Pricing Payment Cards

Özlem Bedre* Emilio Calvano†

First version: April 2008, This version: December 2009‡

Abstract

In a payment card association such as Visa, each time a consumer pays by card, the bank of the merchant (acquirer) pays an interchange fee (IF) to the bank of the cardholder (issuer) to carry out the transaction. This paper studies the determinants of socially and privately optimal IFs in a card scheme where services are provided by a monopoly issuer and perfectly competitive acquirers to heterogeneous consumers and merchants. Different from the literature, we distinguish card membership from card usage decisions (and fees). In doing so, we reveal the implications of an asymmetry between consumers and merchants: the card usage decision at a point of sale is delegated to cardholders since merchants are not allowed to turn down cards once they are affiliated with a card network. We show that this asymmetry is sufficient to induce the card association to set a higher IF than the socially optimal IF, and thus to distort the structure of user fees by leading to too low card usage fees at the expense of too high merchant fees. Hence, cap regulations on IFs can improve the welfare. These qualitative results are robust to imperfect issuer competition and to other factors affecting final demands, such as elastic consumer participation or strategic card acceptance to attract consumers.

JEL Classification: G21; L11; L42; L31; L51; K21.

Keywords: Payment card associations; Interchange fees; Merchant fees

*European School of Management and Technology - Berlin. ozlem.bedre@esmt.org

†Department of Economics, Bocconi University. emilio.calvano@unibocconi.it

‡The authors would like to thank B. Jullien, P. Rey, J.C. Rochet, seminar participants at MIT, U Rochester, Brown, Purdue, HEC-Montreal, U Toulouse, U College London, U Pompeu Fabra, INSEAD, U Maastricht, European School of Management and Technology, Norwegian Business School, Copenhagen Business School, Royal Holloway U London, DIW, U East Anglia (CCP), CEMFI, U Bocconi, U Bogazici, U Koc, European Central Bank, Richmond FED, Central Bank of Norway, Bank of Italy, EIEF, Luiss, Collegio Carlo Alberto, European University Institute for helping us improve this paper.

1 Introduction

Debit cards are expected to overtake cash as the primary method of payment by 2012. Spending at merchants on US credit, debit or prepaid cards topped 3.285 trillion dollars in 2008 (almost a quarter of the US GDP), up 6.1 percent from 2007. US retailers pay their banks on average 1.8 percent of every plastic card transaction to get the payment cleared.[1] Such fees are the second-highest expense for many businesses, after labor costs, exceeding the price of health care insurance for employees. Cardholders instead are typically offered complementary benefits and services to use their cards at checkout counters. In some cases, up to 5 percent of the value of the transaction is returned to consumers under the form of "cash back" bonuses.[2] This asymmetry is mainly due to the payment networks' practice of charging merchants' banks a per transaction fee (called the "swipe" or Interchange Fee) and turn over the proceedings to the cardholders' banks to increase the usage and issuance of cards.

The question that we address in this paper is whether the networks' pricing policies promote an efficient use of these payment instruments. We study this issue for two reasons. First, to date, there is still little consensus in the literature over the answer in spite of a recent flurry of research inspired by the two-sided markets framework.[3] [4] Second, along with the growth of card transactions has come greater scrutiny of the industry practices by public authorities. The concern is that these skewed pricing policies inflate retailers' costs of card acceptance without enhancing the efficiency of the system. Many authorities and legislators, including the European Commission and the U.S. Congress, either already intervened with specific interchange fee regulation or are considering the opportunity of doing so.[5]

This paper shows that the profit maximizing price structure subsidizes card usage too much at the expense of charging inefficiently high fees to merchants. This result is obtained in a very broad setup that nests several previous contributions. In contrast to existing formulations our model studies membership and usage choices simultaneously. This allows us to spell out

[1]Sources: Nilson Report (July 2009, issue 929, pp. 1,9); Nilson Report (April 2007, Issue 895, pp. 7); Hayashi (2009).

[2]For instance as of December 2009 BoA rewards usage of its check cards with frequent flyer miles, cash back bonuses (Keep the Change) and insurance against theft of the merchandise. http://www.bankofamerica.com/checkcard/. The AMEX Blue Cash Card offers 5 percent cash back as of January 2010.

[3]These are businesses that have to attract at least two groups of users to create value from a transaction, e.g., payment card networks, software platforms, dating clubs, etc. See Caillaud and Jullien (2003), Rochet and Tirole (2003, 2005, 2006) and Armstrong (2006) for pioneering contributions. Weyl (2009) provides an important recent contribution.

[4]See Chakravorti and To (2003), Evans and Schmalensee (2005b), and Chakravorti (2010) for an overview of this debate.

[5]After forcing MasterCard to cut its cross-border interchange fees around zero, the European Commission is currently investigating Visa's fees. (see MasterCard case: COMP/34.579, and Visa cases: COMP/29.373, COMP/39.398.). The U.S. congress issued "Credit Card Fair Fee Act of 2008"and is pondering further measures in two new pieces of legislation: the "Credit Card Interchange Fees Act of 2009" and the "Expedited CARD Reform for Consumers Act of 2009". Price cap regulations on various fees (mainly interchange fees) to protect merchants from excess charges, have already been applied in Australia, Canada, Switzerland, Mexico, Chile and Denmark. Other countries (e.g., UK and Sweden) have regulated card networks' rules and agreements, or outlaw them, aiming to reduce merchant fees.

the implications of an important asymmetry between cardholders and merchants: cardholders make two choices (usage and membership) whereas merchants make only one (membership). This asymmetry is shown to be the ultimate root of the above distortion. We argue that this source of inefficiency is particularly important since it follows from a structural feature of the industry rather than from differences in idiosyncratic attributes between the two groups of users.

We develop a model of card pricing in which banks are allowed to charge two-part tariffs to their customers, consumers and merchants, in return for providing access to the payment network (e.g., the Visa network). A two-part tariff consists of a fixed fee, which is paid when a customer signs up or renews a contract with his bank, and a marginal (per-transaction) fee, which is paid every time the card is used. To distinguish membership decisions (and fees) from usage decisions (and fees), the model allows for random shocks on consumers' card usage benefits/costs which are realized on a purchase by purchase basis. These shocks depend on, for instance, the distance to the closest ATM, the size and type of the transaction, the availability of foreign currency and so on. Cardholders learn their "convenience benefit" of paying by card (rather than by other means) at the point of sale and then decide on card usage. At the membership stage consumers pay for the option of being able to pay by card in the future and for the intrinsic benefits associated to membership (e.g., social prestige, insurance). Merchants instead compare their average net benefit from card payments (e.g., easy accounting, safer transactions), with the average merchant fee to decide on card acceptance (membership). We account for the fact that different end-users have different tastes for transactions and membership by allowing for four degrees of end-user heterogeneity and thereby elastic final demands.

We firstly consider a card network's pricing incentives in the simplest setting with one card issuer (i.e., the cardholders' bank) and many perfectly competitive acquirers (i.e., merchants' banks). Alternatively one could think of a "three-party scheme" such as American Express, in which a single company directly contracts with cardholders and merchants (the two setups are formally equivalent). In this context we show that profit maximization always implies allocating an inefficiently high amount of the total per-transaction price to merchants. Intuitively, financing card usage perks through higher charges on merchants not only increases issuance of *new* cards but also fosters usage of *existing* cards. Membership fees make banks the residual claimants of the change in the option value of holding a card. That is, issuing banks fully internalize the *incremental* cardholder surplus due to this additional inefficient usage. The related welfare loss due to lower merchant surplus, is partly internalized through average merchant fees, but merchants' marginal losses cannot be internalized, since merchants cannot affect card usage after becoming an affiliated merchant of the scheme. It follows that starting from the first best price structure shifting price burden towards merchants always benefits cardholders and is therefore profitable for the card scheme, in spite of discouraging card acceptance. Such skewed card prices result in *overusage* of payment cards in the sense that an inefficiently high fraction of sales are settled by cards at affiliated merchants.

It is important to note that the argument does not require membership fees to be positive. Zero or even negative membership fees can be simply explained by fierce competition for cardholders.[6] What matters is that banks internalize the incremental surplus that stems from better card usage terms. Thus even in those cases where card membership is subsidized, by negotiating better usage terms with the networks, issuers are able to reduce the amount of subsidies required to reach their target membership level.

This result is robust to the introduction of imperfect competition for cardholders and for merchants, as well as with network competition (e.g., Visa versus Mastercard) provided that merchants accept different card brands (i.e., multi-home).[7] Finally, we argue that the result is also robust to many other factors affecting final demands, e.g., strategic card acceptance as a quality investment and/or to steal business from a rival.

The literature on payment cards has already identified several potential sources of inefficiencies. Wright (2001) and Schmalensee (2002) firstly emphasize the platform's role in "balancing" the demand of payment services by consumers and merchants. The sign and magnitude of the distortion is shown to depend on asymmetries in costs, in demand elasticities and in the relative intensity of competition for end users on the two sides of the market. Rochet and Tirole (2002) recognized a further source of distortion by formalizing the idea that competing merchants may accept cost increasing cards to steal customers from their rivals. The greater the competitive edge guaranteed by card acceptance, the more likely is that card networks exploit the lower merchant "resistance" (to price increases) by setting an inefficiently high merchant fee. Rochet and Tirole (2003), Guthrie and Wright (2003), Armstrong (2006) reach similar conclusions by studying the effect of competition among payment networks. If merchants accept the cards of multiple card networks (i.e., multi-home), competition increases the distortion even further, as networks try to woo cardholders back from their rivals by lowering their prices. Networks can then charge merchants the monopoly price to provide access to their exclusive turf of cardholders.

Despite shedding a great deal of light on the workings of the industry, these analyses deliver no straightforward normative implications when both consumer and merchant demands are assumed to be elastic (see Wright (2004)). The relationship between socially and privately optimal card prices depend on *quantitative* considerations, e.g., surplus measures hinging on cost and preference attributes. Assessing distortions require a significant amount of information and in principle interventions could go in either direction. We contribute to this literature by identifying another source of distortion arising from the fact that cardholders make two choices (usage and membership) whereas merchants make only one (membership). To illus-

[6] Negative fixed fees (e.g., introductory "bonuses" for card membership) can be rationalized, for instance, when a card is either sold as a part of a bundle (i.e., checking account and other payment services) or provide complementary services (e.g., credit lines) which constitute alternative sources of profits (interest fees).

[7] According to the Nilsen Report the acceptance network of the two major players almost perfectly overlaps with 29 million shops for Visa and 28.5 million for Mastercard. Even if consumers adopt more than one payment card, Rysman (2007) finds empirical evidence that they mostly use only one of them.

trate this distortion we distinguish membership and usage decisions and fees. In contrast to the previous work, the *sign* of the distortion does not depend on fundamental cost and/or preference attributes. Only its *magnitude* does. This result is derived without imposing additional restrictions. In fact, we show that our formulation obtains the baseline characterizations of the equilibrium prices of the above contributions as special cases (that is, imposing simultaneously one or more restrictions on prices and/or information structure). Thus we argue that there is a sense in which such result is stronger

Regarding policy concerns, our model unambiguously predicts that cap regulations on interchange fees can improve social welfare. However, we do not find any support for widely used issuer cost-based cap regulation. In line with the literature, we indeed find that the socially optimal fee structure reflects two considerations: relative demand elasticities (marginal users) and relative net surpluses (average users). We furthermore show that regulating the interchange fee is not enough to achieve the full efficiency in the industry. The interchange fee could affects only the allocation of the total user price between consumers and merchants whereas the first best efficiency requires also a lower total price level due to positive externalities between the two sides.

Section **??** presents our framework. In Section **??** we derive the profit maximizing card fees and merchant fees, as well as illustrate the distortion on the card price structure and describe the optimal regulation. Section **??** characterizes the-first best total price level. In Section **??**, we formally extend our results to imperfect issuer competition and discuss the extension to imperfect acquirer competition. We relate our framework and findings to the existing literature in Section **??**. Section **??** concludes with some policy implications.

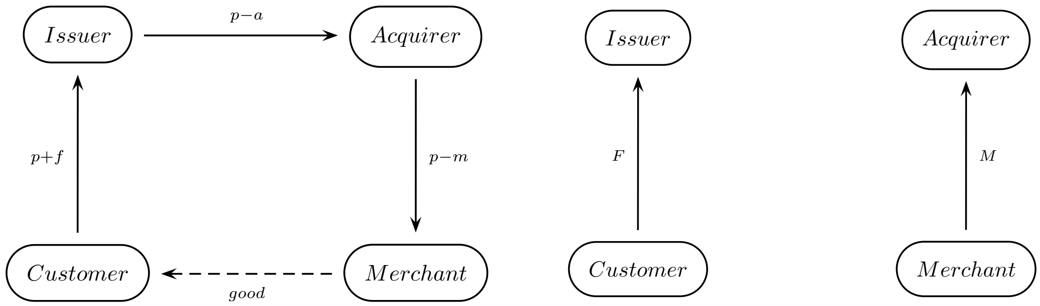
2 A Model of the Payment Card Industry

A payment card association (e.g. Visa) provides card payment services to card users (cardholders and merchants) through issuers (cardholders' banks) and acquirers (merchants' banks). We assume that issuers have market power whereas the acquiring side of the market is competitive.[8] This assumption is meant to fit the payment card industry.[9] Section 7 extends our main argument to the symmetric case of a monopoly issuer and a monopoly acquirer. We also assume that there is a price coherence, i.e. the price of a good is the same regardless it is paid by cash or by card.[10]

[8]By modeling issuers and acquirers as different agents we also implicitly assume that banks are specialized either in issuing or in acquiring.

[9]See Evans and Schmalensee (1999), Rochet and Tirole (2002, 2005), and the EC's report (2007) for a discussion of the cause and the extent of market power in the payment card industry.

[10]Card schemes mostly prohibit merchants from surcharging card payments (the so called No-Surcharge Rule). Although surcharging is allowed in the UK, in Sweden, and in the Netherlands, it is uncommon in practice, probably due to transaction costs of price discrimination among buyers using different forms of payment.



(a) Card Transaction (b) Card Subscription/Acceptance

Fig. 1: Card Payments

Consumption Surplus We consider a continuum (mass one) of consumers and a continuum (mass one) of locally monopoly merchants.[11] Consumers are willing to purchase one unit of a good from each merchant and the unit value from consumption is assumed to be the same across merchants. Let $v > 0$ denote the value of a good purchased by cash, that is the consumption value net of all cash-related transaction costs. A consumer gets $v - p$ from purchasing a unit good by cash at price p and the seller gets p from this purchase.[12]

Card Usage Surplus Consumers (or buyers) get an additional payoff of $b_B - f$ when they pay by card rather than cash. Let b_B denote the net per-transaction benefit[13] and f denote the transaction fee to be paid to the issuer. Similarly, merchants (or sellers) get an additional payoff of $b_S - m$ when paid by card where b_S denotes the net per-transaction benefit of a card payment[14] and m denotes the merchant discount (or fee) to be paid to the acquirer. Note that we do not impose any sign restriction, potentially allowing for negative benefits, i.e. distaste for card transactions, and negative fees, e.g. reward schemes like cash-back bonuses or frequent-flyer miles. For each card transaction, the issuer (respectively the acquirer) incurs cost c_I (respectively c_A). Let c denote the total cost of a card transaction, so $c = c_I + c_A$. The card association requires the acquirer to pay an interchange fee a per transaction to the issuer. The issuer's (respectively the acquirer's) transaction cost is thus $c_I - a$ $(c_A + a)$. Figure 1a summarizes the flow of fees triggered by a card transaction of amount p.

Card Membership Surplus Buyers and sellers are also subject to membership, i.e. transaction insensitive, fees (denoted respectively by F and M) and benefits (denoted respectively by B_B and B_S) upon joining the card association (Figure 1b).[15] To simplify the notation, we

[11] In the extensions, we discuss the robustness of our results to merchant competition.
[12] Retailing costs play no role in the analysis and are wlog set to zero.
[13] Such as foregoing the transaction costs of withdrawing cash from an ATM or converting foreign currency.
[14] Such as convenience benefits from lower cash holdings, faster payments, easy accounting, saved trips to the bank etc.
[15] E.g. cardholders enjoy security of not carrying big amounts of cash, membership privileges (such as access to VIP), travel insurance, ATM services (such as account balance sheets, money transfers, etc.), social prestige (club effects); merchants benefit from safe transactions.

assume that the fixed costs of issuing an extra card and acquiring an extra merchant are zero.

In what follows we assume that consumers and merchants are heterogeneous both in their *usage* and *fixed* benefits from card payments. Specifically, benefits b_B, b_S, B_B and B_S are assumed to be distributed on some compact interval with smooth atomless cumulative distribution functions satisfying the Increasing Hazard Rate Property (IHRP).[16] Benefits b_B and b_S are i.i.d. across transactions.

Timing

Stage i: The payment card association (alternatively a regulator) sets the interchange fee, a.

Stage ii: After observing a, each issuer sets its card fees and each acquirer sets its merchant fees.

Stage iii: Merchants and consumers observe their membership benefits B_S and B_B and decide simultaneously whether to accept and hold the payment card, respectively, and which bank to patronize.

Stage iv: Merchants set retail prices. Merchants and consumers realize their transaction benefits b_S and b_B respectively. Consumers decide whether to purchase. Finally cardholders decide whether to pay by card or cash.

Consumers and merchants maximize their expected payoff. We assume that the card association sets the interchange fee to maximize the sum of the profits earned by its issuers and acquirers. This assumption aims to represent real objectives of for-profit card associations. [17] In principle for-profit card organizations could charge their members non-linear membership fees, and thus could internalize any incremental increase in their members' profits through fixed transfers. In the analysis, this means to define the profit of the association as the total fees collected from members, which could be proxied by the total profits of its member banks allowing the association to charge fixed fees as well as transaction fees to its members. We are looking for a Subgame Perfect Nash Equilibrium (SPNE).

Consumption Surplus versus Card Usage Surplus Let buyer benefit b_B be distributed over interval $\left[\underline{b_B}, \overline{b_B}\right]$ with cumulative distribution function $G(b_B)$ and probability density function $g(b_B)$. Similarly, let seller benefit b_S be distributed over some interval $\left[\underline{b_S}, \overline{b_S}\right]$ with CDF $K(b_S)$ and PDF $k(b_S)$. To simplify the benchmark analysis, we make the following assumption:

$$A1 : v \geq c - \underline{b_B} - \underline{b_S} + \frac{1 - G(\underline{b_B})}{g(\underline{b_B})}.$$

[16]The IHRP leads to *log-concavity of demand functions* (for cardholding, for card usage, and for card acceptance), which is sufficient for the second-order conditions of the optimization problems we solve.

[17]Visa and MasterCard were used to be non-profit organizations, but since 2003 Visa and since 2006 MasterCard are for-profit organizations and their shares are jointly owned by their member banks. See the EC's report (2007) and MasterCard decision (2007).

Guthrie and Wright (2003, Appendix B) show that under A1 monopoly merchants set $p = v$ regardless of whether they accept card payments or not.[18] The assumption guarantees that v is sufficiently high so that merchants never find it profitable to exclude cash users, by setting a price higher than v. In other words A1 rules out the case where merchants try to extract some of the surplus associated with card transactions (e.g., rewards) by increasing retail prices. After solving the benchmark model, we show that relaxing A1 reinforces our results.

2.1 Preliminary Observations

By A1, all merchants set $p = v$ and therefore all consumers purchase a unit good from each merchant. If a merchant accepts cards, a proportion, α_B, of its transactions (to be determined in equilibrium) is settled by card. The net payoff of type B_S merchant from accepting cards is:[19]

$$B_S - M + E\left[b_S - m\right]\alpha_B, \tag{1}$$

which is the sum of the membership and expected transaction surpluses when merchant fees are (M, m). The number of merchants that join the payment card network is thus:

$$\alpha_S \equiv \Pr(B_S - M + E\left[b_S - m\right]\alpha_B \geq 0).$$

Note that α_S depends only on the average merchant benefit and fee, which are defined respectively as:

$$\tilde{b}_S \equiv E\left[b_S\right] + \frac{B_S}{\alpha_B} \qquad \text{and} \qquad \tilde{m} \equiv m + \frac{M}{\alpha_B},$$

and thus $\alpha_S = \Pr(\tilde{b}_S \geq \tilde{m})$. There is therefore one degree of freedom in acquirers' pricing policy. Any $\hat{\alpha_S}$, resulting from some fees $(\hat{M}, \hat{m})$ can also be implemented through a simple linear pricing scheme: $M = 0$, $m = \tilde{m}(\hat{m}, \hat{M})$.[20,21] This observation is due to the fact that the card acceptance decision is sunk when seller learns its benefit b_S, and therefore the card acceptance demand cannot be affected by the realization of b_S. Only the *average* benefit known *before* the acceptance decision matters. For a given α_B, our framework is thus equivalent to a setup where merchants are heterogeneous in their *average* benefits prior to their card acceptance decisions.

Without loss of generality in what follows we focus on a model where $B_S = M = 0$, and merchants are heterogeneous in their average benefit (denoted by b_S) which they know before

[18] Note that this is different than the no-surcharge rule which prevents a merchant from price discriminating between card users and cash users.

[19] Since we assume that b_S is i.i.d across transactions, even when a merchant realizes its benefit $b_S < m$ for one transaction, the merchant does not stop its card membership because it does not know the exact value of its benefit for future transactions.

[20] This would still be the case if we assumed some market power on the acquiring side.

[21] In fact, if merchants were risk-averse it would then be a dominant strategy to charge only for usage since payments are due only if a transaction occurs.

card acceptance decisions. We assume that b_S is continuously distributed on some interval $[\underline{b_S}, \overline{b_S}]$ with CDF $K(b_S)$, PDF $k(b_S)$ and increasing hazard rate $k/(1-K)$. Regarding

Crucially we cannot set $B_B = F = 0$ wlog since buyers make *two* decisions (card membership and usage) at different information sets. Cardholding depends on the average benefit and card fee, whereas card usage depends on the transaction benefit and fee. We assume that B_B is continuously distributed on some interval $[\underline{B_B}, \overline{B_B}]$ with CDF $H(B_B)$, PDF $h(B_B)$. Recall that benefits b_B, b_S and B_B are independently distributed and allowed to be negative.

3 Benchmark Analysis

□ **Usage Decisions**

Cardholders pay by card if and only if their transaction benefit exceeds the usage fee. Buyers' demand for card usage is

$$D_B(f) \equiv \Pr(b_B \geq f) = 1 - G(f),$$

which is the proportion of cardholders paying by card at transaction price f.

□ **Membership Decisions**

Merchant of type b_S accepts cards whenever $b_S \geq m$.[22] The proportion of merchants who accept payment cards is

$$D_S(m) \equiv Pr(b_S \geq m) = 1 - K(m).$$

Define respectively buyers' and sellers' average surpluses from card usage as $v_B(f) \equiv E[b_B - f \mid b_B \geq f]$ and $v_S(m) \equiv E[b_S - m \mid b_S \geq m]$. The expected value of being able to pay by card at a point of sale, which we call as *the option value* of the card, is denoted by Φ_B and equal to

$$\Phi_B(f, m) \equiv v_B(f) D_B(f) D_S(m),$$

where $D_B(f)D_S(m)$ is the volume of card transactions at card fee f and merchant fee m. Note that the option value increases with the expected card usage at affiliated merchants, D_B, and with merchant participation, D_S. Type B_B gets a card if and only if the total benefits from cardholding, i.e. the sum of fixed benefit B_B and the option value of the card, exceed the fixed card fee:

$$B_B + \Phi_B(f, m) \geq F.$$

The number of cardholders, which is denoted by Q, is then

$$\begin{aligned} Q(F - \Phi_B(f, m)) &= \Pr[B_B + \Phi_B(f, m) \geq F] \\ &= 1 - H(F - \Phi_B(f, m)), \end{aligned}$$

which is a continuous and differentiable function of card fees (F, f) and merchant discount m.

□ **Behavior of the Issuer and Acquirers**

[22] Card acceptance is not affected by card usage/membership, i.e., there is no externality imposed by consumers on merchant participation. We could restore this externality by allowing for fixed merchant fees, since the card usage demand then affects the average merchant fee, without changing our conclusions (see the discussion in the previous section).

Taking the IF as given, perfectly competitive acquirers simply pass-through their costs charging $m^*(a) = a + c_A$ per transaction.

For a given IF, the issuer sets card fees by maximizing its profit, which is the sum of the card transaction profits and fixed card fees collected from cardholders:

$$\max_{F,f} \left[(f + a - c_I)D_B(f)D_S(m) + F\right] Q(F - \Phi_B(f,m)). \tag{2}$$

The usual optimality conditions bring the equilibrium fees:

$$f^*(a) = c_I - a, \qquad F^*(a) = \frac{1 - H(F^*(a) - \Phi_B(a))}{h(F^*(a) - \Phi_B(a))}.$$ [23]

The fixed fee is characterized by a Lerner formula. The issuer introduces a monopoly markup on its fixed costs (for simplicity here set to zero), inefficiently excluding some consumers from the market. The usage fee is set at the marginal cost of issuing since the issuer could internalize incremental card usage surpluses through the fixed fee, even though it cannot extract all buyer surplus.

Privately and Socially Optimal Interchange Fees

Taking into account the equilibrium reactions (card fees and merchant fees) of banks to a given IF level, we proceed to define three critical levels of IF: the buyers-optimal IF, a^B, which maximizes the buyer surplus (gross of fixed fees), the sellers-optimal IF, a^S, maximizing the seller surplus, and a^V, which maximizes the volume of card transactions:

$$\begin{aligned}
a^B &\equiv \arg\max_a BS(a) = v_B(f^*)D_B(f^*)D_S(m^*)Q(F^*, f^*, m^*) + \int_{F^*-\Phi_B(f^*,m^*)}^{\overline{B}_B} xh(x)dx \\
a^S &\equiv \arg\max_a SS(a) = v_S(m^*)D_B(f^*)D_S(m^*)Q(F^*, f^*, m^*) \\
a^V &\equiv \arg\max_a V(a) = D_B(f^*)D_S(m^*)Q(F^*, f^*, m^*)
\end{aligned}$$

Lemma 1 *Interchange fees* $\left(a^B, a^S, a^V\right)$ *exist uniquely and satisfy* $a^S < a^V < a^B$.

Proof. *Appendix A.1.*

This lemma highlights the *tension* between buyers' and sellers' interests over the level of IF. An increase in the interchange fee has three effects. On one hand, it induces a higher merchant fee and thus lowers the number of shops where cards are welcome. On the other hand, it results in a lower card usage fee, and thus induces cardholders to settle more transactions by card at each affiliated store. Furthermore, a higher interchange fee changes buyers' expected surplus from card transactions (the option value of the card, Φ_B), and thus changes the net price of the card, $F - \Phi_B$. A unit increase in Φ_B increases the equilibrium fixed fee less than

[23] To simplify the expressions, we write $\Phi_B(a)$ instead of $\Phi_B(c_I - a, c_A + a)$.

one, and therefore lowers the net price of the card resulting in a higher number of cardholders. Given that the number of cardholders, and thus total utility of buyers from cardholding, is increasing in the option value of the card, the IF maximizing the option value also maximizes the buyer surplus (gross of fixed fees). We show that the interchange fee maximizing the option value is higher than the volume maximizing IF which is higher than the sellers-optimal IF, since the average buyer surplus from card transactions, v_B, is decreasing in card usage fee f, so increasing in IF, whereas the average seller surplus, v_S, is decreasing in merchant fee m, so in IF (due to the IHRP). Going above the volume-maximizing IF increases the buyer surplus (gross of fixed fees) at the expense of the seller surplus.

Indeed a^B is also the interchange fee maximizing the net buyer surplus (net of fixed fees). To see this suppose that this is not true and there exists another interchange fee, say $\overline{a} \neq a^B$, which maximizes the buyer surplus. Suppose that we change interchange fee incrementally starting from $\overline{a}$ towards a^B. This change would increase the gross consumer surplus (ignoring fixed fee). This incremental increase would be *partly* captured by the issuer through a fixed fee. In other words, buyers would be better off by this change. This contradicts with the assumption at the beginning that $\overline{a}$ maximizes the net consumer surplus.

□ **Equilibrium Fees**

Given the equilibrium reactions of banks,

$$f^*(a) = c_I - a \qquad \text{and} \qquad m^*(a) = c_A + a,$$

fixing the IF is formally equivalent to allocating the total cost of a transaction between the two sides of the market. Perfect competition on the acquiring side of the market implies that the association sets the IF that maximizes the issuer's profits subject to the equilibrium prices set by banks:

$$\max_{F,f,m} FQ(F - \Phi_B(f,m)) \qquad \textit{st.:} \quad \textit{i.}\ f + m = c \ \ \textit{ii.}\ F = \frac{1 - H\left(F - \Phi_B(f,m)\right)}{h\left(F - \Phi_B(f,m)\right)}. \tag{3}$$

The issuer's profits are clearly increasing in the option value of the card Φ_B. To see this it suffices to apply the envelope theorem to the objective function. Hence the privately optimal allocation is the allocation that maximizes the option value. It is such that the impact of a small variation of f on the option value is equal to the impact of a small variation of m.

From Lemma 1 we know that a^B maximizes Φ_B. We thus conclude that the privately optimal IF is equal to a_B, that is $a^* = a^B$.

□ **Optimal Regulation**

In this section we consider the problem of a regulator seeking to maximize the total surplus in the economy through an appropriate choice of a. Such problem can also be stated as a price allocation problem similar to (??):

$$\max_{F,f,m} \{[v_B(f) + v_S(m)] D_B(f) D_S(m) + E[B_B \mid B_B \geq F - \Phi_B(f,m)]\} Q(F - \Phi_B), \qquad (4)$$

subject to the same set of constraints.

The above formulation makes clear that the only difference between the regulator's problem and the association's problem is in the *allocation* of the total price c across the two sides of the market. As we shall see in the next section, full efficiency indeed requires a total price different than c.

To highlight the discrepancy between public and private incentives we shall restate problem (??) in terms of the indifferent cardholder, $\tilde{B}_B$:

$$\max_{\tilde{B}_B,f,m} (v_B(f) D_B(f) D_S(m) + \tilde{B}_B) Q(\tilde{B}_B) \quad \textit{st.:} \quad \textit{i.} \quad \textit{and} \quad \textit{ii.} \qquad (3')$$

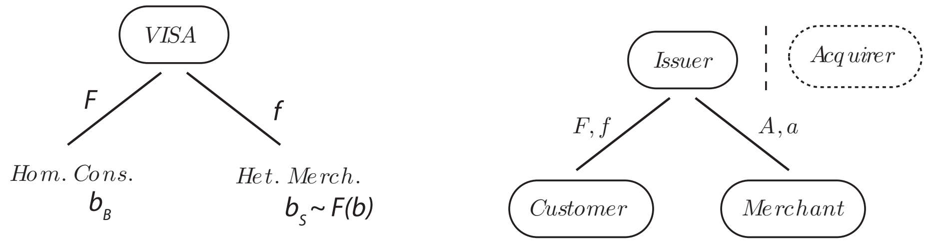
Comparing (??) with the association's objective, (??), highlights the two sources of welfare losses induced by the association's pricing policy. First, the association distorts the allocation of costs between card users and merchants, neglecting the impact of a marginal variation of the interchange fee on the merchant surplus. Starting from any IF between a_S and a_B, a marginal increase of a raises the buyer surplus (gross of fixed fees) at the expense of the merchant surplus (see Lemma 1). Through fixed card fees, the issuer, and thus the association, internalizes all *incremental* card usage surpluses of buyers due to this increase in IF. On the other hand, the lack of term $v_S D_B D_S Q$ in the association's objective reflects the seller surplus that the association fails to account for.

The second source of distortion is due to the monopoly markup of the issuer. Through setting a, the association determines indirectly equilibrium fixed fee, F^*, and the option value of the card, Φ_B^*, which together determines the net price of the card, $F^* - \Phi_B^*$, and thus the equilibrium number of cardholders. Increasing membership on one side implies more surplus on both sides of the market since the number of interactions (i.e. card transactions) increases. For an additional cardholder, the association accounts for the fixed benefit of the marginal cardholder, $\tilde{B}_B$, whereas the social planner internalizes the fixed benefit of the average cardholder, $E\left[B_B \mid B_B \geq \tilde{B}_B\right]$. The fact that the association fails to capture the impact of an extra cardholder on the net benefit of an average cardholder, $E\left[B_B - \tilde{B}_B \mid B_B \geq \tilde{B}_B\right]$, results in an additional discrepancy between private and public interests.

We are now in a position to compare the regulator's choice with the choice of the association:

Proposition 1 *The privately optimal IF is higher than the socially optimal IF. Hence, in equilibrium, cardholders pay too little and merchants pay too much per transaction.*

Proof. *Appendix A.2.*



(a) 3-Party Scheme

(b) w/non linear merchant fees

Fig. 2

For the special case where consumers get no fixed benefits from cardholding, $\underline{B}_B = \overline{B}_B = 0$, there is an intuitive characterization of the efficient fee:[24]

$$\frac{f}{m} = \frac{\eta_B}{\eta_S} \div \frac{v_B}{v_S},$$

where $\eta_B = -\frac{fD'_B}{D_B}$ is the elasticity of buyers' card usage and $\eta_S = -\frac{mD'_S}{D_S}$ is the elasticity of sellers' card acceptance demand. The socially optimal allocation of the total price $f + m = c$ is achieved when relative user prices are equal to the ratio of the relative demand elasticities and the relative average surpluses of buyers and sellers.

So far we have discussed how the discrepancy between private and public interests (respectively (**??**) and (**??**)) affects economic efficiency through the association's pricing policy. In the rest of this section we shall focus on the determinants of such discrepancy.

Observe that interchange fees, which constitute revenues for the issuing side, let the issuer extract (some of) the merchant surplus. It follows that by controlling the association's choice of a, the issuer acts effectively as a single platform owner. In fact one could think of the issuer as directly charging merchants for card services since competitive acquirers simply pass-through interchange fees to merchants. The benchmark framework is therefore formally equivalent to a monopoly platform pricing both sides to maximize its profits (fig. **??**a).[25] The only asymmetry between the two sides of the market is that usage choices (i.e., the choice of the payment instrument) are delegated to consumers. This structural feature of the payment card market is the ultimate foundation of the allocational distortion of proposition 1. The intuition is as follows. Increasing the IF beyond the socially optimal level not only attracts new members through a higher option value but also fosters card usage among *existing* members. The incremental buyer surplus due to this extra, inefficient, usage can be extracted at the membership stage through higher fixed fees, while keeping the consumer participation fixed

[24] An analogous property holds for the optimal access charge between backbone operators or between telecom operators where the access charge allocates the total cost between two groups of users (consumers and web sites in backbone networks, call receivers and call senders in telecommunication networks) (See Laffont et al. (2003)). This condition is first documented by Rochet and Tirole (2003).

[25] Indeed this observation extends our findings to so called proprietary (or 3-party) schemes such as AMEX.

(i.e., keeping the average card fee fixed). The same is not true on the merchant side of the market. The association cannot fully internalize incremental losses in the merchant surplus due to this increase of the IF. As shown in section **??**, considering non-linear charges on merchants (such as non-linear interchange fees or non-linear merchant fees (Fig **??**b)) would not affect the result. Changing the marginal price, m, while keeping the average merchant price constant does not have any impact on the volume of card transactions. This is because merchants make only one decision, that is, whether to become a member of the card association and the number of merchants accepting cards depends uniquely on the average merchant price and benefit from card acceptance. Hence one of the two pricing instruments is redundant on the merchant side.

Rochet and Tirole (2003, 2006b) derive the optimal pricing structure for a monopoly platform setting linear prices to both sides. As opposed to theirs, our equilibrium fees do not maximize the total volume of transactions. We thus cannot conclude that in equilibrium there is *over-provision* of card services simply by noticing that the socially optimal IF is different (in our framework smaller) than the privately optimal one. Improving buyers' usage incentives through a higher IF (inducing for instance reward schemes and cash back bonuses) does not necessarily lead to a higher total volume of transactions, since some merchants abandon the platform in response to higher merchant fees. In our model there is *over-usage* in the sense that, in equilibrium, the proportion of buyers who choose to pay by card at an affiliated merchant is always inefficiently high.

Lemma 1 proves that the volume maximizing IF is above the interchange fee maximizing the seller surplus and below the interchange fee maximizing the buyer surplus (net of fixed fees). The comparison between the volume maximizing IF and the socially optimal IF depends on the trade-off between how much the average seller surplus decreases when we increase the interchange fee above the volume maximizing level versus how much the average buyer surplus increases by this change. In other words, the comparison of the volume maximizing IF with the socially optimal IF depends on distributions of buyer and seller benefits, which depend on hardly observable preferences.

4 Efficient Fees

In this section we characterize the first best (Lindahl) fees. Though it is hard to implement the first best fees in practice, they are informative about the nature of the externalities in this market.

Consider the problem of a public monopoly running the industry in order to maximize the total welfare, which is the total profits of issuer and acquirer banks plus the surplus of buyers and sellers:

$$\max_{F,f,m} W \equiv \{[f + m - c + v_B(f) + v_S(m)]\, D_B(f) D_S(m) + E\left[B_B \mid B_B \geq F - \Phi_B\right]\} Q(F - \Phi_B). \tag{5}$$

Proposition 2 *The first best total price (per transaction) is lower than the total cost of a transaction and equal to* $c - v_B(f^{FB})$*. The socially optimal allocation of such a price is achieved when*

$$v_B(f^{FB}) = v_S(m^{FB}),$$

that is, when the average buyer surplus is equal to the average seller surplus.

Proof. *Appendix B.1.*

Intuitively, each type of user is charged a price equal to the cost of a transaction minus a discount reflecting its positive externality on the other segment of the industry. An extra card user (respectively merchant) attracts an additional merchant (respectively card user) which generates average surplus v_S (respectively v_B).[26] At the optimum, the two externalities must be equalized, the total price is thus

$$f^{FB} + m^{FB} = c - v_S(m^{FB}) = c - v_B(f^{FB}) < c$$

.

A Ramsey planner solves (**??**) subject to an additional constraint: $\Pi_A, \Pi_I \geq 0$, where Π_A and Π_I denote respectively acquirers' and the issuer's profits. The rationale for the latter comes from the problem of a regulator who can control end-user prices but cannot or does not want to run and/or subsidize operations, and therefore has to leave enough profits to keep the industry attractive for private investors. Using an argument analogous to that employed in the proof of proposition 2 it is possible to show that the second best total price is higher than the first best, but still *lower* than the cost of a transaction. Below-cost usage fees can be financed through fixed charges on the consumer side, and thus do not necessarily trigger budget imbalances.

5 Competing Issuers

In this section, we modify our benchmark setup by allowing for imperfect competition between two issuers, denoted by I_1 and I_2, which provide differentiated payment card services within the same card scheme and charge their customers two-part tariff card fees. Consumers have preferences both for payments made by card instead of other means and for the issuer itself, i.e., brand preferences. Brand preferences are due to, for instance, quantity discounts, e.g., family accounts, physical distance to a branch, or consumers' switching costs deriving from the level of informational and transaction costs of changing some banking products, e.g., current accounts.

Card i refers to the payment card issued by I_i, for $i = 1, 2$. We denote the net price of card i by t_i, which is defined as the difference between its fixed fee and the option value of

[26] Such pricing rule was independently found by Weyl (2009).

holding card i: $t_i \equiv F_i - \Phi_B(f_i, m)$. The demand for holding card i is denoted by $Q(t_i, t_j)$ (or Q_i), for $i \neq j$, $i = 1, 2$. We make the following assumptions on Q_i:

$$A2 \; : \; \frac{\partial Q_i}{\partial t_i} < 0 \qquad A3 : \frac{\partial Q_i}{\partial t_j} > 0 \qquad A4 : \left|\frac{\partial Q_i}{\partial t_i}\right| > \frac{\partial Q_i}{\partial t_j}$$

$$A5 \; : \; \frac{\partial^2 InQ_i}{\partial t_i^2} < 0 \qquad A6 : \left|\frac{\partial^2 InQ_i}{\partial t_i^2}\right| > \left|\frac{\partial^2 InQ_i}{\partial t_i \partial t_j}\right|$$

A2 states that the demand for holding a card is decreasing in its net price. A3 ensures the substitutability between the card services provided by different issuers so that the demand for holding card i is increasing in the net price of card j. By A4, we furthermore assume that this substitution is imperfect, and thus the own price effect is greater than the cross price effect. By assuming that Q_i is log-concave in net price t_i, A5 ensures the concavity of the optimization problems. A6 states that own price effect on the slope of the log-demand is higher than the cross price effect.

In Appendix C.1, we provide examples of classic demand functions for differentiated products, such as Dixit (1979), Singh and Vives (1984), Shubik and Levitan (1980), which satisfy all of our assumptions.

□ Behavior of the Issuers and Acquirers

Perfectly competitive acquirers set $m^*(a) = c_A + a$. Taking the IF and card j's fees given, I_i's problem is to set (F_i, f_i) to maximize its profit:

$$\max_{F_i, f_i} \left[(f_i + a - c_I) D_B(f_i) D_S(m) + F_i\right] Q\left(F_i - \Phi_B(f_i, m), F_j - \Phi_B(f_j, m)\right).$$

Like in the benchmark case, both issuers set $f_i^*(a) = c_I - a$ in order to maximize the option value of their card. The option value is therefore equal to $\Phi_B(c_I - a, c_A + a)$ (or compactly $\Phi_B(a)$) regardless of the identity of the issuer. The best reply fixed fee of I_i to its rival's fixed fee, F_j, is implicitly given by

$$\epsilon_i(F_i^*, F_j; a) = 1,$$ [27]

where $\epsilon_i \equiv -F_i \frac{\partial Q_i / \partial F_i}{Q_i}$ refers to the elasticity of I_i's demand with respect to its fixed fee. Assumption A5 (log-concavity of the demand) guarantees that ϵ_i is increasing in F_i, and thus that F_i^* is well-defined. Whenever ϵ_i is greater (respectively less) than 1, I_i has incentives to

[27] Observe that the optimality condition is indeed given by the Lerner formula:

$$markup_i = \frac{1}{\epsilon_i},$$

where the markup of each duopolist issuer is equal to 1 since there is no fixed cost in our setup. If instead each issuer paid fixed cost C_I per card, the solution to I_i's problem would be

$$markup_i \equiv \frac{F_i^* - C_I}{F_i^*} = \frac{1}{\epsilon_i},$$

whereas we simply assume that $C_I = 0$, so we have $markup_i = 1$.

lower (respectively raise) its fixed fee until $\epsilon_i = 1$. An equilibrium of issuer competition is any pair (F_i^*, F_j^*) such that $\epsilon_i(F_i^*, F_j^*; a) = \epsilon_j(F_j^*, F_i^*; a) = 1$.

□ Privately and Socially Optimal Interchange Fees

The association's problem is to set the IF maximizing the sum of the issuers' profits $\Pi_1^* + \Pi_2^*$ where each issuer earns

$$\Pi_i^* = F_i^* Q\left(F_i^* - \Phi_B(a), F_j^* - \Phi_B(a)\right),$$

given that $\epsilon_i(F_i^*, F_j^*; a) = \epsilon_j(F_j^*, F_i^*; a) = 1$. Our claim is that the association sets $a^* = a^B$ maximizing the option value of the card, $\Phi_B(a)$. We prove the claim by showing that equilibrium profits increase with Φ_B. Applying the Envelope Theorem to the issuer profits, we derive

$$\frac{\partial \Pi_i^*}{\partial \Phi_B} = F_i^* \left[-\frac{\partial Q_i}{\partial t_i} - \frac{\partial Q_i}{\partial t_j} + \frac{\partial Q_i}{t_j} \frac{\partial F_j^*}{\partial \Phi_B} \right],$$

which helps us to identify two types of effects on I_i's profit of a marginal increase in the option value:

1. Demand Effect: The direct effect of the net card prices on Q_i, which is composed of own and cross demand effects. The *own demand effect* (the first term in brackets) is positive because the demand decreases in the net price of the card (A2) increasing in the option value of the card. The *cross demand effect* (the second term in brackets) is negative because the demand increases in the net price of the rival's card (A3) decreasing in the option value. The overall demand effect is positive since the positive own demand effect dominates the negative cross demand effect (A4).

2. Strategic Effect: The last term in brackets accounts for the impact of a change in the option value on the rival's pricing policy.

Lemma 2 *Under $A2 - A6$, both equilibrium fees are increasing in Φ_B.*

Proof. *Appendix C.2.*

By Lemma 2 we find that the strategic effect is positive: Increasing the option value of the card softens price competition. As a result the profit of each issuer increases in the option value, Φ_B. A straightforward consequence is that:

Corollary 1 *Under $A2 - A6$, the issuers' incentives over the interchange fee are aligned.*

To maximize the sum of the issuers' profits, the association sets $a^* = a^B$, which maximizes cardholders' surplus from card transactions.

We are now left to compare the profit maximizing interchange fee with the welfare maximizing fee. The regulator would set an IF, a^r, to maximize the total welfare anticipating banks' pricing behavior in equilibrium:

$$\max_a \left\{ \begin{array}{l} [v_B(c_I - a) + v_S(c_A + a)]\, D_B(c_I - a) D_S(c_A + a)\, [Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)] \\ + E\,[B_B \mid B_B \geq F_1^* - \Phi_B]\, Q(F_1^*, F_2^*, a) + E\,[B_B \mid B_B \geq F_2^* - \Phi_B]\, Q(F_2^*, F_1^*, a) \end{array} \right\}.$$

The solution to the usual optimality conditions characterizes a^r. By comparing the association's IF with the regulator's IF, we get our main result:

Proposition 3 *If A2-A6 hold then the privately optimal IF is higher than the socially optimal IF. Hence, in equilibrium, cardholders pay too little and merchants pay too much per transaction.*

Proof. *Appendix C.3.*

Once we acknowledge the fact that the issuers' incentives are aligned to those of cardholders, the logic behind proposition **??** is analogous to that of the previous section: The association sets the buyers' optimal IF, whereas the regulator would set a lower IF since it could internalize buyers' as well as sellers' surpluses and the seller surplus is decreasing in IF.

Finally, we explain the role of issuers' competition. Competition is effective in reducing membership fees and thus in reducing (even eliminating) the distortion due to the issuer market power (see the discussion before proposition 1). This can easily be established contrasting the equilibrium outcome with the outcome that would arise if the issuers were jointly owned. This observation coupled with the marginal cost pricing, $f_i = f_j = c_I - a$, implies that the total surplus is always higher under competition no matter what IF prevails in equilibrium. However, issuer competition fails to reduce the distortion originating from the inefficient allocation of transaction costs between consumers and merchants.

Discussion: Imperfect Acquirer Competition We conjecture that our results are robust to the introduction of market power on the acquiring side of the market. To see this, for instance consider the symmetric case of a monopoly issuer and a monopoly acquirer. Optimal pricing on the acquirer's side involves a markup ($m^* > c_A + a$) which is characterized by a standard inverse elasticity rule over merchants' demand for card services. Such markup allows the acquiring bank to extract some of the surplus that merchants derive from card payments. Assuming that the association maximizes the total profits of its member banks, this creates a countervailing incentive to lower interchange charges. Such conflict between the issuer's and the acquirer's interests is due to the conflict between sellers' and buyers' interests. In particular close to the issuer's optimal IF, the acquirer's profits decrease in a.

Following a reduction of the IF below the buyers' optimal level the acquirer internalizes only *a part* of the incremental surpluses that accrue to the merchant side of the market, since it has to leave some rent to heterogeneous merchants. Therefore, the privately optimal IF would likely be higher than the socially optimal one which takes fully into account both merchants' and consumers' incremental surpluses from card transactions.

6 Comparisons with the Literature

6.1 Cardholding vs Card Usage Decisions and Fees

In Rochet and Tirole (2002, 2003), consumers are fully informed about their benefits before their cardholding decision, so considering linear or non-linear card fees, per-transaction and/or fixed benefits would give the same results in their analysis. In their model consumers get a card if and only if they plan to use it for all future transactions. Such timing implicitly assumes that consumers make only one decision, whether to hold the card or not, by comparing their average benefit with the average card fee. In our formulation consumers get the card in order to secure the *option* of paying by card in the future whenever this happens to be convenient for a particular transaction. Such formulation has mainly two advantages. Firstly it is able to rationalize frequent use of cash by many cardholders. Secondly and most importantly it distinguishes card membership from card usage decisions (and fees), by assuming that these two decisions are made at different information sets. Such timing was firstly introduced by Guthrie and Wright (2003). Their paper however restricts the analysis to linear fees and is therefore formally equivalent to Rochet and Tirole's (2002, 2003), which corresponds to our formulation under the restriction that $F = 0$.

6.2 Homogeneous Merchants

If $\underline{b}_S = \bar{b}_S$ *all* merchants accept cards if and only if $b_S \geq m$. Perfectly competitive acquirers set $m^*(a) = c_A + a$. In this case, Baxter (1983) shows that setting an IF equal to $b_S - c_A$, which we call Baxter's IF, implements efficient card usage if issuers are also perfectly competitive setting $f^*(a) = c_I + a$. Intuitively, the first best could be implemented through the usage fee that induces buyers to internalize the externality they impose to the rest of the economy while paying by card, i.e., $f^{fb} = c - b_S$. His analysis is restricted to be normative since perfectly competitive banks have no preferences over the level of IF. Going beyond Baxter, we assume imperfectly competitive issuers, and thus the privately optimal IF is well-defined in our analysis.

When issuers have market power, card fees are linear and fixed benefits from cardholding are zero (or the same for everyone), Guthrie and Wright (2003, Proposition 2) show that the socially optimal IF results in under-provision of card payment services. The reason is the following. The regulator would like to set an IF above Baxter's IF to induce the optimal card usage in the presence of an issuer markup. But then merchants would not participate (as $m > b_S$). At the second best, the regulator sets Baxter's IF, which is also the privately optimal IF and results in under-provision of card services. Next proposition shows that allowing for fixed card fees prevents inefficient provision of card services by eliminating issuer markups. A formal proof of the proposition is available upon request.

Proposition 4 *When merchants are homogeneous, the privately and the socially optimal IFs* *always coincide. Furthermore,*

i. If imperfectly competitive issuers can charge only linear usage fees, there is under-provision of card payment services.

ii. If membership (fixed) fees are also available, there is socially optimal provision of card payment services.

Intuitively, since issuers could internalize incremental card usage surpluses of buyers through fixed fees, they set the usage fees at their transaction costs, $c_I + a$. Baxter's IF then implements the first best transaction volume.

6.3 Strategic Card Acceptance

By assuming monopoly merchants, we abstract away from business stealing effects of accepting payment cards. Rochet and Tirole (2002) are the first who analyze such effects in a model where merchants accept the card to attract customers from rival merchants who do not accept the card. For a given retail price, card acceptance increases the quality of merchant services associated with the option to pay by card. Consumers are ready to pay higher retail prices for the improved quality as long as they observe the quality.[28] Rochet and Tirole show that when merchants are competing à la Hotelling, they internalize the average surplus of consumers from card usage, $v_B(f)$, so merchants accept cards if and only if $b_S + v_B(f) \geq m$. In other words, merchants pay $m - b_S$ to accept cards since they could recoup v_B through charging higher retail prices for their improved quality of services.

It is important to note that we do not need merchant competition to make this argument. A monopoly merchant would also be willing to incur a cost per card transaction to offer a better quality of services to its customers (who value the option of paying by card), since it could then internalize *some*[29] of the average card usage surplus of buyers by charging higher retail prices.

We make assumption A1 to rule out card acceptance aiming to improve quality. Recall that A1 ensures a high enough consumption value by cash, v, so that merchants who accept cards do not want to exclude cash users by setting a price higher than v. In our setup, merchants accept cards only to enjoy convenience benefits from card payments, and thus they accept cards if and only if $b_S \geq m$. Once we relax A1, a merchant accepting cards might be willing to charge a price higher than v (exclude cash users, sell only to card users) since by increasing its price, it could internalize some of the buyer surplus from card usage. Anticipating this extra revenue from card users, a merchant might accept cost increasing cards. For instance,

[28] The authors assume that only a proportion, α, of consumers observe which store accepts cards before choosing a store to shop. Here, we consider simply their extreme case of $\alpha = 1$, which is sufficient to make our point.

[29] Unlike Hotelling competition, total demand is decreasing in retail price. This is why the monopolist merchant could internalize *some of* the (not *all*) average card usage surplus.

consider simply the case of homogeneous merchants and suppose that a merchant accepting cards prefers to set $p^* > v$, i.e., it gains more from setting $p = p^*$ than $p = v$. If the merchant sets p^*, only card users buy its product and the merchant gets[30]

$$\Pi_S^* = (p^* + b_S - m) D_B(f + p^* - v),$$

If the merchant sets $p = v$, all consumers buy its product and the merchant gets

$$\Pi_S = v + (b_S - m) D_B(f)$$

We assume that $\Pi_S^* > \Pi_S$, and thus the merchant prefers to set $p^* > v$. Since $D_B(f) > D_B(f + p^* - v)$ for $p^* > v$, our assumption ($\Pi_S^* > \Pi_S$) implies also that

$$V(p^*, f) \equiv p^* - \frac{v}{D_B(f + p^* - v)} > 0$$

where $V(p^*, f)$ is a positive function referring to the merchant's extra surplus from increasing its quality (so its retail price) through accepting cards. Putting it differently $V(p^*, f)$ refers to some of the average card usage surplus of buyers. The IHRP implies that p^* is decreasing in f (see the previous footnote). Using this together with the monotonicity of $D_B(.)$, we get that $V(p^*, f)$ is decreasing in f.

If the merchant does not accept cards, it gets $\Pi_S = v$. A merchant thus accepts cards whenever

$$\Pi_S^* = (p^* + b_S - m) D_B(f + p^* - v) \geq v \qquad or$$
$$b_S + V(p^*, f) \geq m$$

Anticipating extra surplus $V(p^*, f)$ from card users, the merchant is willing to pay more than its convenience benefit to be able to accept cards, i.e., it resists less to an increase in m when it expects to get a higher surplus after accepting cards. Furthermore, the reduction in its resistance, $V(p^*, f)$, decreases in card usage fee f, so increases in the IF. When the association raises the IF, the merchant fee increases, which decreases the participation of merchants. Conversely, the increase in the IF decreases the card usage fee increasing $V(p^*, f)$. This in turn increases merchant participation. The latter effect does not exist in our original setup

[30] A monopolist merchant accepting cards sets its price by

$$\max_p (p + b_S - m) D_B(f + p - v) \; st.: \; p \geq v$$

The solution to the unconstrained problem is implicitly given by

$$p^* = m - b_S + -\frac{D_B(f + p^* - v)}{D_B'(f + p^* - v)}$$

The merchant's optimal price is p^* if it satisfies the constraint, i.e., $p^* > v$. Otherwise the merchant sets its price equal to v. We suppose here that the constraint is not binding in equilibrium.

under A1. Hence, merchants would resist less to an increase in the IF if we relaxed A1, in which case the privately optimal IF would be even higher than what we found. Hence, relaxing A1 would reinforce our results: cardholders would pay even less and merchants would pay even more. The same conclusions would hold if we allowed business stealing effects by introducing competition among merchants, since such a modification in our setup would again weaken the resistance of merchants to an increase in IF [see Rochet and Tirole (2002, 2006a)]. For the case of heterogeneous merchants, we could make a similar argument for the marginal merchant: relaxing A1 would make the marginal merchant less resistant to an increase in the IF, and thus the association sets a higher IF.

7 Policy Implications and Concluding Remarks

This paper focuses on a payment card association (e.g. Visa or MasterCard) and analyzes welfare implications of the interchange fee paid by the merchant's bank to the cardholder's bank for every card transaction. We develop a framework taking into account the fact that consumers decide first whether to hold a card, i.e. become a *member* of the card association, and then whether to *use* their card at a particular point of sale, whereas merchants decide *only* whether to become a member. We first illustrate the conflict of interests between merchants and consumers: each side would want the other to bear more of the cost of a card transaction. We show that a card association that seeks to maximize profits of its member banks solves this conflict inefficiently in favor of the cardholders at the expense of too high merchant fees. The association sets an inefficiently high interchange fee since higher interchange fees induce lower card usage fees (or rewards) and thus encourage card usage making the payment card more valuable at the membership stage. The incremental surpluses of cardholders can be extracted though higher annual fees. However, it is not possible to capture incremental card usage surpluses of merchants since they cannot affect card usage once they become a member of the card association. In our model there is *over-usage* in the sense that, in equilibrium, the proportion of consumers who choose to pay by card at an affiliated merchant is always inefficiently high. This inefficiency result is valid also in a third-party card scheme, like AMEX, since in this case the card scheme sets directly inefficiently low card usage fees at the expense of too high merchant fees.

Our results show that there is a scope for improving the social welfare through setting maximum levels (caps) on interchange fees. However, we have not found any reason to apply the widely used cost-based regulation, which sets a cap on the IF that reflects the issuers' (weighted or simple) average cost (such as transaction authorization, processing, fraud prevention etc.). In line with the existing literature we obtain a simple characterization of the socially optimal IF which depends on the relative demand elasticities and the relative average surpluses of consumers and merchants, i.e., end-user preferences.

We also show that regulating the IF is not enough to achieve full efficiency in the payment card industry, since efficiency requires each user fee be discounted by the positive externality of that user on the rest of the industry and one tool (IF) is not enough to achieve efficient usage on both sides. Intuitively, we suggest that if a card scheme charged its member banks fixed membership fees as well as transaction fees[31], the platform could induce both consumers and merchants to internalize their externalities, and thus improve efficiency. We leave the characterization of an efficient IF mechanism for future research.

The qualitative results are robust to imperfect issuer competition and to many factors affecting final demands, such as elastic cardholding and strategic card acceptance to attract consumers. We conjecture that our results could also be extended to the case of imperfect acquirer competition.

Interchange fees are relatively high in the United States, where membership fees are not often used. This observation might look contradictory to the mechanism explained above.[32] Our analysis of competing issuers could indeed explain this. We argue that even when the issuer competition lowers the equilibrium fixed fees to zero (or even to negative), our results would remain valid, i.e., the association sets a too high interchange fee. The fact that the issuers are able to capture incremental card usage surpluses through fixed fees would result in an interchange fee maximizing buyers' card usage surplus regardless of the level of equilibrium fixed fees. To see this consider an incremental change in IF towards the buyers' optimal interchange fee. Since such a change would increase the expected card usage surplus, the issuers could charge higher fixed fees while keeping the number of cardholders fixed. They are thus better off by this change. This is true even when the equilibrium fixed fees are negative, since in this case the issuers lose less from subsidizing membership.

Our assumption that the price of the good is the same whether it is paid by card or by cash is critical to the analysis. We support this assumption by the fact that surcharging card payments is not common in practice. In most countries surcharging card payments is prohibited by payment schemes, especially by international schemes. According to the EC's Retail Banking Sector Inquiry, there is no widespread surcharging even when it is allowed. For instance, prohibition on surcharging was lifted in the UK since 1989, in Sweden since 1995, in Netherlands since 1997, in Switzerland since 1997. In Australia it is lifted since 2003. According the Federal Reserve Bank of Australia, in 2007 only 14% of very large merchants, only 5% of very small merchants surcharge. The theory of interchange fees (Gans and King) shows that when surcharging a payment card is costless, it is optimal for merchants to surcharge card payments, in which case the interchange fee would become neutral for the volume of card transactions. However, in practice surcharging many different types of cards, debit vs credit, Amex vs MasterCard vs Visa, etc., for different types of products would be very costly. This could be one explanation of why we do not observe widespread surcharging even when merchants are allowed to surcharge. Next step would be to develop a theoretical model where

[31]In this case, different transaction fees could be set to issuers versus acquirers.

[32]We would like to thank Mark Armstrong for raising this question.

some merchants in equilibrium would prefer not to surcharge. There might be many factors affecting the attitudes of merchants towards surcharging card payments, such as the scale of the merchant, the type of the product sold at that merchant, the competition between merchants, etc.

It is important to note that in any two-sided market where one group of users make membership and usage decisions at different information sets, whereas the other group makes only membership decisions, and transactions between the end-users are observed by the platform, our results suggest that the platform sets potentially inefficient user prices by favoring too much the side which decides on usage after membership decisions are made. Interesting examples are online search engines, such as google or yahoo!, providing a platform for consumers and content providers. Consumers decide whether to do an online search, which is usually provided for free by the platform, and then whether to click a title appearing on the search result, which is again free for consumers. However, content providers decide only on membership, i.e., whether to put their adds on the web and they pay per transaction, which is click in this example. Our analysis predicts that a search engine subsidizes too much consumers at the expense of content providers paying too high per click prices.

Our setup does not incorporate the implications of competition among card schemes or other payment methods. However, as long as consumers use only one type of card and merchants subscribe to more than one card platform, competing card schemes would like to attract consumers (competitive bottlenecks) (see Rochet and Tirole (2003), Guthrie and Wright (2003)), and thus favor more the consumer surplus than the merchant surplus. In this case, the upward distortion of equilibrium IFs would be greater than the case of a monopoly card scheme. For instance, consider competition between two homogeneous payment schemes. Assume that merchants are multi-homing, i.e., accept cards of both schemes, (which is indeed mostly the case in practice) and that consumers are single-homing, i.e., hold only one type of card. The competitive bottleneck argument of the literature would suggest that the schemes would compete for consumers. They would set the interchange fee which maximizes the total consumer surplus. Hence, our results would be robust to this extension. If we instead assumed that consumers also multi-home, i.e., hold both types of cards, the competition between the schemes would then be to convince cardholders to use their cards. This means more private incentives to raise interchange fees to induce lower card usage fees (e.g., rewards). Hence, our results would be reinforced in this extension, i.e., the privately optimal interchange fees would be further above the socially optimal level. A thorough analysis is needed to see which side is going to use/accept one type of card in equilibrium.

A marginal decrease from the card association's IF is found to be socially desirable, however, we are unable to determine how much the IF should be decreased by. Too stringent price caps could be worse than no cap regulation. Our setup inherits all the practical limitations of setting socially optimal prices that depend on hardly observable characteristics of supply and demand. At this point we provide a theoretical framework which is hopefully rich enough to be used by an empirical analysis to characterize the socially optimal interchange fee.

Appendix

A Benchmark Analysis

A.1 Proof of Lemma 1

We first show that $v'_B(f) < 0$ and $v'_S(m) < 0$ under the Increasing Hazard Rate Property (thereafter IHRP) of distribution functions respectively $G(f)$ and $K(m)$. Consider first $v_B(f)$. Using $D_B(f) = 1 - G(f)$ and integrating by parts, we get

$$v_B(f) \equiv \frac{\int_f^{\bar{b}_B} D_B(x)dx}{D_B(f)}. \tag{6}$$

Define $Y(f) \equiv \int_f^{\bar{b}_B} D_B(x)dx$. Notice that the IHRP is equivalent to say $D'_B/D_B \equiv Y''/Y'$ is a decreasing function. Given that Y''/Y' is decreasing and that $Y(\bar{b}_B) = 0$ and $Y(f)$ is strictly monotonic by definition, we have that Y'/Y is decreasing due to Bagnoli and Bergstrom (1989, Lemma 1).[33] Using (**??**), decreasing Y'/Y is equivalent to $v'_B(f) < 0$. Similarly, we can establish that $v'_S(m) < 0$. Since $v'_B \equiv -\frac{v_B D'_B + D_B}{D_B}$ and $v'_S \equiv -\frac{v_S D'_S + D_S}{D_S}$, inequalities $v'_B(f) < 0$ and $v'_S(m) < 0$ imply respectively that $v_B D'_B + D_B > 0$ and $v_S D'_S + D_S > 0$.

Define functional I as

$$I \equiv -\frac{\left(\text{HR}^{-1}\right)'}{1 - \left(\text{HR}^{-1}\right)'}$$

where HR^{-1} is the inverse of hazard rate, $\frac{1-H}{h}$, and thus decreasing by the IHRP. Note that $0 < I(\cdot) < 1$.

Given the best responses of the issuer ($f^*(a) = c_I - a$ and $F^*(a) = \frac{1-H(F^*(a)-\Phi_B(a))}{h(F^*(a)-\Phi_B(a))}$) and acquirers ($m^*(a) = c_A + a$), we now characterize interchange fees a^B, a^V, a^S which respectively maximize the buyer surplus (gross of fixed fees), the total transaction volume, and the seller surplus subject to the subgame perfection.

Existence and uniqueness of a^B:

First notice that the IHRP and $v'_B < 0$ imply respectively the log-concavity of D_S and $v_B D_B$, and thus Φ_B is log-concave. An important property of continuous log-concave functions is that

[33] The Generalized Mean Value Theorem of calculus ensures, for every x, the existence of a $\xi \in (x, \bar{b}_B)$ such that

$$\frac{Y'(x) - Y'(\bar{b}_B)}{Y(x) - Y(\bar{b}_B)} = \frac{Y''(\xi)}{Y'(\xi)}$$

If Y''/Y' is decreasing, for any $x < \xi$, it should then be the case that

$$\frac{Y'(x) - Y'(\bar{b}_B)}{Y(x) - Y(\bar{b}_B)} < \frac{Y''(x)}{Y'(x)}$$

Since Y is monotone and $Y(\bar{b}_B) = 0$, it must then be that $Y'(x)Y(x) < 0$ whenever $x < \bar{b}_B$. Multiplying both sides of the above inequality by $Y'(x)Y(x)$ gives $Y''(x)Y(x) < (Y')^2 - Y'(\bar{b}_B)Y'(x)$ and thus that $Y''(x)Y(x) - (Y')^2 < 0$, which is equivalent to Y'/Y decreasing.

the first order condition is both necessary and sufficient to have a local (and thus a global) maximum.[34]
Hence there exists a unique IF which maximizes the option value.
The buyers-optimal interchange fee, a^B, is a solution to:

$$\max_a BS(a) = \left[\int_{F^*(a)-\Phi_B(a)}^{\overline{B}_B} xh(x)dx + \Phi_B(a)Q(F^*(a) - \Phi_B(a)) \right],$$

$$where \ \Phi_B(a) = v_B(c_I - a)D_B(c_I - a)D_S(c_A + a).$$

This problem has an interior solution only if $f = c_I - a \leq \overline{b_B}$, which is equivalent to $a \geq c_I - \overline{b_B}$, because otherwise no one pays by card. The quasi-demand D_B is maximized and equal to 1 when $f = c_I - a \leq \underline{b_B}$, that is $a \geq c_I - \underline{b_B}$, and there is no gain from increasing a above $c_I - \underline{b_B}$. Without loss of generality, we thus restrict the domain of a to be $\left[c_I - \overline{b_B}, c_I - \underline{b_B}\right]$. By the Weierstrass Theorem, there exists a maximum of the continuous function $BS(a)$ on the compact interval $\left[c_I - \overline{b_B}, c_I - \underline{b_B}\right]$. By differentiating $F^*(a)$, we get

$$F^{*\prime}(a) = I\left(F^*(a) - \Phi_B(a)\right)\Phi_B'(a),$$

which implies that $[F^* - \Phi_B]' = -(1-I)\Phi_B'$. We therefore conclude that the IF which maximizes $\Phi_B(a)$, minimizes $[F^*(a) - \Phi_B(a)]$, and therefore maximizes $\int_{F^*(a)-\Phi_B(a)}^{\overline{B}_B} xh(x)dx$. Since cardholding demand $Q \equiv 1 - H$ is log-concave by the IHRP, the IF which maximizes $\Phi_B(a)$ also maximizes $\Phi_B(a)Q(F^*(a) - \Phi_B(a))$. We thus conclude that a^B is unique and equal to $\text{argmax}_a \Phi_B(a)$.
The *existence and uniqueness of* a^S*:* The sellers-optimal IF, a^S, is a solution to

$$\max_a SS(a) = v_S(c_A + a)D_S(c_A + a)D_B(c_I - a)Q(F^*(a) - \Phi_B(a))$$

The Weierstrass Theorem guarantees the existence of a^S on $\left[c_I - \overline{b_B}, c_I - \underline{b_B}\right]$. Log-concavity of functions $v_S D_S$ (by $v_S' < 0$), D_B (by the IHRP), and Q (by the IHRP), implies that a^S is uniquely determined by the first-order optimality condition:

$$SS'(a^S) = -D_S(D_B + v_S D_B')Q + (1-I)\Phi_B' h v_S D_S D_B = 0 \tag{7}$$

The existence and uniqueness of a^V*:* The volume-maximizing IF, a^V, is a solution to

$$\max_a V(a) = D_B(c_I - a)D_S(c_A + a)Q(F^*(a) - \Phi_B(a))$$

The Weierstrass Theorem guarantees the existence of a^V on $\left[c_I - \overline{b_B}, c_I - \underline{b_B}\right]$. Since quasi-

[34]To see this notice that by definition a function $f(x)$ is log-concave if $log(f(x))$ is concave, which is equivalent to f'/f decreasing or $f^{''}f - (f^{'})^2 < 0$. It follows that if f is log-concave, at any critical point the SOC must then be verified, i.e., for any x^* such that $f'(x^*) = 0$, we have $f^{''}(x^*) < 0$

demands D_B, D_S and cardholding demand Q are log-concave (implied by the IHRP), the volume of transactions $D_B D_S Q$ is log-concave. The unique interchange fee, a^V, is then implicitly given by the first-order optimality condition:

$$V'(a^V) = \left(-D'_B D_S + D'_S D_B\right) Q + (1-I)\Phi'_B h D_B D_S = 0 \tag{8}$$

Now, our claim is $a^B > a^V$. By using the definition of a^B, i.e., $\Phi'_B(a^B) = D_B D_S + v_B D_B D'_S = 0$, we derive the volume of transactions at a^B:

$$V'(a^B) = -\frac{QD_S}{v_B}\left(v_B D'_B + D_B\right).$$

We have $V'(a^B) < 0$ since $v_B D'_B + D_B > 0$ from $v'_B < 0$. Given that function $V(a)$ is concave (by the IHRP), condition (**??**) implies then that $a^B > a^V$.
Symmetrically, by using the IHRP and $v'_S < 0$, it can be shown that $a^S < a^V$. Hence, we prove that $a^S < a^V < a^B$.

A.2 Proof of Proposition 1

By definition a^B maximizes the surplus of buyers (gross of fixed fees) and a^S maximizes the surplus of sellers. Lemma 1 shows the existence and the uniqueness of a^B and a^S, and that $a^B > a^S$. By the revealed preference argument an interchange fee maximizing the sum of user surpluses, $BS(a) + SS(a)$, necessarily lies in (a^S, a^B).

B Efficient Fees

B.1 Proof of Proposition 2

We decompose the planner's problem of setting transaction prices f, m into a price allocation and a total price setting problem. We have already characterized in Proposition 1 the optimal allocation of total price $f + m = p = c$. We are thus left to generalize the optimal allocation of any total price p and characterize then the optimal p. Let $f(p)$ and $m(p)$ denote the respective fees which implement the optimal allocation of p between buyers and sellers.
The social planner first solves

$$\max_f \left[p - c + v_B(f) + v_S(p-f)\right] D_B(f) D_S(p-f) Q(F - \Phi_B(f, p-f)) + \int_{F-\Phi_B(f,p-f)}^{\overline{B}_B} x h(x) dx,$$

which characterizes implicitly $f^{FB}(p)$ and $m^{FB}(p) = p - f(p)$ as follows:

$$\begin{gathered}\left[(p-c)(D'_B D_S - D_B D'_S) - v_B D_B D'_S + v_S D'_B D_S\right] Q - \\ (p - c + v_B + v_S) D_B D_S Q' \partial_f \Phi_B + (F - \Phi_B)\, h\,(F - \Phi_B)\, \partial_f \Phi_B = 0\end{gathered} \tag{9}$$

where $Q' < 0$ and $\partial_f \Phi_B$ denotes the derivative of the option value, $\Phi_B(f, p - f)$, with respect to f..

The planner next determines the socially optimal total price by

$$\max_p \left[p - c + v_B(f(p)) + v_S(p - f(p))\right] D_B(f(p)) D_S(p - f(p)) Q(F - \Phi_B) + \int_{F-\Phi_B}^{\overline{B}_B} x h(x) dx,$$

Using $[v_i D_i]' = -D_i$ and the Envelope Theorem, we get the first order condition:

$$(p - c + v_B) D_B D_S' Q - (p - c + v_B + v_S) D_B D_S Q' \partial_p \Phi_B + (F - \Phi_B) h(F - \Phi_B) \partial_p \Phi_B = 0 \quad (10)$$

Finally the socially optimal membership fee $F^{FB}(p, f(p))$ is characterized by:

$$(p - c + v_B + v_S) D_B D_S Q' = (F - \Phi_B) h(F - \Phi_B) \quad (11)$$

Plugging (**??**) into **??** gives:

$$(p - c + v_B) D_B D_S' Q = 0 \quad (12)$$

which is verified if and only if $p^{FB} = c - v_B(f^{FB})$. Plugging (**??**) and p^{FB} into condition (**??**) we get:

$$(v_S - v_B) D_B' D_S Q = 0 \quad (13)$$

which implies that $v_S(p^{FB} - f^{FB}) = v_B(f^{FB})$.

C Competing Issuers

C.1 Examples of Demand Functions

The following examples of demand functions for differentiated products satisfy assumptions A2-A6.

(1) Linear symmetric demands of form, for $i = 1, 2, i \neq j$,

$$q_i = \frac{1}{1+\sigma} - \frac{1}{1-\sigma^2} p_i + \frac{\sigma}{1-\sigma^2} p_j$$

where q refers to demand, p refers to price, and σ measures the level of substitution between the firms (here, for imperfectly competitive issuers we have $\sigma \in (0, 1)$). These demands are driven from maximizing the following quasi-linear and quadratic utility function

$$U(q_i, q_j) = q_i + q_j - \sigma q_i q_j - \frac{1}{2}\left(q_i^2 + q_j^2\right)$$

subject to the budget balance condition, namely

$$p_i q_i + p_j q_j \leq I$$

(2) Dixit (1979)'s and Singh and Vives (1984)'s linear demand specification, for $i = 1, 2, i \neq j$,

$$q_i = a - bp_i + cp_j$$

where $a = \frac{\alpha(\beta-\gamma)}{\beta^2-\gamma^2}$, $b = \frac{\beta}{\beta^2-\gamma^2}$, $c = \frac{\gamma}{\beta^2-\gamma^2}$, and the substitution parameter is $\varphi = \frac{\gamma^2}{\beta^2}$, under the assumptions that $\beta > 0$, $\beta^2 > \gamma^2$, and $\varphi \in (0, 1)$ for imperfect substitutes.
(3) Shubik and Levitan (1980)'s demand functions of form, for $i = 1, 2, i \neq j$,

$$q_i = \frac{1}{2}\left[v - p_i(1+\mu) + \frac{\mu}{2}p_j\right]$$

where $v > 0$, μ is the substitution parameter and $\mu \in (0, \infty)$ for imperfect substitutes.
Special case: Hotelling Demand, for $i = 1, 2, i \neq j$,

$$q_i = \frac{p_j - p_i}{2t} + \frac{1}{2}$$

satisfies the assumptions except for A4 and A6 since the own price effect is equal to the cross price effect, that is

$$\left|\frac{\partial q_i}{\partial p_i}\right| = \frac{\partial q_i}{\partial p_j} \qquad \left|\frac{\partial^2 Inq_i}{\partial p_i^2}\right| = \frac{\partial^2 Inq_i}{\partial p_i \partial p_j}$$

which imply that the equilibrium fixed fees are independent of the option value, and thus independent of the IF. In this case, the issuers would not have any preferences over IF. Hence, the privately optimal IF is not well defined.

C.2 Proof of Lemma 2

Consider the FOC of I_i's problem:

$$FOC_i : Q\left(F_i - \Phi_B, F_j - \Phi_B\right) + F_i \frac{\partial Q_i}{\partial F_i} = 0$$

Solving FOC_i and FOC_j together gives us the equilibrium fees as functions of the option value, i.e., $F_i^*(\Phi_B)$ and $F_j^*(\Phi_B)$. The second-order condition holds by A5:

$$SOC_i : 2\frac{\partial Q_i}{\partial F_i} + F_i^* \frac{\partial^2 Q_i}{\partial F_i^2} < 0$$

The solution of the issuers' problems gives us the symmetric equilibrium $F_i^* = F_j^*$. By taking the total derivative of the first-order conditions, we derive

$$\frac{\partial F_j^*}{\partial \Phi_B} = \frac{\partial F_i^*}{\partial \Phi_B} = 1 - \frac{\partial Q_i / \partial Fi}{SOC_i + \frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j}}$$

If $\frac{\partial^2 InQ_i}{\partial t_i \partial t_j} < 0$, we have

$$\frac{(\partial^2 Q_i/\partial F_i \partial F_j)Q_i - (\partial Q_i/\partial F_i)(\partial Q_i/\partial F_j)}{Q_i^2} < 0$$

so that

$$\frac{\partial Q_i}{\partial F_j} - \frac{Q_i}{\partial Q_i/\partial Fi}\frac{\partial^2 Q_i}{\partial F_i \partial F_j} < 0$$

From FOC_i we have, $F_i^* = -\frac{Q_i}{\partial Q_i/\partial Fi}$, so we get

$$\frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j} < 0$$

Moreover, the log-concavity of Q_i (A5) implies that $SOC_i < \partial Q_i/\partial Fi$. Thus, we get $0 < \frac{\partial F_i^*}{\partial \Phi_B} < 1$.
If $\frac{\partial^2 InQ_i}{\partial t_i \partial t_j} > 0$, we have

$$\frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j} > 0.$$

Assumption A6 becomes $-\frac{\partial^2 InQ_i}{\partial t_i^2} > \frac{\partial^2 InQ_i}{\partial t_i \partial t_j}$, which implies that

$$-\left[\frac{\partial Q_i}{\partial F_i} + F_i^* \frac{\partial^2 Q_i}{\partial F_i^2}\right] > \frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j},$$

Using SOC_i, we get

$$\partial Q_i/\partial Fi > SOC_i + \frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j},$$

proving that $0 < \frac{\partial F_i^*}{\partial \Phi_B} < 1$.

C.3 Proof of Proposition 3

Following the lines of our benchmark analysis, we first define three important IF levels: the buyers-optimal IF, the sellers-optimal IF, and the volume maximizing IF,which we denote respectively by a^{Bc}, a^{Sc}, and a^{Vc}, where superscript c refers to issuer competition:

$$\begin{aligned}
a^{Bc} &\equiv \arg\max_a \left\{ \begin{array}{c} v_B(c_I - a)D_B(c_I - a)D_S(c_A + a)\left[Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)\right] + \\ \int_{F_1^* - \Phi_B(a)}^{\overline{B}_B} xh(x)dx + \int_{F_2^* - \Phi_B(a)}^{\overline{B}_B} xh(x)dx \end{array} \right\} \\
a^{Sc} &\equiv \arg\max_a v_S(c_A + a)D_B(c_I - a)D_S(c_A + a)\left[Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)\right] \\
a^{Vc} &\equiv \arg\max_a D_B(c_I - a)D_S(c_A + a)\left[Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)\right]
\end{aligned}$$

From Lemma 2, we have $0 < \frac{\partial F_i^*}{\partial \Phi_B} = \frac{\partial F_j^*}{\partial \Phi_B} < 1$. Consider now the derivative of $Q(F_i^*, F_j^*, a)$

with respect to a:

$$Q_i'(a) = \left[\frac{\partial Q_i}{\partial F_i} \left(\frac{\partial F_i^*}{\partial \Phi_B} - 1 \right) + \frac{\partial Q_i}{\partial F_j} \left(\frac{\partial F_j^*}{\partial \Phi_B} - 1 \right) \right] \Phi_B'(a)$$

The first term inside the brackets represents the direct effect of the option value on Q_i, through changing the net price of card i, $F_i^* - \Phi_B$, and the second term represents the indirect effect of the option value on Q_i, through changing the net price of card j, $F_j^* - \Phi_B$. Imperfect issuer competition (A3 and A4) implies that the direct effect of the option value on Q_i dominates its indirect effect so that the term inside the brackets is positive. We therefore conclude that when two differentiated issuers are competing with symmetric demands, the demand for holding card i is maximized at $a = a^B$, which is the interchange fee maximizing the option value of the card, $\Phi_B = v_B D_B D_S$.

Following the lines of Lemma 1, we then conclude that the IF maximizing the option value of the card also maximizes the buyer surplus (gross of fixed fees) when the issuers are imperfect competitors, i.e., $a^{Bc} = a^B$. Recall that the association sets $a^* = a^B$ to maximize the issuers' payoffs. Hence, the privately optimal IF coincides with the buyers-optimal IF.

Since the average surplus of buyers and the average surplus of sellers are decreasing in their own usage fees, i.e., $v_B'(f), v_S'(m) < 0$ (see the proof of Lemma 1), we have $a^{Sc} < a^{Vc} < a^{Bc}$. The regulator wants to maximize the sum of buyers' and sellers' surpluses, the socially optimal IF is therefore lower than the privately optimal one.

The formal proof of proposition 4 is available upon request.