Pricing Payment Cards

Online Appendix

Özlem Bedre-Defolie and Emilio Calvano

1 Introduction

This is a supplementary appendix for the paper "Pricing Payment Cards." It provides some additional formal analysis and robustness results anticipated in the article. Section 2 characterizes the first best (Lindahl) fees and the second-best (Ramsey) fees which are informative about the nature of the externalities in the market for plastic payments. Section 3 and 4 deal with alternate market structures: imperfect issuer competition and platform competition, respectively. In the context of competing issuers, we illustrate that zero or even negative membership fees can be explained simply by intense issuer competition for cardholders. Even in those cases where card membership is subsidized, better usage terms for cardholders enable issuers to reduce the amount of subsidies required to reach their target membership level. In the extension of platform competition, we characterize a unique equilibrium with two-part tariffs and, as an important side dish, we solve the previous literature's problem of the multiplicity of equilibrium.[1] Section 5 extends our model to allow for elastic consumption demand. All formal proofs are in the Appendix.

---

[1] This problem is firstly identified by Armstrong (2006). In an independent work, Reisinger (2011) shows the uniqueness of equilibrium with two-part tariffs by modeling the idea that only an exogenous fraction of participants transact with each other. See also Weyl and White (2010) for a general treatment of platform competition. Cantillon and Yin (2010) is the first empirical paper to focus on membership as separate from usage in platform competition.

2 Efficient fees

Consider the problem of a public monopoly running the industry in order to maximize the total welfare:

$$\max_{F,f,m} W \equiv \{[f + m - c + v_B(f) + v_S(m)] D_B(f)D_S(m) + E[B_B \mid B_B \geq F - \Phi_B]\}Q(F - \Phi_B). \quad (1)$$

Proposition 1 *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 the average buyer surplus is equal to the average seller surplus: $v_B(f^{FB}) = v_S(m^{FB})$.*

An extra card user (respectively, merchant) attracts an additional merchant (respectively card user) which generates average surplus v_S (respectively v_B) and the platform earns m^{FB} (respectively f^{FB}) from this additional user. 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:[2]

$$f^{FB} = c - \left[v_S(m^{FB}) + m^{FB}\right], m^{FB} = c - \left[v_B(f^{FB}) + f^{FB}\right].$$

A Ramsey planner would solve (1) subject to the issuer's and acquirers' budget balance constraints: $\Pi_A, \Pi_I \geq 0$, Using an argument analogous to that employed in the proof of Proposition 2 it is possible to show that

Proposition 2 *The second-best total price (per transaction) is higher than the first-best, but lower than the cost of a transaction.*

The Ramsey prices distorts the total price level, but less than the card network. Intuitively, below-cost usage fees can be financed through fixed charges on the consumer side, and thus do not necessarily trigger budget imbalances.

3 Issuer competition

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

---

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

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, that is, brand preferences. Brand preferences are due to, for instance, quantity discounts (family accounts), physical distance to a branch, or consumers' switching costs that derive from the level of informational and transaction costs of changing some banking products (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 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:

$$
\begin{array}{llll}
A2 & : & \dfrac{\partial Q_i}{\partial t_i} < 0 \qquad & A3 : \dfrac{\partial Q_i}{\partial t_j} > 0 \qquad & A4 : \left| \dfrac{\partial Q_i}{\partial t_i} \right| > \dfrac{\partial Q_i}{\partial t_j} \\[3mm]
A5 & : & \dfrac{\partial^2 InQ_i}{\partial t_i^2} < 0 \qquad & A6 : \left| \dfrac{\partial^2 InQ_i}{\partial t_i^2} \right| > \left| \dfrac{\partial^2 InQ_i}{\partial t_i \partial t_j} \right|
\end{array}
$$

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 further 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.

The standard Hotelling model satisfies these assumption. In Appendix C.1., we provide other examples of classic demand functions for differentiated products, such as Dixit (1979), Singh and Vives (1984), Shubik and Levitan (1980), satisfying 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 as given, I_i's problem is to set (F_i, f_i) to maximize its profit:

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

As 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 the Lerner formula where the markup of each duopolist issuer is equal to 1 (since there is no fixed cost):

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

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

Lemma 1 *Suppose that two differentiated issuers are competing in prices. Under A2-A6, in equilibrium the option value of holding one issuer's card is the same as the other's and the fixed fees are increasing in the option value.*

In the monopoly issuer setup, we have seen that one unit increase in the option value increases the equilibrium fixed fee by less than one unit (due to the heterogeneity of consumer benefits). With issuer competition, an increase in the option value increases the equilibrium fixed fees even less than the monopoly case, since the issuers compete away some part of the increase in the fixed fees. As long as issuer competition is not perfect, the equilibrium fixed fees are increasing in the option value of the card.

☐ **Privately and socially optimal interchange fees**

The network'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 network sets $a^* = a^B$, maximizing the option value of the card, $\Phi_B(a)$. We prove the claim by showing that the 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 identifies two 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 inside the 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 inside the brackets accounts for the impact of a change in the option value on the rival's pricing policy. By Lemma 2, the equilibrium fixed fee of the rival is increasing in the option value, and therefore, the strategic effect is positive, that is, increasing the option value of the card softens price competition.

The option value of a card can be seen as the quality of the card. Higher quality increases the price of the product. This is the demand effect. Moreover, a higher quality increases the opportunity cost of a price cut and thereby softens price competition. This is the strategic effect. As a result, the profit of each issuer increases in the option value, Φ_B. The direct consequence is that:

Corollary 1 *Suppose that two differentiated issuers are competing in prices. Under A2-A6, the issuers' incentives over the interchange fee are aligned.*

To maximize the sum of the issuers' profits, the network sets $a^* = a^B$, which maximizes the buyer surplus.

Anticipating the banks' pricing behavior in equilibrium, the regulator would set the IF maximizing the total welfare, denoted by a^r:

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

By comparing the network's IF with the regulator's IF, we get our main result:

Proposition 3 *Suppose that two differentiated issuers are competing in prices. Under A2-A6, the privately optimal IF is higher than the socially optimal IF.*

The logic behind Proposition 3 is analogous to that of the benchmark model: The network sets the buyers' optimal IF, whereas the regulator would set a lower IF since it internalizes buyers' as well as sellers' surpluses and the seller surplus is decreasing in IF.

Competition is effective in reducing membership fees and thus in reducing (even eliminating) the distortion due to the issuer market power (see the discussion prior to Proposition 1 in the original paper). This observation coupled with the marginal cost pricing, $f_i = f_j = c_I - a$, implies that the total welfare is always higher under issuer competition no matter what IF prevails in equilibrium. On the other hand, issuer competition fails to reduce the distortion originating from the inefficient allocation of the total transaction price between buyers and sellers.

Interchange fees are relatively high in the US, where membership fees are not often used. This observation might look contradictory to the mechanism explained above.[3] Our analysis of issuer competition indeed provides an explanation to this: even when issuer competition lowers the equilibrium fixed fees to zero (or makes them negative), the card network sets a too high interchange fee and distorts the price structure. To see this, consider an incremental change in interchange fee from the socially optimal level towards the buyers' optimal level. Since such a change would increase the expected card usage surplus, issuers could raise fixed card 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 issuers lose less from subsidizing membership.

4 Network competition

The impact of introducing network competition (e.g., Visa versus MasterCard) depends on which side of the market users adopt multiple networks (*multi-homing*) rather than a single one (*single-homing*).[4] Intuitively, competition has a bias favoring the single-homing side since steering cus-

---

[3]We would like to thank Mark Armstrong for raising this point.

[4]See Rochet and Tirole (2003), Guthrie and Wright (2003, 2007), Armstrong (2006), Armstrong and Wright (2007).

tomers towards an exclusive relationship lets platforms extract monopoly rents from the multi-homing side users (*competitive bottleneck*).

Casual observation suggests that, at least for the two major card networks, merchants do indeed multi-home. The global card acceptance network of Visa and MasterCard almost perfectly overlaps with 29 million merchants accepting Visa cards and 28.5 million accepting MasterCard in 2009.[5] Multi-homing is encouraged by a widespread practice called "blending", which describes the case where acquirers charge one price for accepting different cards of different networks.[6]

□ Consumers single-home and merchants multi-home

In our setup, assuming that consumers get only one card (single-home) and merchants multi-home, network competition would fail to mitigate the distortion we describe. To see this in the simplest setup, consider two three-party card networks competing (e.g., American Express and Diners Club), denoted by 1 and 2, and assume that consumers single-home, whereas merchants multi-home. Assume furthermore that the networks are homogenous with respect to the card services they provide, but are allowed to be differentiated due to, for example, the brand preferences of users or other differentiated services (such as travel insurance) provided by the card networks. This implies that a consumer receives her membership benefit B_B from holding a card and convenience benefit b_B from paying by card at a point of sale regardless of the card platform she uses. Similarly, a merchant receives the same convenience benefit b_S from being paid by card regardless of the card network it is affiliated to. Similar to the benchmark we focus on per-transaction merchant benefit and per-transaction merchant fee without loss of generality. We make the same assumptions on the distribution functions of end user benefits and keep the same timing except for the following modifications in the first two stages of the game: First, the platforms simultaneously set their user prices, (F_1, f_1) and (F_2, f_2), to consumers and, m_1 and m_2, to merchants. Second, consumers realize their membership benefit and decide whether to hold a card and, if so, choose one card network. Simultaneously, merchants observe their transaction benefit and decide whether to accept the cards of each platform.

The analysis of this setup is very similar to our extension of issuer competition in the four-

---

[5]Nilson Report, June 2009.

[6]See the EC's Sector Inquiry, 2007, pp.16.

party network: issuers correspond to competing networks, so we need to replace the interchange fee by $m_i - c_{Ai}$ for network i and by $m_j - c_{Aj}$ for network j in the analysis of Section 3. Another difference is that competing networks set m_i and m_j competitively, whereas in issuer competition the network can coordinate the merchant fees through setting an interchange fee (strategic effect). However, this difference is not crucial for our qualitative results.

Network i's problem is to set (F_i, f_i, m_i) maximizing its profit:

$$\max_{F_i, f_i, m_i} \left[(f_i + m_i - c) D_B(f_i) D_S(m_i) + F_i \right] Q \left(F_i - \Phi_B(f_i, m_i), F_j - \Phi_B(f_j, m_j) \right),$$

taking its rival's fees, (F_j, f_j, m_j), as given. The solution to the first-order conditions[7] show that the network sets the total user price at its cost of a payment transaction:

$$f_i^* + m_i^* = c, \tag{2}$$

and, similar to the case of issuer competition, allocates the total cost between buyers and sellers to maximize the option value of its cards:

$$m_i^* = \arg\max_{m_i} \Phi_B(c - m_i, m_i) \equiv m_i^B. \tag{3}$$

Each network captures the option value of its cards by a fixed fee given by the Lerner Formula on the network's residual demand:

$$F_i^{BR}(t_j) = \qquad\qquad\qquad -\frac{Q_i}{\frac{\partial Q_i}{\partial t_i}}. \tag{4}$$

---

[7]More precisely, the first-order condition with respect to F_i is

$$Q_i + \left[(f_i + m_i - c) D_B(f_i) D_S(m_i) + F_i \right] \frac{\partial Q_i}{\partial t_i} = 0,$$

and the first-order condition with respect to f_i is

$$\left[D_B(f_i) D_S(m_i) + (f_i + m_i - c_i) D_B'(f_i) D_S(m_i) \right] Q_i$$

$$+ \left[(f_i + m_i - c_i) D_B(f_i) D_S(m_i) + F_i \right] \left[-\frac{\partial Q_i}{\partial t_i} \frac{\partial \Phi_{Bi}}{\partial f_i} \right] = 0$$

where

$$\frac{\partial \Phi_{Bi}}{\partial f_i} = -D_B(f_i) D_S(m_i)$$

using the definitions of $\Phi_{Bi} = v_B(f_i) D_B(f_i) D_S(m_i)$ and $v_B(f_i) = E\left[b_i - f_i \mid b_i \geq f_i \right]$.

Observe that network i has a unique best reply F_i^{BR} to its rival's prices, (F_j, f_j, m_j). The simultaneous solution to the networks' best replies determines the Nash equilbrium fixed fee of network competition:

Proposition 4 *Suppose that two card networks compete, consumers single-home and merchants multi-home. There is a unique Nash equilibrium when the networks compete in two-part tariffs on the buyer side and per-transaction fees on the seller side. For $i = 1, 2$, equations, (2) and (3) characterize the equilibrium transaction fees, (f_i^*, m_i^*), and the simultaneous solution to equations (4) characterizes the equilibrium fixed fees, F_i^*.*

This result provides one solution to the more general problem of multiplicity of equilibria in the literature:[8] Distinguishing between extensive and intensive margins on the buyer side results in a unique best reply two-part tariff for each network and thereby generates a unique equilibrium in network competition.

Similar to our benchmark analysis, for each network, we define three important price structures allocating the total price, c, between the two sides: the buyers-optimal merchant fee, the sellers-optimal merchant fee, and the volume maximizing merchant fee, respectively,

$$
\begin{aligned}
m_1^{Bc}, m_2^{Bc} &\equiv \arg\max_{m_1, m_2} \sum_{i=1,2} \left[\begin{array}{c} v_B(c - m_i) D_B(c - m_i) D_S(m_i) Q(t_i^*(m_i), t_j^*(m_j)) \\ + \int_{F_i^* - \Phi_B(m_i)}^{\overline{B}_B} x h(x) dx \end{array} \right] \\
m_1^{Sc}, m_2^{Sc} &\equiv \arg\max_{m_1, m_2} \sum_{i=1,2} \left[v_S(m_i) D_B(c - m_i) D_S(m_i) Q(t_i^*(m_i), t_j^*(m_j)) \right] \\
m_1^{Vc}, m_2^{Vc} &\equiv \arg\max_{m_1, m_2} \sum_{i=1,2} \left[D_B(c - m_i) D_S(m_i) Q(t_i^*(m_i), t_j^*(m_j)) \right]
\end{aligned}
$$

Following the lines of Lemma 1 (in the original paper), one could prove that the price structure maximizing the option value of a network's card also maximizes the buyer surplus (gross of fixed fees) when there is network competition, that is, $m_i^{Bc} = m_i^B$ for $i = 1, 2$.

Since the average surplus of buyers and the average surplus of sellers are decreasing in their own usage fees, that is, $v_B'(.), v_S'(.) < 0$ (see the proof of Lemma 1 in the original paper), we have

---

[8]Independently, by formalizing the idea that on each side only a fraction of participants transact, Reisinger (2011) obtains a unique equilibrium when two platforms compete in two-part tariffs on both sides. In our context of the payment card industry this is the case only on the buyer side and the fraction of participants (cardholders) transacting with the other side is $D_B(f)$, which is endogenous, as opposed to his model assuming an exogenous fraction of trade on each side.

$m_i^{Sc} < m_i^{Vc} < m_i^{Bc}$, for $i = 1, 2$.

We previously showed that a network's optimal merchant fee coincides with the one maximizing the option value of that platform's card, and therefore with the buyers-optimal merchant fee: $m_i^* = m_i^B = m_i^{Bc}$ for $i = 1, 2$. On the other hand, the regulator would set m_i^r implementing a price structure that maximizes the sum of buyers' and sellers' surpluses. Hence, we conclude that $m_i^r < m_i^{Bc} = m_i^*$:

Proposition 5 *Suppose that two card networks compete, consumers single-home and merchants multi-home. The privately optimal price structure would lead to a higher merchant fee than the socially optimal price structure.*

□   **Both consumers and merchants multi-home**

Consider now two competing card networks and assume that both consumers and merchants multi-home (this adoption behavior can be rationalized in equilibrium only if the networks are differentiated). If one network increases its IF, this increases its merchant fee and lowers its card usage fee. Suppose first that sellers do not internalize the card usage surplus of buyers, and so accept cards only because they get convenience surplus from card acceptance, like in our benchmark. Then, the increased merchant fee does not affect the demand of the sellers for the rival network (since sellers multi-home), however the lowered card usage fee decreases buyers' point-of-sale usage of the rival card and therefore decreases the buyer demand for the rival network. As a result, the rival network's card usage volume decreases. Since competing networks do not internalize the negative externality of setting a higher interchange fee on the rival's demand, they set a higher interchange fee than a monopoly network. Hence, the equilibrium interchange fees would be distorted further upward than our benchmark of a monopoly network. If sellers do internalize some of the buyers' card usage surplus (for instance, due to business stealing effects of accepting cards, like in Rochet and Tirole, 2002), the negative impact of the lowered card usage fee on card usage demand for the rival card would be partially internalized by merchants and thereby merchants would become less willing to accept the rival network's cards. This implies that the rival network's card usage volume decreases even further and so negative externality of one network's interchange fee on the other would be even higher. Hence, the upward distortion on interchange fees would be even more than

10

the case where sellers do not internalize the card usage surplus of buyers.

5   Elastic consumption and externalities on the market for goods

So far we envisioned the market for one type of services, that is, card services, whose availability at the point of sale (and associated fees) did not affect purchase decisions by consumers. This restriction allowed us to study the market for card transactions in isolation. Externalities of payment cards and fees on purchase decisions by consumers could occur either directly through raising willingness to pay, since card users enjoy convenience benefits, or indirectly through changing the retail prices of card-accepting merchants. As we discuss below, considering simultaneously the effect of a change in IF on the two markets poses a challenge. We provide a first pass towards a unified model accounting for these effects. We show that the trade-off identified in our benchmark does not vanish as a result of allowing for these externalities, we characterize some countervailing effects and discuss when these are likely to be dominated by the source of inefficiency we focus on. We capture these additional features by assuming that the value of the good, v, is a random variable *i.i.d.* across buyers and drawn from a compact interval $[\underline{v}, \overline{v}]$ according to cdf $Z(.)$ satisfying the IHRP. The baseline timing is extended as follows. At stage 3, given b_S, merchants make their membership choices and set the retail price. Upon entering a store, consumers observe their b_B and v as well as the retail price and make their purchase and card usage decisions simultaneously. Let p^* denote the equilibrium retail price charged by cash-only merchants and p^c denote the equilibrium retail price charged by card accepting merchants. To simplify the analysis we focus on the simplest case in which consumers are all alike at the membership stage: $\underline{B}_B = \overline{B}_B = 0$. So in equilibrium all consumers hold a card.

First, we reconsider card usage and consumption choices at a store accepting cards. As in the benchmark case, if $v \geq p^c$, the item is sold regardless of the realization of b_B and a card transaction occurs if and only if $b_B \geq f$. Different from the benchmark, consumers whose valuation is less than the price, $v < p^c$, would purchase the good if the relative surplus from paying by card rather than cash is sufficiently high to compensate the losses from consumption: $b_B - f \geq p^c - v$. Hence, the

total number of card transactions at a store accepting cards is $T(f, p^c)$:

$$T(f, p^c) = (1 - Z(p^c)) D_B(f) + \int_{\underline{v}}^{p^c} D_B(f + p^c - v) w(v) dv. \qquad (5)$$

The latter term reflects the *additional* consumption due to card acceptance. Next consider the incentives of merchants. Assuming zero retailing costs for simplicity, retail profits at a cash only and a card accepting store are, respectively:

$$\max_{p} \; p(1 - Z(p)), \qquad (6)$$

$$\max_{p} \; (p + (b_S - m) D_B(f))(1 - Z(p)) + \int_{\underline{v}}^{p} (p + b_S - m) D_B(f + p - v) z(v) dv.$$

The integral captures the expected extra revenues from additional sales due to card acceptance. For the merchants who have zero surplus from card transactions, type $b_S = m$, retail profits are higher with card acceptance due this additional business. It follows that the marginal merchant type will incur losses per card transaction ($\widetilde{b_S} < m$) and therefore will *not* be indifferent over card usage demand. The threshold $\widetilde{b_S}$ depends on all prices. This is different from the benchmark where the marginal merchant type is equal to the merchant fee. Here, sellers' additional profit from card acceptance at prices f and m is the sum of three terms: the relative seller revenue from charging p^c rather than p^*, the convenience surplus from card transactions and the seller revenue from the additional sales due to card availability:

$$\Phi_S\left(f, m, p^c, \widetilde{b_S}\right) = \int_{\widetilde{b_S}}^{\overline{b_S}} \left[\begin{array}{c} p^c (1 - Z(p^c)) - p^* (1 - Z(p^*)) + (b_S - m) T(f, p^c) \\ + \int_{\underline{v}}^{p^c} p^c D_B(f + p^c - v) w(v) dv \end{array} \right] k(b_S) db_S. \qquad (7)$$

Buyers' additional surplus from cardholding, so called *the option value of a card*, is the sum of the expected convenience benefits from card usage plus the expected additional consumption surplus:

$$\Phi_B\left(f, p^c, \widetilde{b_S}\right) = \int_{\widetilde{b_S}}^{\overline{b_S}} \left[\begin{array}{c} (1 - Z(p^c)) v_B(f) D_B(f) + \\ \int_{\underline{v}}^{p^c} v_B(f + p^c - v) D_B(f + p^c - v) z(v) dv \end{array} \right] k(b_S) db_S. \qquad (8)$$

Since buyers are homogenous at the membership stage, Φ_B is fully extracted by the monopoly issuer through $F^* = \Phi_B$. The issuer profit is equal to the sum of the option value and the revenue

from card transactions:

$$\Pi_I(f, p^c, \widetilde{b_S}) = \Phi_B(f, p^c, \widetilde{b_S}) + (f - c_I + a) \int_{\widetilde{b_S}}^{\overline{b_S}} T(f, p^c)k(b_S)db_S. \tag{9}$$

As in the benchmark, perfectly competitive merchants set $m^* = c_A + a$ and their surplus is equal to zero. Note that f affects the card usage demand *and* the consumption demand. In equilibrium a change in card usage demand affects merchant participation. Besides, changes in f shift the consumption demand at card accepting merchants, and so affect the retail price p^c. For these reasons f^* need not be equal to the opportunity cost $c_I - a$. It follows that $m^*(a) + f^*(a)$ is neither equal to $c_I + c_A$ nor constant over a. In what follows we assume that a unit increase in interchange fee reduces merchant participation, increases retail prices at stores accepting cards and that card accepting stores set higher prices than cash-only stores, respectively:

$$\frac{d\widetilde{b_S}}{da} > 0, \ \frac{dp^c}{da} > 0, \ p^c \ge p^* \text{ for all } \ b_S \ge \widetilde{b_S}. \tag{10}$$

Although strong, these assumptions are in line with common intuition and can easily be verified.

The gap between the total welfare and the issuer's profit is equal to the aggregate merchant profits (net of fees) plus the utility of cash users (net of prices):

$$\begin{aligned}
W - \Pi_I &= \Phi_S + p^*(1 - Z(p^*)) + K(b_S) \int_{p^*}^{\overline{v}} (v - p^*)z(v)dv \\
&\quad + \int_{\widetilde{b_S}}^{\overline{b_S}} \int_{p^c}^{\overline{v}} (v - p^c)z(v)dvk(b_S)db_S.
\end{aligned} \tag{11}$$

A first important observation is that the association fails to internalize the impact of the IF on merchant profits and on cash users' surplus (since $p^c \ne p^*$). To better understand the divergence of private incentives from social incentives, we spell out how changes in IF affect merchants and cash users. Observe that we get our benchmark by setting $\underline{v} = \overline{v}$ and assuming A1. For this special case the derivative of $W - \Pi_I$ with respect to a takes a simple and intuitive form because all merchants set the same prices and therefore there is no impact on cash users:

$$\frac{d(W - \Pi_I)}{da}\bigg|_{\underline{v}=\overline{v}} = \frac{d\Phi_S}{da}\bigg|_{\underline{v}=\overline{v}} = -D_B D_S - \int_m^{\overline{b_S}} (b_S - m)k(b_S)db_S D_B'. \tag{12}$$

13

The first term captures merchants' direct costs from a unit increase in a: the merchant fee increases by one unit decreasing merchants' margin by the amount of card transactions. The second term captures merchants' indirect benefits from a unit increase in a: the card usage fee decreases by one unit (recall that $f^* = c_I - a$) increasing buyers' card usage demand and so merchants' convenience surplus from card transactions. Next we derive the expression for the more general case with elastic demand:

$$\frac{d(W - \Pi_I)}{da}\bigg|_{\underline{v} < \overline{v}} = \frac{d\Phi_S}{da}\bigg|_{\underline{v} < \overline{v}} + k(b_S)\frac{\widetilde{b_S}}{da}\left(\int_{p^*}^{\overline{v}}(v - p^*)z(v)dv - \int_{p^c}^{\overline{v}}(v - p^c)z(v)dv\right)$$
$$- \int_{\widetilde{b_S}}^{\overline{b_S}}\int_{p^c}^{\overline{v}}\frac{dp^c}{da}z(v)k(b_S)dvdb_S. \qquad (13)$$

The last two terms capture the net impact of a small variation of a on cash-users, which is ambiguous. On one hand, a higher a implies less card accepting stores, which is good for cash users as long as $p^c > p^*$ (the second term). On the other hand, a higher a leads to higher prices at all infra-marginal card-accepting stores (the third term). The first term can be decomposed as follows:

$$\frac{d\Phi_S}{da}\bigg|_{\underline{v} < \overline{v}} = \int_{\widetilde{b_S}}^{\overline{b_S}}\left(-T + \frac{df^*}{da}\left[(b_S - m)\frac{dT}{df} + p^c\int_{\underline{v}}^{p^c}D'_B(f + p^c - v)z(v)dv\right]\right)k(b_S)db_S. \qquad (14)$$

Again the trade-off is between merchants' direct losses from a unit increase in m^*, which sum up to the total number of card transactions (the first term) and merchants' indirect benefits from the change in the cardholders' equilibrium choices due to a lower f^* (the term inside the brackets). Contrasting (14) with (12) we note that here the indirect benefits comprises of 2 terms (rather than 1) and are weighted by $\frac{df^*}{da}$. The first term (inside the brackets) is analogous to the indirect benefits in the benchmark: A higher IF decreases f^* increasing the per-merchant card usage volume, T, and so merchants' convenience surplus from card usage. The second term inside the brackets is new and reflects additional merchant profits due to the extra sales arising from the lower card usage fee.

In Section 4 of the original paper we showed that (12) is necessarily negative at the privately optimal fee. That is, following a unit increase in a, the merchants' direct losses from a lower margin offset the indirect benefits from a higher card usage volume. This is not necessarily true when consumption demand is elastic. This is due to three reasons: Compared to the inelastic case, 1) the direct losses might be lower since the card usage demand of buyers is lower than their

14

demand in the inelastic case, $T(f, p^c) < D_B(f)$, as $p^c > \underline{v}$, but there are more merchants accepting cards, $\widetilde{b}_S < m$, 2) the indirect gains might be higher, even if the pass-through of a onto f^* would be below one, since the total card usage demand by buyers might react to card usage fee more than the inelastic case, $\frac{dT(f,p^c)}{df} < D'_B(f)$, (this would happen if $D_B(f)$ is concave), 3) the indirect gains are augmented by an amount proportional to the average price p^c: the higher retail margins, the more merchants would like their customers to have access to cheaper cards. Given these in addition to the ambiguous effect of a on cash-users, we cannot conclude that there would be an upward distortion on the IF.

Appendix

A Efficient Fees

☐ **Proof of Proposition 1**

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 (without arguments):

$$\left[(p - c)(D'_B D_S - D_B D'_S) - v_B D_B D'_S + v_S D'_B D_S \right] Q - \tag{15}$$
$$(p - c + v_B + v_S) D_B D_S Q' \frac{d\Phi_B}{df} + (F - \Phi_B) h \frac{d\Phi_B}{df} = 0.$$

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 $[v_i D_i]' = -D_i$ (see Lemma 1 for the derivation of this property) and the Envelope Theorem, we get the first-order condition:

$$(p - c + v_B) D_B D'_S Q - (p - c + v_B + v_S) D_B D_S Q' \frac{d\Phi_B}{dp} + (F - \Phi_B) h \frac{d\Phi_B}{dp} = 0. \tag{16}$$

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

$$(p - c + v_B + v_S) D_B D_S Q' = (F - \Phi_B) h. \tag{17}$$

Plugging (17) into 16 gives:

$$(p - c + v_B)D_B D'_S Q = 0, \tag{18}$$

which is verified if and only if $p^{FB} = c - v_B(f^{FB})$. Plugging (17) and p^{FB} into condition (15) we get:

$$v_S(p^{FB} - f^{FB}) = v_B(f^{FB}). \tag{19}$$

B   Competing issuers

□   **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, $p_i q_i + p_j q_j \leq I$.

(2) Dixit's (1979) and Singh and Vives's (1984) 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's (1980) 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

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

which implies 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.

□ **Proof of Lemma 1**

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, that is, $F_i^*(\Phi_B)$ 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 $\frac{\partial^2 In Q_i}{\partial t_i \partial t_j} < 0$, 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 $0 < \frac{\partial F_i^*}{\partial \Phi_B} < 1$.

□ **Proof of Proposition 2**

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:

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

$$a^{Sc} \equiv \arg\max_a v_S(c_A + a)D_B(c_I - a)D_S(c_A + a)\left[Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)\right]$$

$$a^{Vc} \equiv \arg\max_a D_B(c_I - a)D_S(c_A + a)\left[Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)\right]$$

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, that is, $a^{Bc} = a^B$. 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, that is, $v_B'(f), v_S'(m) < 0$ (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.