min_{\sigma} \left\{T_S(m, f, \sigma, 0)\right\}.$$

The next proposition characterizes the solution to this problem.

Proposition 4 ***(equilibrium cash discount)*** *In any equilibrium where cash discounting is allowed:*

1. *The equilibrium cash discount is $\sigma^*(m, f) = 0$ if and only if*

$$m \leq M^d(f) \equiv b_S + \frac{H(f)}{h(f)},$$

where the threshold $M^d(f)$ is weakly increasing in f.

2. *If $m > M^d(f)$, the equilibrium cash discount $\sigma^*(m, f) > 0$ is the unique solution of the following equation in σ:*

$$\sigma + \frac{H(f+\sigma)}{h(f+\sigma)} = m - b_S. \tag{13}$$

The equilibrium cash discount is strictly increasing in m, and weakly decreasing in f.

[25]Following an out-of-equilibrium price announcement, beliefs about cash discounts and card surcharges do not change. As will be clear shortly, this results from the fact that the merchant-optimal cash discounts and card surcharges do not depend on the retail price.

3. *In the case of an open payment system, let $\tilde{m}^d$ solve $\tilde{m}^d = M^d(c - \tilde{m}^d)$. Then*

$$\sigma^*(m, f) = \max\left\{m - \tilde{m}^d, 0\right\}.$$

Accordingly, the equilibrium cash discount is positive, and varies one-for-one with the merchant fee, whenever $m \geq \tilde{m}^d$.

According to Proposition 4, cash discounts are positive if and only if the merchant fee is sufficiently higher than b_S, that is, when the fee m entails a "large" violation of the tourist test. In this case, cash discounts are characterized by the first-order condition (13). This condition equalizes the gains on transaction costs from steering the marginal consumer to use cash with the losses associated with "giving away" discounts to infra-marginal consumers (i.e., those who strictly prefer to pay by cash at the discount $\sigma^*(m, f)$). Conversely, for "small" violations of the tourist test, the merchant finds it too costly to actively subsidize consumers to use cash. Thus, a *sufficient* condition for the absence of cash discounts is that the tourist test be satisfied.

Interestingly, when the payment system is open, there is a complete pass-through of merchant fees into cash discounts. That is, the cash discount varies one-for-one with the merchant fee (and therefore with the interchange fee). Moreover, for both open and closed systems, discounts increase if the distribution H increases in the sense of reverse-hazard-rate dominance.[26] Accordingly, one should expect higher cash discounts when the card option is credit rather than debit.

For any fee structure (m, f) for which $m \neq b_S$, allowing for cash discounts does not lead to the welfare-maximizing outcome (as discussed in Lemmas 2 and 3). Indeed, as implied by (13), the effective merchant fee $m - \sigma^*(m, f)$ is, at the optimum, too high relative to the convenience benefit of card transactions, b_S. Accordingly, for a fixed fee structure (m, f), cash discounts reduce but do not eliminate the inefficient *overuse* of cards that would prevail under uniform pricing.

Finally, by virtue of Assumption 1, the cash discount $\sigma^*(m, f)$ generates no missed sales in equilibrium. Once in the store, those consumers who find themselves with high inconvenience costs of cash payments will pay by card rather than refrain from buying the good.

4.3 Card surcharge

We now consider the scenario where the merchant is allowed to use card surcharges, but (as will turn out without loss of generality) is not allowed to use cash discounts. We shall assume that missed sales do not occur on the equilibrium path. This simplification deserves some justification, as missed sales are likely to occur when card surcharges are high (recall that when $\tau = \infty$, which is the cash-only scenario, missed sales *always* occur). First, the next proposition

[26] The distribution $\hat{H}$ dominates H according to reverse-hazard-rate dominance if $\frac{\hat{h}(b_B)}{\hat{H}(b_B)} \geq \frac{h(b_B)}{H(b_B)}$ for all b_B.

will provide, for any given fee structure (m, f), necessary and sufficient conditions in terms of the shopping cost for missed sales not to occur when merchants optimally surcharge card transactions. Second, as the analysis of Section 5 reveals, card surcharges, under mild qualifications, do not occur at the *equilibrium* fee structure, in which case it is natural to rule out missed sales (by assuming sufficiently high shopping costs).

Given the fee structure announced by the platform at stage 1, the merchant at stage 2 chooses a retail price p and a card surcharge τ so as to maximize

$$(1 - G(V(p, f + \tau^a, 0))) \cdot (p - T_S(m, f, 0, \tau)),$$

where, similarly to the case of cash discounts, the threshold $V(p, f + \tau^a, 0)$ is evaluated at the anticipated card surcharge τ^a (rather than at the actual card surcharge practiced by the merchant). This again reflects the fact that consumers do not observe the merchant's cash/card policy before visiting the store. Because consumers have rational expectations, in equilibrium, the card surcharge anticipated by consumers will coincide with the actual card surcharge practiced by the merchant.

The equilibrium card surcharge, $\tau^*(m, f)$, then minimizes the expected transaction costs faced by the merchant:

$$\tau^*(m, f) \in \arg\min_{\tau} \{T_S(m, f, 0, \tau)\}.$$

The next proposition characterizes the solution to this problem.

Proposition 5 ***(equilibrium card surcharge)*** *In any equilibrium where card surcharging is allowed, and assuming no missed sales:*

1. *The equilibrium card surcharge is $\tau^*(m, f) = 0$ if and only if*

$$m \leq M^s(f) \equiv b_S - \frac{1 - H(f)}{h(f)},$$

where the threshold $M^s(f)$ is weakly increasing in f.

2. *If $m > M^s(f)$, the equilibrium card surcharge $\tau^*(m, f) > 0$ is the unique solution of the following equation in τ:*

$$\tau - \frac{1 - H(\tau + f)}{h(\tau + f)} = m - b_S. \tag{14}$$

The equilibrium card surcharge is strictly increasing in m, and weakly decreasing in f.

3. *In the case of an open payment system, let $\hat{m}^s$ solve $\hat{m}^s = M^s(c - \hat{m}^s)$. Then*

$$\tau^*(m, f) = \max\{m - \hat{m}^s, 0\}.$$

Accordingly, the equilibrium card surcharge is positive, and varies one-for-one with the merchant fee, whenever $m \geq \hat{m}^s$.

Missed sales indeed do not occur in equilibrium if and only if $s \geq S^s(m, f)$, where the threshold $S^s(m, f)$ satisfies condition (3) evaluated at $\sigma = 0$ and $\tau = \tau^*(m, f)$ with equality.

According to Proposition 5, card surcharges are positive if and only if the merchant fee is not too low relative to the convenience benefit b_S. In particular, card surcharges arise *even when* the tourist test is satisfied. When positive, card surcharges are characterized by first-order condition (14). This condition balances two effects. First, marginally increasing the surcharge raises the platform's revenue from infra-marginal consumers (i.e., those who strictly prefer to pay by card at the surcharge $\tau^*(m, f)$). Second, marginally increasing the surcharge steers the marginal consumer towards cash, generating a loss to the platform (which consists of the foregone surcharging revenue plus the net benefit of a card transaction, $b_S - m$). Only when the merchant fee is sufficiently low, the platform does not surcharge. Thus, a *necessary* condition for the absence of card surcharges is that the tourist test be satisfied.

Interestingly, when the payment system is open, there is a complete pass-through of merchant fees into card surcharges. That is, the card surcharge varies one-for-one with the merchant fee (and therefore with the interchange fee). Moreover, for both open and closed systems, surcharges increase if the distribution H increases in the sense of hazard-rate dominance. Accordingly, one should expect higher surcharges for credit than for debit cards.

For any fee structure (m, f), allowing for card surcharges does not lead to the welfare-maximizing outcome (as discussed in Lemmas 2 and 3). As revealed by (14), at the optimum, *the effective merchant fee*, $m - \tau^*(m, f)$, is too low relative to the convenience benefit of card transactions, b_S. Accordingly, for any fee structure (m, f), card surcharges necessarily lead to an *underuse* of cards (relative to first best). The reason is that merchants use card surcharges as a price discrimination device (screening consumers with different costs of cash payments). As such, the merchant always "overshoots" in the surcharge, and surcharges consumers more than what efficiency would dictate.

Finally, the card surcharge $\tau^*(m, f)$ generates no missed sales if and only if the shopping cost is high enough. In this case, those consumers who find themselves with high inconvenience costs of cash payments will pay by card even when subject to a large surcharge (rather than refrain from buying the good).

4.4 Comparison: cash discount or card surcharge?

Merchant associations often lobby for the right to surcharge card transactions. By contrast, merchants are usually allowed to discount for cash payments, but rarely do so. The next corollary combines the two previous propositions to compare cash discounts and card surcharges. In accordance with recent antitrust litigation, it shows that merchants unambiguously prefer

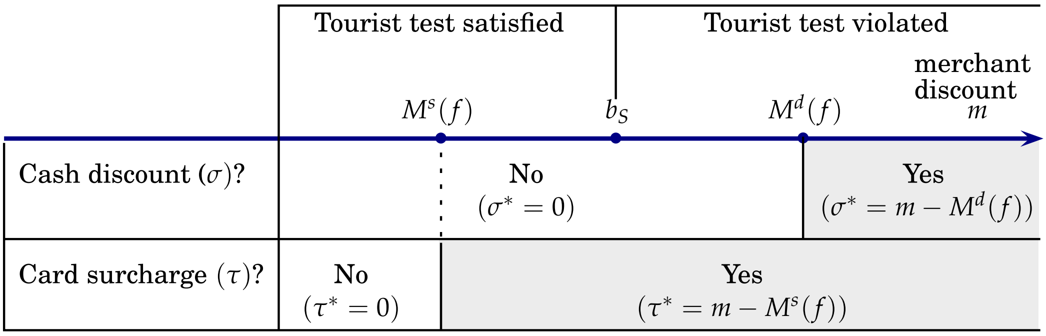

Figure 3: Discount and surcharge behavior in an open system

surcharges to discounts. Thus, if both surcharges and discounts are allowed, merchants will always employ the former and never the latter.

Corollary 2 ***(comparison between cash discounts and card surcharges)*** *Fix the fee structure (m, f). In any equilibrium with no missed sales:*

1. *On a stand-alone basis, surcharges are always larger than discounts: $\tau^*(m, f) \geq \sigma^*(m, f)$,*

2. *The merchant prefers surcharges to discounts (and so set discount $\sigma = 0$ if both are allowed).*

Moreover, missed sales are more likely to occur under card surcharges than under cash discounts.[27]

The intuition for this comparison is straightforward. Because consumers imperfectly observe the merchant's cash/card policy, surcharges are a rent extraction device (due to hold-up), while discounts are a giveaway. As a consequence, steering consumers to use cash rather than card is always more costly for the merchant under discounts than under surcharges. Therefore, for a given fee structure (m, f), surcharges are always larger than discounts, and generate higher profits for the merchant. An immediate consequence is that missed sales are more likely to occur under surcharges than under discounts, in the sense that the range of shopping costs such that missed sales do not occur is larger under discounts than under surcharges. Figure 3 summarizes the relationship between discounts and surcharges on the one hand, and compliance with the tourist test on the other hand (the specific passthrough formulae for $\sigma^*(m, f)$ and $\tau^*(m, f)$ apply only for an open system).

We are now in a position to study whether, *for a given fee structure* (m, f), society is best served by allowing or prohibiting surcharging (in Section 6, we will consider finer modes of regulation). To get intuition, consider the case of an open system. As implied by Proposition 5, the effective merchant fee under surcharging is equal to $\min\{m, \hat{m}^s\}$. The quasi-convexity of transaction costs around the minimizer b_S, and Lemma 3, then yield the following result.

[27]In the sense that, for every shopping cost s, if missed sales do not occur under card surcharges, then they do not occur under cash discounts.

Corollary 3 (unconstrained card surcharging vs. uniform pricing) *Consider a platform that operates as an open system. For a given fee structure (m, f), there exists a threshold $\bar{m}^s > b_S$ such that:*

1. *If $m \leq \hat{m}^s$, prohibiting surcharges has no impact on welfare.*

2. *If $\hat{m}^s < m < \bar{m}^s$, prohibiting surcharges strictly increases welfare.*

3. *If $\bar{m}^s < m$, allowing surcharges strictly increases welfare.*

By contrast, allowing for cash discounts always weakly increases welfare.

The result above assumes an *exogenous* fee structure (m, f). As such, its welfare statements are most useful for short-run analyses or in environments where the fee structure is rigid (due to long-term contracts or direct regulation). Arguably, the platform should adjust its fee structure in response to a regulatory intervention allowing merchants to price discriminate. This is the topic of the next section.

5 Platform's choice of merchant fee

How does the platform adjust its fee structure in response to laws allowing merchants to discount cash transactions or to surcharge card transactions? To answer this question, we will now derive a perfect Bayesian equilibrium of the game described in Figure 1. As before, we adopt the equilibrium refinement of Proposition 2, which selects the Pareto-superior equilibrium of the game defined from stage 2 onwards.

When discriminatory pricing is allowed, the platform in its choice of fee structure must account not only for its effect on retail prices, but also for its impact on the cash discount/card surcharge practiced by the merchant (that, together with the cardholder fee f, determines the share of card transactions). To study the two forms of discriminatory pricing in a unified manner, let us define the functions

$$V^*(m, f) \equiv V(p^*(m, f, \sigma^*(m, f), \tau^*(m, f)), f + \tau^*(m, f), \sigma^*(m, f))$$

and

$$H^*(m, f) \equiv H(\tau^*(m, f) + \sigma^*(m, f) + f),$$

which describe the equilibrium threshold valuation and the share of cash transactions, as a function of the fee structure (m, f). Obviously, by setting $\tau^* \equiv 0$ or $\sigma^* \equiv 0$, we recover the scenarii in which only cash discounts or only card surcharges are permitted.

A platform that operates as an open system therefore solves

$$\max_{\{m,f\}} \quad (1 - G(V^*(m,f))) \cdot (1 - H^*(m,f)), \tag{15}$$

subject to the zero-profit constraint, $m + f = c$, and, in the case of cash discounts, the card acceptance constraint, $m \leq \mathbb{M}^d(f)$. The next proposition derives the platform's optimal fee structure when merchants practice discriminatory pricing.

Proposition 6 ***(equilibrium under discriminatory pricing)*** *Consider an open system that maximizes the number of card transactions.*

1. *When cash discounts are allowed, the equilibrium effective merchant fee is*

$$\hat{m}^d = \min\left\{\tilde{m}^d, m^u\right\}.$$

In particular, the effective merchant fee strictly violates the tourist test: $\hat{m}^d > b_S$. Relative to uniform pricing, the effective merchant fee weakly decreases, $\hat{m}^d \leq m^u$, the effective cardholder fee weakly increases, $\hat{f}^d \geq f^u$, and social welfare weakly increases.

2. *When card surcharges are allowed, and assuming the absence of missed sales ($s \geq S^s(\hat{m}^s, c - \hat{m}^s)$), the equilibrium effective merchant fee is $\hat{m}^s$. In particular, the effective merchant fee strictly satisfies the tourist test: $\hat{m}^s < b_S$. Relative to uniform pricing, the effective merchant fee strictly decreases, $\hat{m}^s < m^u$, and the effective cardholder fee strictly increases, $\hat{f}^s > f^u$. Social welfare strictly decreases if $m^u < \bar{m}^s$, and strictly increases if $m^u > \bar{m}^s$.*

Proof (sketch). Let us consider first the scenario of card surcharges. It is convenient to think of the platform as choosing *effective* merchant fees (as opposed to nominal merchant fees), and then mapping the optimal effective merchant fee into its nominal counterpart. In the absence of missed sales, Proposition 5 reveals that, in equilibrium, the effective merchant fee $\hat{m}$ satisfies $\hat{m} \leq \hat{m}^s$. In particular, this condition is satisfied as a strict inequality if the *nominal* merchant fee satisfies $m < \hat{m}^s$ (in which case surcharges do not happen and $\hat{m} = m$), and as an equality if $m \geq \hat{m}^s$ (in which case surcharges are weakly positive). Intuitively, the merchant's surcharging behavior imposes an upper bound on the effective merchant fee that the platform can implement.

From Lemmas 2 and 3 the volume of consumers visiting the store is an inverse-U-shaped function of $\hat{m}$ with strict maximum at $\hat{m} = b_S$. Because increasing $\hat{m}$ also increases the share of card transactions (as $\hat{f} = c - \hat{m}$ goes down), it follows that the *volume* of card transactions is maximized at the effective merchant fee $\hat{m} = \hat{m}^s < b_S$. Setting the *nominal* merchant fee to satisfy $m \geq \hat{m}^s$ obviously implements the effective merchant fee $\hat{m}^s$ (by part 3 of Proposition 5).

Finally, that $s \geq S^s(\hat{m}^s, c - \hat{m}^s)$ guarantees that no missed sales happen on the equilibrium path. The proof contained in the Appendix completes the arguments above by showing that no fee structure that generates missed sales can be optimal for the platform. We also relegate to the Appendix the proof of claim 1, pertaining to cash discounts, as the arguments in this proof are analogous to the ones for card surcharging. ■

As the proof sketch reveals, allowing for card surcharges has the same effect as that of an *upper bound on the effective merchant fee that the platform can implement*. The reason is that the merchant, by surcharging card transactions, has full control over his effective fee, $m - \tau$, and the effective cardholder fee, $f + \tau$. As the merchant "overshoots" in surcharges relative to efficiency (in order to extract more rents from consumers), the platform is unable to implement an effective merchant fee *at or above the tourist test level*. As a consequence, allowing for card surcharges leads to an inefficiently low volume of card transactions.

As a result, lifting the no-surcharge rule substitutes one inefficiency (underuse of cards due to inefficiently high card surcharges) for another (overuse of cards due to low merchant resistance under uniform pricing). Therefore, welfare can either decrease or increase, depending on how m^u compares with the threshold $\bar{m}^s$, identified in Corollary 3.

By contrast, allowing cash discounts is unambiguously welfare-enhancing. As in the case of card surcharges, allowing cash discounts has the same effect as that of an upper bound on the effective merchant fees that the platform can implement. Because merchants "undershoot" in discounts relative to efficiency, this upper bound, $\tilde{m}^d$, is strictly above the tourist test level. As a result, cash discounts reduce, but do not eliminate, an inefficiently high volume of card transactions. Welfare *strictly* increases with cash discounts if the upper bound $\tilde{m}^d$ is binding for the platform's problem in a world of uniform pricing. If this upper bound is slack, allowing for cash discounts affects neither the platform nor the behavior of the merchant (who does not find profitable to practice cash discounts).

Remark 4 (no card surcharging on the equilibrium path). Proposition 6 shows that, under discriminatory pricing, the platform's *effective* fee structure is uniquely determined in equilibrium. However, the platform's *nominal* fee structure is not uniquely pinned down, as, for instance, in the case of card surcharges, any (m, f) with $m \geq \hat{m}^s$ and $f = c - m$ maximizes the volume of card transactions. Yet, if surcharges involved any additional (unmodelled) costs to the merchant, the platform should select the fee structure $(m, f) = (\hat{m}^s, c - \hat{m}^s)$, which maximizes the volume of card transactions and *generates no card surcharges on the equilibrium path*. This consideration is likely to be relevant in explaining why card surcharges are rarely observed when allowed.[28]

Finally, the case of a closed system can be treated similarly to that of an open system. For brevity, consider the case of card surcharges. In this case, the platform chooses a fee structure

[28]Similarly, if cash discounts involved (unmodelled) practical inconveniences to the merchant, the platform should select the fee structure $(m, f) = (\hat{m}^d, c - \hat{m}^d)$, which maximizes the volume of card transactions and generates no cash discounts on the equilibrium path.

(m, f) so as to maximize profit:

$$\max_{\{m,f\}} \ (1 - G(V^*(m, f))) \cdot (1 - H^*(m, f)) \cdot (m + f). \tag{16}$$

Let, as before, $\tilde{c}$ denote platform's aggregate fee at the optimum. From Proposition 5, the effective merchant fee satisfies

$$\hat{m} \le M^s(\tilde{c} - \hat{m}),$$

as implied by the merchant's surcharging behavior. This constraint is the same as that for an open system, after replacing the total cost c by the aggregate fee $\tilde{c}$ (which accounts for the platform's markup).

It then follows from the same arguments in the proof sketch that the equilibrium effective merchant fee has to strictly satisfy the tourist test, and that card usage is inefficiently low. The welfare comparison relative to uniform pricing is however ambiguous, as, (i) relative to efficiency, surcharging leads to underuse of cards, while uniform pricing generates overuse of cards, and (ii) the platform's markup over each card transaction decreases in the presence of card surcharging.

6 Surcharge regulation

The analysis above shows that allowing merchants to practice cash discounts is unambiguously welfare-improving relative to uniform pricing. The same conclusion does not hold, however, in the case of card surcharges, which can be welfare-enhancing or detrimental depending on the magnitude of the "card-underuse" distortion that surcharging generates.

In the case of uniform pricing, one natural (and widely practiced) regulatory intervention is to cap interchange fees (or equivalently, merchant fees). It is worth pointing out that, if the no-surcharge rule is lifted, direct regulation of merchant fees *is not* an effective tool for increasing welfare. To understand why, consider an open-system platform that maximizes the number of card transactions. Even in the absence of caps on merchant fees, the merchant's surcharging behavior prevents the platform from inducing an effective merchant fee at or above the tourist test, and results in under-provision of card services. Therefore, further constraining the platform's choice of fee structure (by means of a binding cap on merchant fees) can only magnify the under-provision of card services. As a consequence, *when surcharging is allowed, capping interchange fees (or, equivalently, merchant fees) is unambiguously detrimental to welfare.*[29]

This observation suggests that, once the no-surcharge rule is lifted, regulation should focus on merchants' behavior rather than on the platform's behavior. In what follows, we discuss two

[29]More precisely, it is strictly detrimental if regulation caps the merchant fee at some $\mathring{m} < \hat{m}^s$, and neutral otherwise.

forms of merchant regulation that have received much attention in the policy debate regarding surcharges.

6.1 Surcharging caps

A simple, and warmly debated, regulatory intervention is to impose by law an upper bound on the surcharge that merchants can practice. In Denmark, for example, where merchants are allowed to surcharge credit card transactions, the government capped surcharges by 3.75% of the retail price. As discussed in the introduction, the US, the UK and Australia all have or will have rules that limit the surcharge to a level set at or above m (depending on the interpretation).[30]

In light of Proposition 6, one should expect caps on surcharges to reduce the downward pressure on equilibrium effective merchant fees. Setting the "right" cap seems therefore a crucial component of a successful regulatory intervention. The next proposition shows that the welfare-maximizing outcome can be attained if surcharges are capped by the "net merchant fee", $m - b_S$.

Proposition 7 ***(optimal surcharging cap regulation)*** *Consider a platform that operates as an open system, and assume that card surcharges are allowed. Let the regulator constrain the merchant's ability to surcharge according to the following cap regulation:*

$$\tau \leq \max \{m - b_S, 0\} . \tag{17}$$

Then the social welfare achieved in equilibrium is maximal.

Recall from Proposition 5 that, absent a cap, merchants choose inefficiently high surcharges in order to extract rents from consumers with high inconvenience of cash payments. The idea behind the cap regulation from Proposition 7 is precisely to eliminate the surcharge "overshooting" by merchants. To see how it works, note that, for any $m > b_S$, merchants choose the maximal allowed surcharge, leading to an effective merchant fee that exactly satisfies the tourist test:

$$m - (m - b_S) = b_S.$$

[30] Since March 2013, the Reserve Bank of Australia (RBA) empowered card platforms to limit a merchant's surcharge to "the reasonable cost of acceptance". In the UK, the Office of Fair Trading promulgated a ban on excessive credit and debit card charges that came into force on April 2013. Merchants are now forbidden to charge customers more than what it costs them to process the payment. The European Consumer Rights Directive 2011/83/UE (article 19) similarly states that above-cost surcharges will be prohibited for all payment methods. In the US, Visa and MasterCard agreed in November 2012 to alter their rules concerning "check-out fees". Namely, these platforms must allow surcharging for credit cards (but not for debit cards), but can limit surcharging to the cost incurred by the merchant from a card payment (and never above 4%). Systems such as Amex are not directly affected by this rule, but merchants who surcharge on Visa or Master Card must apply an equivalent surcharge on all credit cards that charge equal or higher merchant fees.

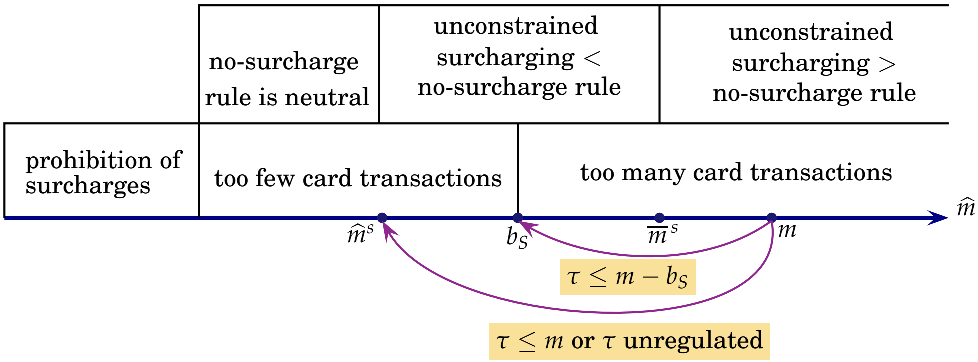

Figure 4: Effective merchant fee as a function of the nominal merchant fee under unconstrained surcharging and cap regulation

As a consequence, at a fee $m > b_S$, the effective card fee faced by consumers is at its efficient level:

$$f + (m - b_S) = c - b_S,$$

where the equality follows from the platform's zero-profit condition. By preventing surcharges from exceeding the net merchant fee $m - b_S$, the regulator eliminates surcharge overshooting, and achieves the welfare-maximizing outcome. In particular, under cap regulation, setting $m = b_S$ is optimal for the platform, in which case surcharges would not happen on the equilibrium path. Moreover, this choice of merchant fee is the unique optimum if surcharges involve any additional (unmodelled) costs to the merchant. This logic is illustrated in Figure 4.

There is an obvious caveat to the applicability of surcharging caps. Setting the "right" surcharge cap, at $m - b_S$, requires that the regulator knows the merchant's convenience cost of cash benefits, b_S. Knowledge of b_S, however, is not easily available to regulators, and its measurement is subject to debate, as the recent experience of interchange fee regulation according to the tourist test attests.

Finally, in the case of a closed system, the use of surcharging caps entails an effect that is absent in the open-system case. Namely, when surcharging is constrained, the platform has an incentive to increase its aggregate fee, $m + f$, over each card transaction (double marginalization). The market power enjoyed by the card platform then generates a countervailing effect that moves the equilibrium away from first best. In this case, determining the optimal cap on surcharges is more difficult, as it requires making assumptions about the degree of market power enjoyed by the card platform vis-a-vis cardholders and merchants. If the platform is subject to a high degree of competition, though, a surcharging cap close the net merchant fee, $m - b_S$, is likely to produce large welfare gains.

6.2 Mandated transparency

We assumed so far that, before incurring the shopping cost s, the consumer observes the retail price p, but does not observe the merchant's cash/card policy. We argued that this is a reasonable assumption in a number of contexts (for example, when markets are characterized by one-time or infrequent shopping, or when consumers are rationally inattentive). This assumption also fits well with the "hold-up" narratives that have led authorities around the world to reconsider their policies toward surcharging. Most importantly, we showed that, due to "hold-up" at the point of sale, merchants are able to impose high card surcharges. This naturally impacts the equilibrium fee structure: In response to the *possibility* of card surcharges, in equilibrium, the effective merchant fee decreases and the effective cardholder fee increases (relative to uniform pricing).

In the last subsection, we argued that properly chosen surcharging caps are an effective tool for achieving first best, as they eliminate the inefficiently high surcharges imposed by merchants. Antitrust authorities are contemplating an alternative antidote to excessive surcharges, "mandated transparency", in which the merchant would be obliged to post the level of surcharge *together* with the price.

To understand the effects of transparency, we will now study the merchant's problem in a setting of perfect consumer information. Because consumers observe surcharges together with price, the volume of consumers who visit store is sensitive to the *actual* (as opposed to the anticipated) surcharge practiced by the merchant. Accordingly, in the absence of missed sales, the merchant chooses a retail price p and a surcharge τ to maximize

$$(1 - G\left(V(p, f + \tau, 0)\right)) \cdot (p - T_S(m, f, 0, \tau))\,. \tag{18}$$

The next proposition, a variant of Rochet and Tirole (2002), shows that mandated transparency leads to efficient card usage.

Proposition 8 ***(mandated transparency)*** *Consider a platform that operates as an open system, and assume that card surcharges are allowed. Let the regulator oblige merchants to post the card surcharge jointly with the retail price. Then, in equilibrium, card usage is efficient.*

Transparency regulation gives the merchant the ability to commit to a card surcharge. This commitment technology enables the merchant to implement the following pricing strategy: He employs the card surcharge to minimize the aggregate transaction costs, and employs the retail price to extract rents from buyers. As a result, card usage is efficient (but, of course, retail prices display the usual market-power distortion). This result echoes, in our payment-choice application, the classical insight of Spence (1975), according to which "quality" (in our case, payment choice) is efficiently provided when "marginal" and "infra-marginal" consumers have similar tastes for quality (payment choice).

The independence between the consumers' valuations for the good and their inconvenience costs of cash payments is crucial to the efficiency result of Proposition 8. Arguably, one would expect richer consumers (i.e., those with a less elastic demand curve, and therefore a higher valuation) to also be on average those who have the highest opportunity cost of fetching and managing cash. This suggests that, under positive correlation, card surcharges can be used by the merchant as an instrument to capture consumer surplus (even with transparency). The next proposition analyzes an interesting tractable case that confirms this intuition.

Proposition 9 ***(Spencian distortions and transparency regulation)*** *Assume that consumers' valuations are perfectly positively correlated with their inconvenience costs of cash payments, and consider an equilibrium where both payment methods are used. Transparency regulation then results in inefficiently high card surcharging and inefficiently low card usage.*

The understand the result above, note that that, under perfect positive correlation between v and b_B, the lowest-valuation consumers to visit the store strictly prefer paying by cash than by card (as they have the lowest b_B among all those who purchase). As a consequence, at the optimum, card surcharges do not affect the volume of consumers visiting the store, but only the fraction of high-valuation consumers who decide to pay by card. By raising card surcharges above the efficient level, the merchant can therefore extract more rents from high-valuation consumers (who consider cash payments way too inconvenient).[31]

More generally, whenever "marginal" consumers choose cash payments more often than the "infra-marginal" ones, card surcharging *cum* transparency regulation will lead to inefficiently low use of cards.[32] In addition, the magnitude of the distortion relative to first best can only be assessed with precise knowledge of the joint distribution of v and b_B.

The limits of voluntary and mandated transparency

Interestingly, Proposition 8 implies that if the merchant could *costlessly* communicate the card surcharge to consumers (before visiting choices are made), he would find profitable to do so. Indeed, by committing not to surcharge abusively, the merchant is able to attract more consumers to the store and increase profit. By contrast, in reality merchants rarely communicate surcharges *before* the point of sale.

There are several reasons for this discrepancy. The existence of unaware consumers, who do not factor in card surcharges when deciding whether or not to visit the store, is likely to be an important one. Disclosing card surcharges works as an "eye-opener" to unaware consumers, who, following disclosure, make store visiting decisions with correct expectations about effec-

[31]We introduced correlation between v and b_B in order to qualify the efficiency result of Proposition 8. It should be noted that the qualitative insights from the baseline model (pertaining cash discounts, card surcharges, equilibrium fees and surcharging caps) are robust to the distributions G and H being positively correlated.

[32]For brevity, Proposition 9 considered the case where the correlation between v and b_B is positive and perfect. Under appropriate regularity conditions, this result can be easily extended to the case of imperfect correlation.

tive cardholder fees. As studied in section 7, disclosure is likely to reduce demand and profits for the merchant.

Another prominent reason for why voluntary disclosure of surcharges is rare (as well as for for why mandated transparency regulation might have limited effects) is the existence of considerable menu costs for the merchant, who needs to post next to the price, or more realistically as a general policy displayed prominently in the store, the surcharges for the different types of cards he accepts (debit/credit, various card associations and proprietary systems). Furthermore, inattentive consumers may overlook the card surcharging announcement, or be overloaded by its information. For instance, consumers may enter the shop thinking of a purchase for which they have enough cash and then discover that they want to buy a more expensive or an additional item.

Another issue is that price publicity may not come from the merchant. For example, the product manufacturer may run a national advertising campaign, where it is infeasible to disclose the policies of all retailers carrying the product. A similar issue arises when the consumer learns the retailer's price through a price-comparison engine; as is well-known, such comparisons are multidimensional as they depend on the type of purchase/consumer. So, if websites pick the price for the bare-bones product (including a payment by cash) as they often do, there is still an element of hold-up in the retailer's choice of surcharge. In the end, neither surcharge capping nor mandatory transparency is a perfect regulatory response to the inefficiencies attached to surcharging, and we therefore can think of the two regulations as complements.

7 Discussion

In this section, we investigate how our conclusions are affected by (i) alternative assumptions about consumer rationality, and (ii) the possibility that consumers adjust their choice of cash holdings in response to cash/card policy anticipated from the merchant. For brevity, we shall restrict attention to the card surcharge scenario in an open system.

7.1 Unaware Consumers

The results of this paper were derived under the strong assumption that consumers correctly anticipate the cash discount/card surcharge practiced by the merchant. This modeling choice was deliberate, as regulatory interventions under rational expectations are less likely to be needed than under weaker forms of consumer rationality.

We now investigate how the insights of our analysis change if we assume that consumers are *unaware* of card surcharges. Formally, this means that consumers make their decisions to visit the store assuming that card surcharges are always zero, i.e., $\tau^a \equiv 0$. To simplify the technical exposition, let us assume in this subsection that the distribution G has a constant hazard rate $\lambda^g > 0$.[33]

Unlike in the case of rational expectations, unaware consumers misperceive their effective cardholder fee, as they do not factor in surcharges and mistake the nominal cardholder fee for its effective counterpart. This generates an incentive for the merchant to conceal surcharges before the point of sale, and for the platform to set low nominal cardholder fees, so as to attract a large volume of consumers to visit the store.

In particular, the platform unambiguously gains by decreasing nominal cardholder fees at any fee structure that induces *no* missed sales. To understand why, note that, in the *absence of missed sales*, the merchant's optimal card surcharge is the same as in the case of rational expectations (characterized in Proposition 5).[34] Therefore, absent missed sales, the *effective* fee structure is independent of the *nominal* fee structure, and equals $(\hat{m}^s, \hat{f}^s)$. This implies that the share of card transactions is constant in the nominal cardholder fee, as it only depends on the effective cardholder fee $\hat{f}^s$ that consumers face *once* in the store. By reducing the nominal cardholder fee, the transactions costs anticipated by consumers strictly decrease. Because retail prices remain constant at

$$p^\star = T_S(\hat{m}^s, \hat{f}^s, 0, 0) + \frac{1}{\lambda^g},$$

the number of consumers visiting the store, and the volume of card transactions, strictly increases as the nominal cardholder fee decreases.

As consequence, no fee structure (m, f) such that the no-missed-sales constraint (3) is slack can be optimal for the platform. This implies that the platform-optimal nominal cardholder fee satisfies $f \leq \bar{f}$, where the threshold $\bar{f}$ is such that the no-missed-sales constraint (3) is satisfied with equality:

$$s + \mathbb{E}\left[\min\left\{b_B - \bar{f}, 0\right\}\right] = 0.$$

For $f < \bar{f}$, the merchant incurs missed sales, and the effective cardholder fee $\hat{f}^s$ is strictly increasing in the nominal fee f (i.e., the pass-through is imperfect). As a result, at any $f < \bar{f}$, the platform faces a genuine trade-off between increasing the share of card transactions (by means of a lower effective cardholder fee), and decreasing the volume of consumers visiting the store (as retail prices increase as the transaction costs faced by merchants increase).[35]

Therefore, relative to rational expectations, the welfare-detrimental effects of hold-ups are magnified (as too many consumers visit the store). The case for surcharging caps is therefore stronger when consumers are unaware of card surcharges than under rational expectations.

[33]This is equivalent to G being an exponential distribution with mean $\frac{1}{\lambda^g}$. A constant hazard rate guarantees that aware and unaware consumers have the same demand semi-elasticity as long as the missed-sales constraint is not binding.

[34]This results from the fact that, because consumers do not observe the merchant's cash/card policy before visiting the store, card surcharges are chosen to minimize the merchant's transactions costs *per sale*. As a consequence, the optimal card surcharge for the merchant is independent of how consumers form expectations about his cash/card policy.

[35]We cannot however rule out a kink solution at $\bar{f}$.

7.2 Endogenous Cash Holdings

We have assumed that the distribution of costs b_B of cash payments is exogenous. In fact, it results from consumer choice, at least for brick-and-mortar purchases: Consumers can decide how much cash to hoard in their wallet or whether to carry a checkbook. More interestingly, that decision hinges on the *anticipated* cash/card policy of the merchant. Intuitively, if the merchant is expected to levy steep card surcharges, consumers are more likely to carry cash, which will limit the merchant's ability to impose surcharges.

To formalize the consumers' cash-holdings decision, let us assume that consumers select an intensity θ of cash-holdings attention, which determines the ex-post distribution of the inconvenience cost of cash transactions, denoted by $\tilde{H}(b_B|\theta)$. We let $\tilde{h}(b_B|\theta)$ be the density of $\tilde{H}(b_B|\theta)$ and assume that its hazard rate $\frac{\tilde{h}(b_B|\theta)}{1-\tilde{H}(b_B|\theta)}$ is continuously differentiable and strictly increasing in θ. Accordingly, a larger cash holding intensity leads to a decrease in the distribution of the inconvenience cost of cash transactions in the sense of hazard-rate dominance (and therefore also in the sense of first-order stochastic dominance). A cash holding attention intensity θ generates a nuisance costs $I(\theta)$ to consumers. The function I is strictly increasing, weakly convex, and continuously differentiable with $I'(0) = 0$.

We assume that consumers take the decisions of whether to visit the store and of how much cash to hold simultaneously. An *equilibrium with endogenous cash holdings* adds to the equilibrium definition of Section 2 the requirement that the consumers' cash holdings decision and the merchant's cash/card policy constitute mutual best responses.

To simplify the exposition, let us consider the case where missed sales do not occur (which is assured to hold provided the shopping cost s is large enough). Accordingly, when consumers anticipate the card surcharge τ^a, they choose θ so as to solve

$$\max_{\theta}\left\{-I(\theta) - \int_{\underline{b}_B}^{f+\tau^a} b_B d\tilde{H}(b_B|\theta) - \left(1 - \tilde{H}(f+\tau^a|\theta)\right)\cdot[f+\tau^a]\right\}.$$

The objective above is continuously differentiable in θ and exhibits strictly increasing differences with respect to $(\theta, f+\tau^a)$. Therefore, the optimal choice of cash holdings, denoted by $\theta^*(f+\tau^a)$, is strictly increasing in the effective consumer fee *anticipated* by consumers, $f+\tau^a$.

In turn, as established in Proposition 5, the merchant optimal card surcharge is

$$\tau^{\ddagger}(m, f, \theta^a) = \max\{m - \check{m}^s(\theta^a), 0\},$$

where the threshold $\check{m}^s(\theta^a)$ solves

$$\check{m}^s(\theta^a) = b_S - \frac{1-\tilde{H}(c-\check{m}^s(\theta^a)|\theta^a)}{\tilde{h}(c-\check{m}^s(\theta^a)|\theta^a)}.$$

Because $\check{m}^s(\theta^a)$ is strictly increasing in θ^a, card surcharges are weakly decreasing in the *anticipated* cash holding intensity θ^a. In equilibrium, the consumers and the merchant correctly anticipate each others' choices: $\theta^a = \theta^*(f + \tau^\ddagger(m, f, \theta^a))$. It is straightforward to show that as the cost of cash holding intensity decreases (as captured, for instance, by a decreasing α in the parametrized cost function $\alpha I(\theta)$), consumers hold larger cash holdings and merchants levy smaller card surcharges.[36]

Most importantly, the effective merchant fee prevailing in any equilibrium with endogenous cash holdings is, as before, strictly below the tourist test level b_S. Accordingly, the analysis of Proposition 6 and the regulatory prescriptions of Section 6 hold unchanged, modulo the qualification that the distribution the inconvenience cost of cash transactions is now endogenous. In particular, as access to cash becomes easier, the merchant's ability to hold up consumers at the point of sale is diminished. As a consequence, the welfare loss generated by hold up at the point of sale is reduced.

8 Predictions and policy implications

Our analysis unveils a rich set of predictions that are consistent with evidence, as well as policy insights that question current regulatory practices. This overview particularly emphasizes surcharging, both because the theory predicts that cash discounts will be infrequent if surcharging is prohibited and inexistent if surcharging is allowed, and because surcharging, unlike discounting, is a hotly debated policy concern.

Consequences of the no-surcharge rule

1) *Missed-sales concerns generate must-take cards. That is, merchants will accept the card even when the merchant fee exceeds their convenience benefit of card payments. This must-take-cards channel, which relies on consumers being imperfectly informed, is novel and complements that based on the merchants' attractiveness concerns, which requires that consumers are well-informed about individual merchants' card-acceptance policy.*

The importance of missed sales is underscored by the analysis of Bolt et al (2010), who, using survey data from the Netherlands, document that 5% of consumers reported leaving a merchant's store without purchasing when faced with card refusal or steep card surcharges. Given the relative magnitude of markups and merchant fees, such a fraction is likely to raise a significant concern for merchants.

2) *There is less merchant resistance (the card acceptance threshold increases) when cards are more valuable to consumers relative to cash. Therefore, for a given fee structure, there is less merchant resistance to credit than to debit cards.*

[36] Note that there might be multiple equilibria with endogenous cash holdings. The comparative statics above holds for each equilibrium.

3) *Under weak conditions, the payment platform optimally sets the merchant fee at the merchants' acceptance threshold, thereby making cards must-take cards. Moreover, as a result of lower merchant resistance, the platform charges a merchant fee higher for credit than for debit cards.*

This prediction is in line with the fact that Visa's and MasterCard's interchange fees (and therefore the merchant fee) have always been higher under the no-surcharge rule for credit than for debit.

Consequences of discriminatory pricing

4) *Merchants always opt for a card surcharge (a levy) over a cash discount (a giveaway).*

Bolt et al (2010) provide empirical support to this prediction. Using consumer and retailer survey data from the Netherlands (where both cash discounts and card surcharges are legal), they show that around 22% of Dutch retailers practice card surcharges, while no retailers in their sample practice cash discounts. They also provide empirical support for abusive surcharges by reporting an average debit card surcharge of 2.3% among merchants who surcharge. As merchant fees are around 1% for debit cards, the effective merchant fee faced by these merchants is on average negative.

5) *Surcharging always generates too few card transactions, both from the point of view of the payment platform, which therefore prefers to prohibit surcharging, and from the point of view of the social planner.*

Cash/card steering effects are crucial for this conclusion. Using payment data from Dutch merchants, Bolt et al (2010) provide evidence of significant cash/card steering: They estimate that removing card surcharges should increase the share of card payments on average from 36% to 44%.

6) *Card surcharging is more frequent when cards are more valuable to consumers.*

Prediction 6 is broadly consistent with available evidence.[37] First, credit card surcharges are more frequent than debit card surcharges. Second, Internet transactions, for which card alternatives are costly to the consumer, are prone to surcharging. Card surcharges are for instance commonplace for low-cost online bookings, in which customers sink a substantial time cost only to find out in the last (payment) step that a surcharge is levied. A dramatic illustration is provided by eDreams, which used to surcharge up to 18% for customers paying by card. Third, transactions that involve customers on business cards (who therefore have a deflated personal cost of surcharges and a cost of cash payments inflated by the need to obtain reimbursement) are also conducive to surcharging.

Moreover, the industries where credit card surcharges are most often experienced are air travel, holiday travel, restaurants, taxis and gas stations (see Bolt et al (2010) and Choice (2014) for empirical evidence of this point in the Netherlands and Australia, respectively). These

[37]Much is this evidence concerns Australia, which authorized surcharging for debit and credit cards in 2003. See, e.g., Bagnall et al (2011), Evans (2011), and Choice (2014).

industries exhibit one-time or infrequent shopping, and appear to fit well the assumption that consumers have imperfect information regarding the merchant's cash/card policy.

7) *In response to laws allowing card surcharging, effective merchant fees decrease and effective cardholder fees increase. Moreover, if surcharges involve any extra convenience cost for merchants, the payment platform should optimally choose merchant and cardholder fees in a way that surcharges do not occur. Consequently, public regulations authorizing card surcharging do not generate much actual surcharging.*

Prediction 7 sheds light on the puzzling fact that surcharging is rather rare when allowed.[38] The model with homogenous merchants predicts no surcharging at all; merchant heterogeneity would deliver some surcharging, but still a limited amount of it. This suggests that the (somewhat sparse) evidence of card surcharging is only the "tip of the iceberg", and that it therefore understates the global impact of card surcharging in the payment system.

Relatedly, interchange fee regulation and authorization of surcharging are substitute instruments. Take the 2003 reduction in the interchange fee mandated by the Reserve Bank of Australia (RBA). For credit cards, the three bank associations, which had set the interchange fee at around 0.95 percent of the transaction value, were forced to reduce their interchange fee to around 0.55 percent.[39] This reduction was concomitant with a regulation authorizing surcharging. Our theory predicts that, up to some caveats, a partial decrease in the interchange fee might have occurred anyway in reaction to the introduction of surcharging.[40]

Policy Implications

Finally, our policy conclusions challenge received wisdom concerning regulation:

8) *Surcharging authorization cum regulation is redundant if the merchant fee is regulated at its tourist test level, namely the merchant's convenience benefit from card payments b_S.*

9) Allowing card surcharging increases social welfare if and only if the merchant fee under uniform pricing much exceeds the tourist test level.

10) *If the no-surcharge rule is lifted, interchange fee (or merchant fee) regulation is detrimental to welfare. Regulation should focus on merchants' behavior, rather than on the platform's behavior.*

11) *If surcharging is to be allowed, the optimal cap is equal to the merchant fee minus the merchant's convenience benefit from card payments: $\tau \leq m - b_S$. Recent cost-based regulations are more lenient, as they require that $\tau \leq m$ or that $\tau \leq m + X$, with $X > 0$;*

[38] See Evans (2011, page 23).

[39] See Evans (2011, chapter 3) for a detailed analysis of the impact of the interchange fee reduction in Australia.

[40] We are cautious in this statement for two reasons. First, the 0.95 percent rate probably was lower than the associations' desired interchange fee as they may have (unsuccessfully) tried to avoid regulation. Second, merchants in Australia often demand a "blended surcharge". As Amex and Diners Club charge substantially higher merchant fees, the level of the interchange fee has less impact on the surcharge than is predicted by our theory. Overall, only 5% of transactions in Australia involve a surcharge. Also consistent with the theory is the fact that reward cards have become less attractive in Australia.

12) *Mandated transparency regulation eliminates hold-ups for attentive consumers; however, it (i) may not be feasible (as when the consumer learns price through a national advertising campaign or a price comparison engine), (ii) may involve transaction costs for the merchant, (iii) does not address the existence of inattentive consumers, and (iv) does not prevent inefficient surcharging if the consumer's willingness to pay is correlated with his desire to use the card (as we argue is likely to be the case). Thus transparency is not a perfect regulatory response to the inefficiencies attached to surcharging, and the two regulations may well be complementary.*

Appendix: Omitted Proofs

The appendix collects all proofs omitted in the text.

Proof of Lemma 2. Using the definition of $\hat{m}$ and $\hat{f}$, it follows that

$$T(m, f, \sigma, \tau) = \int_{\underline{b}_B}^{\hat{f}} (b_B + b_S)\, dH(b_B) + [1 - H(\hat{f})] \cdot [\hat{m} + \hat{f}].$$

Let us fix the platform's aggregate fee: $m + f = \hat{c}$. Therefore the effective cardholder fee that minimizes $T(m, f, \sigma, \tau)$ solves

$$\min_{\hat{f}} \int_{\underline{b}_B}^{\hat{f}} (b_B + b_S)\, dH(b_B) + [1 - H(\hat{f})] \cdot \hat{c}.$$

Fixing $\hat{f}$, the objective above is strictly increasing in $\hat{c}$, what implies that aggregate transaction costs decrease with the platform's aggregate fee. The derivative of the objective above with respect to $\hat{f}$ is

$$h(\hat{f}) \cdot \left(\hat{f} - \hat{c} + b_S\right),$$

which is strictly positive if $\hat{f} - \hat{c} = \hat{m} > b_S$, strictly negative if $\hat{f} - \hat{c} = \hat{m} < b_S$ and zero if $\hat{f} - \hat{c} = \hat{m} = b_S$. Therefore, any quadruple (m, f, σ, τ) satisfying (7) is a global minimizer of (6). Note that Assumption 1 guarantees that missed sales do not happen whenever (7) holds. The argument above then establishes that condition (7) leads to minimal aggregate transaction costs among all quadruples (m, f, σ, τ) such that missed sales do not happen. Because $V(p, f + \tau^a, \sigma^a) > p$, any quadruple (m, f, σ, τ) such that missed sales happen leads to aggregate transactions that are no less than $T(b_S, c - b_S, 0, 0)$. It then follows that any quadruple (m, f, σ, τ) satisfying (7) is a global minimizer of aggregate transaction costs. ■

Proof of Lemma 3. Using (2) and the fact that missed sales do not occur, let us rewrite the merchant's profit function as

$$(1 - G\left(s + p + T_B\left(\infty, f + \tau, \sigma\right)\right)) \cdot (p - T_S(m, f, \sigma, \tau))$$

$$= (1 - G\left(s + p_B\right)) \cdot \left(p_B - T(m, f, \sigma, \tau)\right),$$

where the equality from the first to the second line follows from defining the *buyer effective price* $p_B \equiv p + T_B\left(\infty, f + \tau, \sigma\right)$. The optimum buyer effective price p_B^* can then be written as a univariate function of $T(m, f, \sigma, \tau)$. Moreover, that G has a weakly increasing hazard rate implies that $\dfrac{dp_B^*}{dT} \in (0, 1]$. Therefore, the threshold valuation

$$V(p^*(m, f, \sigma, \tau), f + \tau, \sigma) = s + p_B^*\left(T(m, f, \sigma, \tau)\right)$$

is a univariate and strictly increasing function of the aggregate transaction costs $T(m, f, \sigma, \tau)$. It is then obvious from (9) that social welfare is a univariate and strictly decreasing function of $T(m, f, \sigma, \tau)$. ■

The proof of the Proposition 1 uses the following convenient mathematical fact.

Lemma 5 *Let Φ be an absolutely continuous distribution with support $[\underline{v}, \infty)$ and density φ. For any $x > 0$, define the distribution Φ_x, with support $[\underline{v}, \infty)$, according to*

$$\Phi_x(v) \equiv \frac{\Phi(v + x) - \Phi(x + \underline{v})}{1 - \Phi(x + \underline{v})}.$$

Then the distribution Φ dominates Φ_x according to first-order stochastic dominance if

$$\frac{\varphi(v)}{1 - \Phi(v)}$$

is weakly increasing in v.

Proof. Because $\Phi_0(v) = \Phi(v)$, it suffices to show that $\Phi_x(v)$ is weakly increasing in x. A straightforward computation shows that the derivative of $\Phi_x(v)$ with respect to x is weakly positive if and only if

$$\frac{\varphi(v + x)}{1 - \Phi(v + x)} \geq \frac{\varphi(\underline{v} + x)}{1 - \Phi(\underline{v} + x)}$$

for all $v \geq \underline{v}$. That the inequality above holds is implied by the increasing hazard rate property, as stated. ■

Proof of Proposition 1. 1. Consider retail prices p and p' such that $p' > p$. We will show that $\Delta\pi(p, m, f) \geq 0$ implies that $\Delta\pi(p', m, f) > 0$. To this end, note that

$$\Delta\pi(p', m, f) = \int_{V(p,f,0)}^{\infty} \left[p' - T_S(m, f, 0, 0) - H(\tilde{v} - p) \cdot (p' - b_S)\right] dG(\tilde{v} + (p' - p)),$$

after the change of variables $\tilde{v} \equiv v - (p' - p)$. Therefore,

$$\Delta\pi(p', m, f) > \int_{V(p,f,0)}^{\infty} [p - T_S(m, f, 0, 0) - H(\tilde{v} - p) \cdot (p - b_S)]\, dG(\tilde{v} + (p' - p)),$$

where the inequality follows from the fact that $H(\tilde{v} - p) < 1$. Together with the fact that the integrand above is strictly decreasing in $\tilde{v}$, Lemma 5 implies that

$$\frac{\Delta\pi(p', m, f)}{1 - G\left(V(p, f, 0) + (p' - p)\right)} > \frac{\int_{V(p,f,0)}^{\infty} [p - T_S(m, f, 0, 0) - H(\tilde{v} - p) \cdot (p - b_S)]\, dG(\tilde{v})}{1 - G\left(V(p, f, 0)\right)}$$

$$= \frac{\Delta\pi(p,m,f)}{1-G\left(V(p,f,0)\right)}.$$

Therefore, $\Delta\pi(p,m,f) \geq 0$ implies that $\Delta\pi(p',m,f) > 0$. As such, for each (m,f), there exists $P^u(m,f)$ such that a card acceptance equilibrium exists if and only $p \geq P^u(m,f)$. Inspecting the definition of $\Delta\pi(p,m,f)$ reveals that $P^u(b_S,f) = b_S$ for all f. That $P^u(m,f)$ is strictly increasing in m follows from the fact that $\frac{\partial\Delta\pi}{\partial m}(p,m,f) < 0$.

2. In a card refusal equilibrium, all consumers with valuation $v \geq V(p,\infty,0)$ go to the store, where the threshold $V(p,\infty,0)$ is given by (2). Moreover, merchants have to find profitable to refuse cards given the consumers expectations of card refusal. Equivalently, the profit differential

$$\nabla\pi(p,m,f) \equiv \int_{V(p,\infty,0)}^{\infty} \left\{H(v-p)\cdot(p-b_S) - (p - T_S(m,f,0,0))\right\} dG(v) \geq 0,$$

where $T_S(m,f,0,0)$ is the merchant's expected transaction cost per sale under uniform pricing (as defined in (5)).

Consider retail prices p and p' such that $p' < p$. We will show that $\Delta\pi(p,m,f) \geq 0$ implies that $\Delta\pi(p',m,f) > 0$. To this end, note that

$$\nabla\pi(p',m,f) = \int_{V(p,\infty,0)}^{\infty} \left[H(\tilde{v}-p)\cdot(p'-b_S) - \left(p' - T_S(m,f,0,0)\right)\right] dG(\tilde{v}+(p'-p)),$$

after the change of variables $\tilde{v} \equiv v-(p'-p)$. Therefore,

$$\nabla\pi(p',m,f) > \int_{V(p,\infty,0)}^{\infty} \left[H(\tilde{v}-p)\cdot(p-b_S) - (p - T_S(m,f,0,0))\right] dG(\tilde{v}+(p'-p)),$$

where the inequality follows from the fact that $H(\tilde{v}-p) < 1$. Together with the fact that the integrand above is strictly increasing in $\tilde{v}$, Lemma 5 implies that

$$\frac{\nabla\pi(p',m,f)}{1-G\left(V(p,\infty,0)+(p'-p)\right)} > \frac{\int_{V(p,\infty,0)}^{\infty} \left[H(\tilde{v}-p)\cdot(p-b_S) - (p - T_S(m,f,0,0))\right] dG(\tilde{v})}{1-G\left(V(p,\infty,0)\right)}$$

$$= \frac{\nabla\pi(p,m,f)}{1-G\left(V(p,\infty,0)\right)}.$$

Therefore, $\nabla\pi(p,m,f) \geq 0$ implies that $\nabla\pi(p',m,f) > 0$. As such, for each (m,f), there exists $Q^u(m,f)$ such that a card acceptance equilibrium exists if and only $p \leq Q^u(m,f)$. Inspecting the definition of $\nabla\pi(p,m,f)$ reveals that $Q^u(b_S,f) = b_S$ for all f. That $Q^u(m,f)$ is strictly increasing in m follows from the fact that $\frac{\partial\nabla\pi}{\partial m}(p,m,f) > 0$.

3. First note from (2) that $V(p,\infty,0) > V(p,f,0)$, which implies that the card acceptance equilibrium Pareto dominates the card refusal equilibrium. We will now show that

$P^u(m,f) \leq Q^u(m,f)$. Because the integrand in $\Delta\pi(p,m,f)$ is strictly decreasing in v, it follows that $\nabla\pi(P^u(m,f),m,f) > 0$. As a result, $P^u(m,f) \leq Q^u(m,f)$, as claimed. ■

Proof of Corollary 1. Consider first the card acceptance equilibrium. As the distribution of consumers' inconvenience of cash payments shifts from H to $\hat{H}$, the valuation threshold $V(p,f,0)$ and the seller transaction costs $T_S(m,f,0,0)$ remain constant, as $H(b_B) = \hat{H}(b_B)$ for all $b_B \in [\underline{b}_B, f]$. Because $H(v-p) \geq \hat{H}(v-p)$ for all $b_B \in (f,+\infty)$, it follows that the profit differential $\Delta\pi(p,m,f)$ is higher under $\hat{H}$ than under H. This immediately implies that $P^u(m,f) \leq \hat{P}^u(m,f)$. The arguments for the card refusal are analogous and therefore omitted. ■

Proof of Proposition 2. Let us define

$$\pi(p,m,f,\sigma,\tau) \equiv (1 - G(V(p,f+\tau,\sigma))) \cdot (p - T_S(m,f,\sigma,\tau))$$

to be the merchant's profit when the retail price is p, the fee structure is (m,f) and the cash/card policy is (σ,τ). Notice that, by construction

$$\pi(p^*(m,f,0,0),m,f,0,0) \geq \pi(p,m,f,0,0)$$

for any $p \geq P^u(m,f)$. So consider deviations to a retail price $p < P^u(m,f)$. The most profitable deviation of this kind achieves a profit

$$\bar{\pi} \equiv \max_p \int_{V(p,\infty,0)}^{\infty} H(v-p) \cdot (p - b_S) dG(v),$$

where the threshold $V(p,\infty,0)$ is given by equation (2).

Note that

$$\pi(p^*(b_S,f,0,0),b_S,f,0,0) > \bar{\pi},$$

as $V(p,f,0) < V(p,\infty,0)$, $T_S(m,b_S,0,0) = b_S$, and $H(v-p) < 1$.

Let $\mathbb{M}^u(f)$ be the unique solution to the following equation in m:

$$\pi(p^*(m,f,0,0),m,f,0,0) = \bar{\pi}.$$

That the equation above has a unique solution follows from the fact that its left-hand side is strictly decreasing in m and that it converges to zero as $m \to \infty$. Clearly, $\mathbb{M}^u(f) > b_S$. Accordingly, no deviation involving a retail price different from $p^*(m,f,0,0)$ is profitable if and only if $m \leq \mathbb{M}^u(f)$. To conclude our equilibrium construction, note that $p^*(m,f,0,0) \geq$

$P^u(m, f)$ (as implied by the definition of $P^u(m, f)$). ■

Proof of Proposition 3. For the case of a closed system, let $c^u \equiv m^u + f^u$. It then follows that the pair (m^u, f^u) has to solve

$$\max_{\{m,f\}} \quad (1 - G(V(p^*(m, f, 0, 0), f, 0))) \cdot (1 - H(f))$$

subject to $c^u - f = m \leq \mathbb{M}^u(f)$. The problem above is that of an open system with cost c^u. The arguments in the proof sketch (regarding open systems) then establish the claim. ■

Proof of Lemma 4. 1. Because consumers do not observe the merchant's cash/card policy before visiting the store, the optimal cash discount, denote it σ^*, is independent of the retail price p. Therefore, for a fixed fee structure (m, f), a card acceptance equilibrium can be constructed identically as in Proposition 1, after one replaces b_B by its effective counterpart, $b_B + \sigma^*$, and b_S by its effective counterpart, $b_S - \sigma^*$. Therefore, there exists a threshold $P^d(m, f)$ such that a card acceptance equilibrium exists if and only if $p \geq P^d(m, f)$. That $P^d(m, f) \leq P^u(m, f)$ follows from the fact that the merchant's profits when cash discounts are possible are weakly larger than the profits under uniform pricing. The same arguments used in the proof of Proposition 2 then imply the existence of $\mathbb{M}^d(f) > b_S$ such that the merchant accepts cards if and only if $m \leq \mathbb{M}^d(f)$. That $\mathbb{M}^d(f) \geq \mathbb{M}^u(f)$ follows from (i) $P^d(m, f) \leq P^u(m, f)$, and (ii) that having the possibility to use cash discounts weakly increases the merchant's profit, relative to uniform pricing, at any fee structure (m, f).
2. When card surcharges are allowed, it is a weakly dominant strategy for merchants to accept cards, as they can always set $\tau = m - b_S$ to recover any losses from card transactions. ■

Proof of Proposition 4. The merchant chooses σ to solve:

$$\min_{\{\sigma \geq 0\}} \{H(\sigma + f)[\sigma + b_S] + [1 - H(\sigma + f)] \cdot m\}.$$

The latter objective function is quasi-concave, and so a necessary and sufficient condition for $\sigma^*(m, f) = 0$ is

$$h(f)(m - b_S) - H(f) \leq 0,$$

which holds if and only if $m \leq M^d(f)$, as announced in the proposition. Otherwise, $\sigma^*(m, f)$ is given by the first-order condition (13).

In the case of an open system, $m + f = c$. Therefore, cash discounts occur if and only if $m \geq M^d(c - m)$, which is equivalent to $m \geq \tilde{m}^d$. In this case, we have that

$$m - \sigma^*(m, f) - \frac{H(c - (m - \sigma^*(m, f)))}{h(c - (m - \sigma^*(m, f)))} = b_S,$$

which follows from (13). The equation above implies that, whenever $m \geq \tilde{m}^d$, $m - \sigma^*(m, f)$ is (i) independent of m, and (ii) equals $\tilde{m}^d$, as we wanted to show. ■

Proof of Proposition 5. The merchant chooses τ to solve:

$$\min_{\{\tau \geq 0\}} \{H(\tau + f)b_S + [1 - H(\tau + f)] \cdot [m - \tau]\}.$$

The latter objective function is quasi-concave, and so a necessary and sufficient condition for $\tau^*(m, f) = 0$ is

$$1 - H(f) - h(f)(b_S - m) \leq 0,$$

which holds if and only if $m \leq M^s(f)$, as announced in the proposition. Otherwise, $\tau^*(m, f)$ is given by the first-order condition (14).

In the case of an open system, $m + f = c$. Therefore, card surcharges occur if and only if $m \geq M^s(c - m)$, which is equivalent to $m \geq \hat{m}^s$. In this case, we have that

$$b_S - (m - \tau^*(m, f)) = \frac{1 - H(c - (m - \tau^*(m, f)))}{h(c - (m - \tau^*(m, f)))},$$

which follows from (14). The equation above implies that, whenever $m \geq \hat{m}^s$, $m - \tau^*(m, f)$ is (i) independent of m, and (ii) equals $\hat{m}^s$, as we wanted to show. ■

Proof of Corollary 2. Claim 2 follows from the comparison of

$$\min_{\{\sigma \geq 0\}} \{H(\sigma + f)[\sigma + b_S] + [1 - H(\sigma + f)] \cdot m\}$$

and

$$\min_{\{\tau \geq 0\}} \{H(\tau + f)b_S + [1 - H(\tau + f)] \cdot [m - \tau]\}.$$

The seller can always select $\tau = \sigma^*(m, f)$ (the optimal cash discount) and do at least as well as under a surcharge than under a discount. That the optimal card surcharges are weakly greater than the optimal cash discounts follows from the statements of Propositions 4 and 5. ■

Proof of Corollary 3. The proof follows immediately from the quasi-convexity of $T(m, c - m, 0, 0)$ around b_S. ■

Proof of Proposition 6. 1. It is convenient to think of the platform as choosing *effective* merchant fees (as opposed to nominal merchant fees), and then mapping the optimal effective merchant fee into its nominal counterpart. In the absence of missed sales, Proposition 4 reveals that, in equilibrium, the effective merchant fee $\hat{m}$ satisfies $\hat{m} \leq \tilde{m}^d$. In particular, this condition is satisfied as a strict inequality if the *nominal* merchant fee satisfies $m < \tilde{m}^d$ (in which case discounts do not happen and $\hat{m} = m$), and as an equality if $m \geq \tilde{m}^d$ (in which case discounts are weakly positive).

Consider first the case where $m^u < \tilde{m}^d$. In this case, cash discounts do not happen under the merchant fee m^u, and the constraint $\hat{m} \leq \tilde{m}^d$ is slack. Accordingly, the possibility of cash discounts plays no role in the platform's problem.

Now consider the case where $m^u > \tilde{m}^d$. In this case, the quasi-concavity of the objective (10) implies that the volume of card transactions is maximized at the effective merchant fee $\tilde{m}^d$. That total welfare goes up follows from the fact that aggregate transaction costs are weakly lower at the effective fee structure $(\hat{m}^d, c - \hat{m}^d)$ than at $(m^u, c - m^u)$. Finally, missed sales do not happen on the equilibrium path because $\hat{f}^d < c - b_S$ and Assumption 1.

The argument above establishes that $(\hat{m}^d, \hat{f}^d)$ maximizes the platform's objective among all effective fee structures such that missed sales do not happen. Because $V^*(m, f) > p$, any effective fee structure $(\hat{m}, \hat{f})$ such that missed sales happen leads to aggregate transactions that are no less than $T(b_S, c - b_S, 0, 0)$. Therefore, relative to $(b_S, c - b_S)$, a fee structure that leads to missed sales induces a smaller number of consumers visiting the store, and, because $m + f = c$, a smaller share of card transactions. Because the effective fee structure $(\hat{m}^d, \hat{f}^d)$ produces more card transactions than the efficient fee structure $(b_S, c - b_S)$, we conclude by transitivity that $(\hat{m}^d, \hat{f}^d)$ is a global maximum, as stated.

2. Because $s \geq S^s(\hat{m}^s, c - \hat{m}^s)$, the argument in the text establishes that $(\hat{m}^s, \hat{f}^s)$ maximizes the platform's objective among all effective fee structures $(\hat{m}, \hat{f})$ such that missed sales do not happen. As a consequence of $V^*(m, f) > p$, any effective fee structures $(\hat{m}, \hat{f})$ such that missed sales happen leads to aggregate transactions that are no less than $T(\hat{m}^s, \hat{f}^s, 0, 0)$. Therefore, relative to $(\hat{m}^s, \hat{f}^s)$, a fee structure that leads to missed sales induces a smaller number of consumers visiting the store, and, because $m + f = c$, a smaller share of card transactions. We then conclude that $(\hat{m}^s, \hat{f}^s)$ is a global maximum, as stated. ■

Proof of Proposition 7. The proof follows immediately from the arguments in the text. ■

Proof of Proposition 8. Let us fix some fee structure (m, f), with $m > b_S$, and assume (to be verified later) that missed sales do not happen on the equilibrium path. Define the buyer price p_B according to

$$p_B \equiv p + T_B(\infty, f + \tau, 0).$$

The merchant's problem can then be recast as

$$\max_{p_B, \tau} \left(1 - G\left(s + p_B\right)\right) \cdot \left(p_B - T(m, f, 0, \tau)\right).$$

Therefore, the merchant-optimal card surcharge is independent of the equilibrium price, and solves

$$\min_{\tau} T(m, f, 0, \tau).$$

Lemma 2 then implies that

$$\tau^{\dagger}(m, f) = m - b_S.$$

Therefore, for any fee structure (m, f) that violates the tourist test, it follows from Lemma 3 that card usage is efficient and the social welfare is maximal.

In equilibrium, by the same arguments of Proposition 6, the platform is indifferent between any fee structure such that $m \geq b_S$. As a result of the merchant's surcharging behavior, the effective merchant fee satisfies the tourist test with equality. Assumption 1 then implies that missed sales do not happen on the equilibrium path. ■

Proof of Proposition 9. For simplicity, let $b_B = \varepsilon v$, where $1 > \varepsilon > 0$. The willingness to pay v is as earlier distributed according to distribution G. A consumer of type v then purchases whenever:

$$v \geq p + \min \{f + \tau, \varepsilon v\}.$$

Thus, there are two cutoffs v^* and $v^{**} > v^*$, such that consumers in $[v^*, v^{**})$ purchase by cash and those with $v \geq v^{**}$ pay by card.

The merchant solves:

$$\underset{\{p,\tau\}}{\text{Max}} \left\{ [p - (m - \tau)] \left[1 - G\left(\frac{f + \tau}{\varepsilon}\right) \right] + [p - b_S] \left[G\left(\frac{f + \tau}{\varepsilon}\right) - G\left(\frac{s + p}{1 - \varepsilon}\right) \right] \right\},$$

yielding first-order conditions:

$$\frac{p - b_S}{1 - \varepsilon} = \frac{1 - G\left(\dfrac{s + p}{1 - \varepsilon}\right)}{g\left(\dfrac{s + p}{1 - \varepsilon}\right)},$$

and

$$m - \tau = b_S - \varepsilon \frac{1 - G\left(\dfrac{f + \tau}{\varepsilon}\right)}{g\left(\dfrac{f + \tau}{\varepsilon}\right)}.$$

The equation above reveals that the effective merchant fee has to strictly satisfy the tourist test in equilibrium, establishing the result. ■

References

Agarwal, S., Chomsisengphet S., Mahoney N., and J. Stroebel (2014) "Regulating Consumer Financial Products: Evidence from Credit Cards," mimeo.

Armstrong, M., and J. Vickers (2012) "Consumer Protection and Contingent Charges," *Journal of Economic Literature*, 50(2): 477–93.

Armstrong, M., Vickers, J., and J. Zhou (2009a) "Prominence and Consumer Search", *RAND Journal of Economics*, 40: 209-233.

Armstrong, M., Vickers, J., and J. Zhou (2009b) "Consumer Protection and the Incentive to Become Informed", *Journal of European Economic Association*, 7: 399-410.

Bagnall, J., Chong, S., and K. Smith (2011) "Strategic Review of Innovation in the Payments System: Results of the Reserve Bank of Australia's 2010 Consumer Payments Use Study," Reserve Bank of Australia.

Baxter, W. (1983) "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics*, 26: 541-588.

Bedre-Defolie, O., and E. Calvano (2013) "Pricing Payment Cards," *American Economic Journal: Microeconomics*, 5(3): 206-231.

Bolt, W., Jonker, N., and C. van Renselaar (2010) "Incentives at the Counter: An Empirical Analysis of Surcharging Card Payments and Payment Behavior in the Netherlands," *Journal of Banking and Finance*, 34: 1738-1744.

Carlton, D. W., and A. S. Frankel (1995) "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, 63(2): 643-668.

Chakravorti, S., and T. To (2007) "A Theory of Credit Cards," *International Journal of Industrial Organization*, 25: 583-595.

Choice (2014) "Credit Card Surcharging in Australia," report prepared on behalf of NSW Government Fair Trading.

Economides, N., and D. Henriques (2011) "To Surcharge or not to Surcharge? A Two-Sided Market Perspective of the No-Surcharge Rule," NET Institute Working Paper 11-03.

Edelman, B., and J. Wright (2014) "Price Coherence and Adverse Intermediation," mimeo.

Ellison, G. (2005) "A Model of Add-on Pricing," *Quarterly Journal of Economics*, 120: 585-637.

Evans, D. (2011) *Interchange Fees: The Economics and Regulation of What Merchants Pay for Cards*, Competition Policy International.

Evans, D., and R. Schmalensee (2005) "The Economics of Interchange Fees and Their Regulation: An Overview." Proceedings of a conference on "Interchange Fees in Credit and Debit Card Industries" (Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005): 77-120.

Gabaix, X., and D. Laibson (2006) "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121(2): 505-540.

Gans, J., and S. King (2003) "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis & Policy*, 3: 1-16.

Grubb, M. (2014) "Consumer Inattention and Bill-Shock Regulation," *Review of Economic Studies*, forthcoming.

Heidhues, P., Koszegi, B., and T. Murooka (2012) "Exploitive Innovation", mimeo.

Heidhues, P., Koszegi, B., and T. Murooka (2014) "Inferior Products and Profitable Deception", mimeo.

Kosfeld, M., and U. Schüwer (2011) "Add-on Pricing, Naive Consumers, and the Hidden Welfare Costs of Education", IZA Discussion Paper No. 6061.

Rochet, J.C., and J. Tirole (2002) "Cooperation among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics*, 33(4): 549-570.

——— (2011) "Must-Take Cards: Merchant Discounts and Avoided Costs," *Journal of the European Economic Association*, 9(3): 462-495.

Rochet, J.C., and J. Wright (2010) "Credit Card Interchange Fees," *Journal of Banking and Finance*, 34: 1788-1797.

Schmalensee, R. (2002) "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, 50: 103-122.

Schwartz, M., and D. Vincent (2006) "The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network," *Review of Network Economics*, 5(1): 71-102.

Shapiro, C. (1995) "Aftermarkets and Consumer Welfare: Making Sense of Kodak," *Antitrust Law Journal*, 63: 483-512.

Spence, M. (1975) "Monopoly, Quality, and Regulation," *Bell Journal of Economics*, 6: 417-429.

Tirole, J. (2009) "Cognition and Incomplete Contracts," *American Economic Review*, 99(1): 265-294.

——— (2011) "Payment Card Regulation and the Use of Economic Analysis in Antitrust," *Competition Policy International*, 7(1): 137-158.

Vickers, J. (2005) "Public Policy and the Invisible Price: Competition Law, Regulation and the Interchange Fee." Proceedings of a conference on "Interchange Fees in Credit and Debit Card Industries" (Federal Reserve Bank of Kansas City, Santa Fe, New Mexico, May 4-6, 2005): 231-247.

Wright, J. (2003) "Optimal Card Payment Systems," *European Economic Review*, 47: 587-612.

Wright, J. (2004) "The Determinants of Optimal Interchange Fees in Payment Systems," *Journal of Industrial Economics*, 52 (1): 1-26.

Wright, J. (2012) "Why Payment Cards Fees Are Biased Against Retailers." *Rand Journal of Economics*, 43 (4): 761-80.