A Welfare Ranking of Two-Sided Market Regimes*

Mats A. Bergman**

Sveriges Riksbank Working Paper Series
No. 185
Revised version September 2005

Abstract

Two-sided network effects in card payment systems are analysed under different market structures, e.g., competition, one-sided monopoly, bilateral monopoly and duopoly; with and without an interchange fee; for the so-called Baxter's case of non-strategic merchants. A partial ranking of market structures according to their welfare effects is provided.

Some support is found for the policy adopted by the EU Commission in the competition law case concerning Visa's interchange fees.

Keywords: Two-sided markets, card payments, payment systems, acquiring, issuing, market structure
JEL code: G21, L11, L44

*The paper was previously circulated under the title "Two-Sided Network Effects, Bank Interchange Fees, and the Allocation of Fixed Costs".

**Swedish Competition Authority and Department of Economics, Södertörn University College, SE-141 89 Huddinge, Sweden. E-mail: mats.bergman@sh.se and mats.bergman@kkv.se.
I am grateful for comments from Gabriela Guibourg, Björn Segendorff, from seminar participants at the Research Institute of Industrial Economics, Swedish Riksbank and Umeå Univeristy, from participants at the NORIO V conference and at the conference in tribute to Jean-Jacques Laffont, as well as for financial support from Sveriges Riksbank. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

1 Introduction

Payment systems, such as credit and debit-card networks, have features that make economic analysis particularly challenging, for regulators as well as for theoretical analysts. The markets are two-sided, in the sense that two different types of agents - e.g., card holders and merchants - must adopt the system for it to be fully functional. At the same time, payment systems are often set up jointly by a large number of banks, suggesting the possibility that the systems are designed in such a way that competition between the banks is limited.

In two-sided markets in general, two types of agents (or consumers) interact through a common platform. In order to maximize the utility that can be derived from such a system (or market), a sufficient number of consumers of both types must adopt the platform. In the particular case of debit and credit cards, this means that adopting the card must be attractive both for potential cardholders and for merchants that consider installing card-reading facilities (EFTPOS terminals). An optimal encouragement for both types generally requires that one side subsidizes the other. In a so-called proprietary system - i.e., a system where a single firm (bank) operates the system and provides services to both sides of the market - this is achieved by simply setting appropriate (and, typically, different) prices on the two sides of the market.[1] However, it will often be necessary that a large number of competing service providers (e.g., banks) join a common platform, in order to share costs and so as to create a sufficiently large mass of users within a single system. In systems with multiple providers, known as four-party systems, competitive pricing may not result in an optimal balancing between the two sides of the market. To achieve optimal network effects, the providers can introduce a fee structure for their internal transactions, that serves the purpose of subsidising one side of the market, while taxing the other.

In the context of payment-card system, this is achieved through so-called multilateral interchange fees (to be explained futher in see Section 2.) However, this fee structure is, by its nature, an agreement between competitors that influences their individual pricing decisions towards the two types of consumers. In general, regulatory authorities have well-founded reasons for being suspicious of close cooperation between competitors, in particular concerning pricing decisions. Competition between independent firms fosters efficiency and tends to bring prices down to costs. In the provision of payment services, however, completely independent competition will not be a feasible alternative. The best available option may be competition at the retail level in combination with cooperation at the upstream (system) level. On the other hand, close cooperation at one level may give the banks the opportunity to cleverly design system fees and multilateral fees in such a way that downstream collusion is induced. This could, for example, be achieved by raising the appropriate marginal costs, so that incentives are created for the banks to raise final-customer prices and so

[1] A possible complication is that it may be optimal to set prices on one side of the market below (variable) costs. If the provider is dominant, this may (incorrectly) be interpreted as predatory pricing.

that excess profits are generated elsewhere in the system. Considerations of this type has drawn the attention of regulators, resulting in a substantial amount of regulatory activity. The EU Commission, for example, forced Visa and (indirectly) Mastercard[2] to reduce their multilateral interchange fees (see below for a discussion of interchange fees; see Bergman, 2003, and Chakravorti, 2003, for references to cases).

For much the same reasons, payment systems have spawned a considerable economics literature in recent years. In a wider context, these contributions can be seen as a part of the growing literature on two-way network effects (Rochet and Tirole, 2003; Armstrong, 2004). A key insight of this literature is that network effects between two different types of consumers have to be analysed in a system-wide context.

Under the assumptions of a single platform, non-strategic final users (cardholders and merchants) and linear prices, this paper explores the welfare effects of different market structures. The modelling assumptions resemble those of Schmalensee (2002), but the analysis is extended to a large number of market structures: perfect competition, a proprietary system (a two-sided monopoly), one-sided monopoly and bilateral monopoly; with and without interchange fees that may be set in order to maximize welfare or profit. In addition, some results are provided for duopoly markets under quantity competition.

The main findings are reported in Proposition 3. The introduction of market power in a market with two-sided network effects has two opposing effects. On the one hand, the firm with market power will try to extract profit. This tends to increase margins and to reduce welfare. On the other hand, a firm that can influence prices will have incentives to balance the network effects between the two sides of the markets, which tends to increase welfare. The net effect is indeterminate: welfare can be either higher or lower under competition that under a proprietary system. Similarly, given that one firm has monopolized one side of the market, while the other side remains competitive, the introduction of an interchange fee can either increase or reduce welfare. The former will happen if the main consequence is a better balance between the two sides; the latter if the ability of the one-sided monopoly to extract profits from the other side dominates.

The trade-off between the two types of inefficiences suggests that an intermediate policy concerning the interchange fee may be called for, one that allows some degree of cross subsidies between the two sides of the market, while still limiting the scope for monopoly pricing. To some extent, this supports the policy adopted by the EU Commission in the Visa case - i.e., to allow a positive interchange fee, but to cap its level - although establishing the optimal level of the cap would most likely require a detailed market simulation, using estimates of the demand structure on the two sides, detailed knowledge of the cost structure and an assessment of how price-cost margins will be set by banks on the

[2] The EU Commission has a case open concerning Mastercard, but has at the time of writing (April 2005) not yet taken a decision. See the XXXIIIrd Report on competition policy (p. 45), DG Competition, EU Commission, which states that Mastercard will have to abide by the principles formulated in the earlier Visa-card decision.

two sides.

The outline of the paper is as follows. Sections 2 discusses network effects in general and (card) payment systems in particular, while Section 3 sets up the basic model. In Sections 4 through 8 this model is applied to different market structures: perfect competition and welfare maximum (Section 4), second-best regulation of the interchange fee (Section 5), bilateral monopoly (Section 6), a proprietary system (Section 7) and a one-sided monopoly (Section 8). Section 9 makes welfare comparisons between these market configurations. In section 10, some aspects of oligopoly interaction are introduced and further references to the literature are provided. Finally, section 11 concludes.

2 Network effects and payment markets

In a market without network effects, the consumer cares only about his or her own level of consumption and about the price. In a market with network effects, the consumer cares - directly or indirectly - also for other consumers' levels of consumption. In the simplest setting, the number of other consumers of the same product has a direct effect on the (marginal) utility of consuming a unit of the product. For example, a given consumer's utility from having a phone or a fax increases with the number of other consumers that also have phones and faxes, respectively. This type of network effect is sometimes called a one-sided network effect.

A somewhat more complex situation arrises when there are two types of agents that interact on one "platform". Either type cares for the number of agents of the *other* type that uses the platform, but not (directly) about the number of users of its own type. Some examples are buyers and sellers in advertising markets and marketplaces for trading (e.g., stock markets), as well as matchmaking markets (dating agencies, real estate agents, business-to-business websites *et cetera*). A buyer does not benefit from the presence of other buyers - and may indeed suffer from the increased competition for the sellers' product that additional buyers bring. On the other hand, the buyer derives benefit from the presence of additional sellers, while the sellers derive benefit from the presence of additional buyers. Hence, buyers may indirectly benefit from there being a large number of other buyers, as this will attract a large number of sellers - and vice versa. This phenomenon is known as a two-sided network effect. Another example is the market for operative systems for personal computers: the operative system is a platform that is used by software manufacturers and by users of personal computers. An operative system such as Windows, that has a large installed base of users, is an attractive platform for software developers. Conversely, if a large number of applications have been developed for an operative system, that system will be attractive for new users. More generally, many manufacturing standards (computers and peripherals, CD players and CDs, et cetera) and communication protocols are examples of markets with two-sided network effects. Yet another example is shopping malls, which must attract customers as well as retailers.

In the financial markets, a payment-card system is an example of a two-way market: cardholders cannot interact with other cardholders and nor can merchants interact with other merchants (e.g., using their EFTPOS terminals), but cardholders can interact with merchants.[3]

Sometimes a third type of network effect is identified: indirect network effects in one-sided markets. Possible examples are public-transport networks and (single-bank) ATM networks. A higher number of passengers and a higher number of cardholders on the ATM network, respectively, will result in more frequent departures and a denser (or wider) ATM network. This increases welfare for the average customer, even though congestion effects may imply that the *direct* effect on a given passenger's utility of another passenger may be negative, and similarly for an additional ATM cardholder. This type of network effect is very reminiscent of ordinary scale (or density) economies: as the number of customers in a retail outlet increases, the retailer can expand its range of products, it can extend opening hours and it can often reduce prices. Similarly, the manufacturer of some widget will often be able to reduce average costs when the scale of production increases. For this reason, the concept of indirect network effects appears to be lacking in rigor. However, if *different* banks join the same ATM network, or if *different* airlines, say, use the same airport, then this can be seen as an example of a market with a platform and two-sided network effects.[4]

The market that fits the modelling assumptions of this article is the payment card market (the market for debit and credit cards). This is an example of a two-sided market with four types of participants, as shown in Figure 1: cardholders (or customers), issuers, acquirers and merchants. Cardholders use cards to pay for goods and services provided by merchants. The cards are provided by issuing banks, while acquiring banks provide services to merchants. When a card payment is registered with a merchant, the acquiring bank charges the issuing bank, which in turn charges the cardholder's account. For its services, the acquiring bank subtracts a fraction, the so-called merchant fee, from the payment to the merchant's account. Typically, the acquiring bank will have to pay a fee to the issuing bank, the so-called interchange fee, which is therefore subtracted from the payment from the issuer to the acquirer. The merchant fee and the interchange fee can be either a percentage of the transaction value, or a fixed fee. In most cases, the cardholder pays an annual fee, while in some cases, he or she will (also) have to pay per-transaction fees. Credit cards may be free of charge for the cardholder, in particular in the USA, as long as the accumulated debt is paid monthly.

Relative to cash payments, card payments potentially generate benefits for customers, as well as for merchants. For cardholders, these benefits will increase

[3] Although, for some purposes, it may be appropriate to view a giro system as a two-sided market, with business and non-business customers, a giro system is perhaps better thought of as a market with one-sided network effects. Some account holders (non-business customers) may not be able to receive payments, but all account holders (business customers as well as non-business customers) are able to make payments.

[4] Armstrong (2004) provides further examples and analyses network effects in two-sided markets in a general setting.

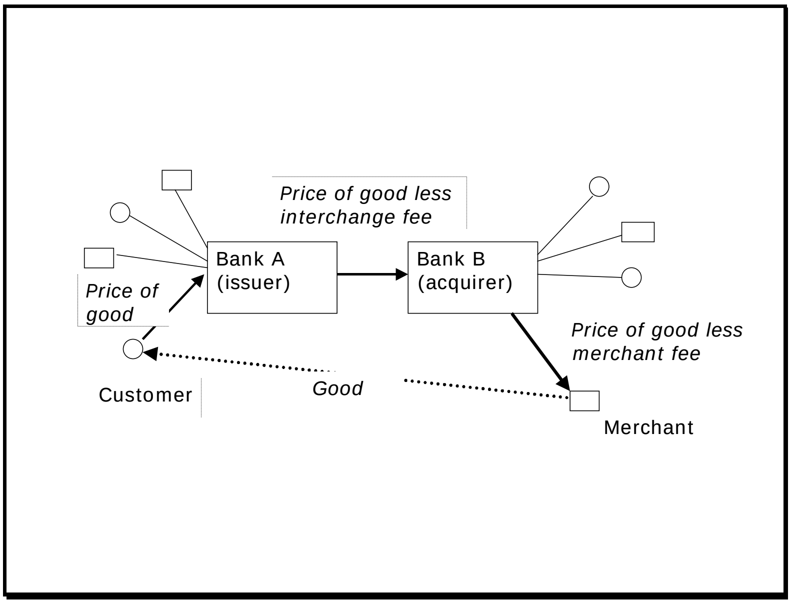

Figure 1: Multilateral interchange fees

with the number of merchants that accept cards, while merchants will be more willing to accept a card that has a larger base of cardholders. This interdependence creates two-way network effects, which will be discussed more extensively in Section 2. In the presence of network effects, the competitive equilibrium may be inefficient.

The interchange fee, which is set multilaterally by the banks, provides an instrument for improving efficiency. A higher interchange fee tends to reduce cardholders' fees and to increase the merchant fee. The first-generation literature on two-way network effects in payment markets (Baxter, 1983, Schmalensee, 2002) analysed these issues under the assumption that cardholders' and merchants' demand for card-payment services reflect the intrinsic benefits - such as convenience and safety - that they derive from card payments (relative to cash payments).[5] A typical result is that there will be underprovision of card services, because private agents do not internalize positive network effects that accrue to other agents.

An important insight, however, was that merchants with market power (with a positive price-cost margin) may have *strategic* reasons to accept cards. By accepting cards, they will attract customers away from other merchants that do not accept cards. Conversely, if they do not accept cards, they may loose customers to other merchants that do. These strategic motives do not correspond to social gains. It follows that there may potentially be *overprovision* of card

[5] These assumptions were dubbed "Baxter's case" by Rochet and Tirole, 2002.

services (Katz, 2001). In Rochet and Tirole (2002) these strategic motives are explicitly accounted for by incorporating merchant competition in the models (see further Section 10).

3 The model

In the following, it will be assumed that cardholders, as well as merchants that accept cards, pay on a per-transaction basis, rather than an annual fee and that non-linear pricing schemes cannot be used. The cardholders pay p_1 per transaction, while merchants pay p_2. The number of consumers that chooses to hold a card is given by the inverse demand function.[6]

$$p_1 = \phi_1(n_1) \tag{1}$$

and the number of merchants that accepts cards is given by inverse demand function

$$p_2 = \phi_2(n_2) \tag{2}$$

where n_1 is the number of cardholders and n_2 is the number of merchants that accepts cards, p_i is the per-transaction price for cardholders ($i = 1$) and merchants ($i = 2$), with $\phi' < 0$ for both customer groups. Let $\phi_i(0) = \overline{p}_i$ for $i = 1, 2$ be the maximum price any consumer (merchant) is willing to pay and let $\phi_i(N_i) = 0$, where N_i is the total number of consumers and merchants with non-negative valuations, for $i = 1$ and 2 respectively. Assume that N_1 and N_2 are large numbers and that each identical consumer wishes to make an equal number of transaction with every firm, normalized to 1 for every consumer-merchant pair.[7] As mentioned in the introduction, merchant (and consumer) demand is assumed to correspond to social benefits.

Given that consumer n_i has adopted a payment card, he will use it with any merchant that accepts cards. Given the assumed (inverse) demand functions above, if consumer n_i uses cards, so will consumers $1, ..., n_i - 1$; given that merchant n_j accepts cards, so will merchants $1, ..., n_j - 1$. Under the assumption that each consumer buys once from each merchant, the total number of card transactions will be $n_i n_j$, where n_i and n_j now represent the highest-number consumer and merchant, respectively, that holds a card or accepts card payments. (That is, consumer $n_i + 1$ do not use a payment card and merchant $n_j + 1$ do not accept card payments.)[8]

[6] Rochet and Tirole, 2003 (see their note 6), refer to the corresponding type of demand functions as "quasi-demand functions", since these functions characterizes demand for a certain type of transaction from one side of the market only and since the total number of transactions will be given by the product of n_1 and n_2, as will be explained below.

[7] The key assumption is that the number of transactions made between a given consumer (potential cardholder) and a given merchant (with or without card facilities) is independent of the two agents' valuations of card transactions relative to cash payments. See Rochet and Tirole, 2003, p. 995.

[8] The model resembles that of Schmalensee (2002).

Assume that issuing banks have equal and constant marginal costs c_1 per transaction and that acquiring banks' constant marginal cost is c_2 per transaction. Possibly, there is an interchange fee a, that adds to the marginal cost of the acquiring banks and subtracts from the marginal cost of issuing banks (or conversely, for negative values of a). An underlying assumption is that there is just one "platform" (i.e., card system) and that this platform either operates under a not-for-profit basis, or is vertically integrated with issuing and acquisition (i.e., a proprietary system) or with issuing only. That is, issues concerning rivals' access to a bottleneck facility controlled by a vertically integrated company are assumed away.

The number of merchants that accepts card payments is irrelevant for the decision to become a cardholder and the number of individuals that holds cards is irrelevant for the merchant's decision to accept card payments. This is so, first, since both groups pay purely on a per-transaction basis (i.e., there are neither per-customer fees, nor fixed customer costs associated with holding a card or maintaining the capacity to accept card payments) and, second, because it is assumed that there are no strategic effects ("Baxter's case"). That is, it is assumed that merchants do not choose to accept cards in order to induce cardholders to use *their* outlet, rather than another one.[9]

However, there will be network effects between the two consumer groups. The number of transactions a cardholder makes, and therefore his utility, depends on the number of merchants that accepts cards. Conversely, the number of cardholders affects the number of transactions a merchant makes, as well as her utility. One additional cardholder will increase the consumer surplus of every merchant that accepts cards, while an additional card-accepting merchant will increase the consumer surplus of every cardholder. In other words, although network effects will be irrelevant for the adoption decisions of potential cardholders and potentially card-accepting merchants, and hence for the equilibrium outcome in a competitive market (in the absence of subsidy schemes), they appear in the welfare analysis and they will be relevant for parties with market power.

If a fixed annual fee were introduced, the number of merchants that accepts cards would of course be relevant when a consumer decides whether to adopt a card or not. A consumer i would then become a cardholder only if $(\phi_1(i) - p_1)n_2 \geqslant F$, where the term within bracket equals consumer i's per-transaction surplus, the left-hand side term equals his total surplus and F is the (fixed) annual fee. A corresponding condition would hold for merchants, if they were to pay a fixed annual fee. Although a fixed fee would add to the realism of the model, it will in the following be assumed that only per-transaction fees are used; including fixed fees in the analysis would greatly complicate the analysis.

[9] Such a business-stealing effect is analysed by Rochet and Tirole, 2002, and Wright, 2004. See also the discussion in section 10 below.

4 Perfect competition and welfare maximum

The outcome under perfect competition is straightforward. Since no bank has any market power, prices will be driven to marginal costs. In the absence of an interchange fee, $p_1 = c_1$ and $p_2 = c_2$ will hold. With an interchange fee (issued by a not-for-profit platform), the outcome will be $p_1 = c_1 - a$ and $p_2 = c_2 + a$.

Total welfare will be the sum of the cardholders' and the merchants' valuations of card transactions, less the sum of the costs at the issuing and acquiring sides of the market that these transactions give rise to. That is, welfare is given by

$$W = \int_0^{n_1} (\phi_1(x_1) - c_1)\, n_2 dx_1 + \int_0^{n_2} (\phi_2(x_2) - c_2)\, n_1 dx_2 \tag{3}$$

where n_1 and n_2 again represent the number of cardholders and the number of merchants that accept cards, respectively. The first term of the right-hand side integrand - the per-transaction consumer surplus of cardholder x_1 - is multiplied with n_2, since n_2 merchants accept cards and, therefore, each consumer that holds a card will be able to use the card for n_2 transactions. Similarly, every merchant will meet n_1 cardholders that wishes to make one transaction each.

Evaluating the integral, we find that

$$W = (\Phi_1(n_1) - c_1 n_1)\, n_2 + (\Phi_2(n_2) - c_2 n_2)\, n_1 \tag{4}$$

where $\Phi_i(n_i) = \int_0^{n_i} \phi(x_i) dx_i$. Differentiating eq. (4) with respect to n_1 and n_2, we find the following first-order conditions for welfare maximization[10]

$$\begin{aligned} (\phi_1(n_1) - c_1)\, n_2 + \Phi_2(n_2) - c_2 n_2 &= 0 \\ (\phi_2(n_2) - c_2)\, n_1 + \Phi_1(n_1) - c_1 n_1 &= 0 \end{aligned} \tag{5}$$

or, using (1) and (2),

$$\begin{aligned} (p_1 - c_1) n_2 + \Phi_2(n_2) - c_2 n_2 &= 0 \\ (p_2 - c_2) n_1 + \Phi_1(n_1) - c_1 n_1 &= 0 \end{aligned} \tag{6}$$

or

$$\begin{aligned} p_1 &= c_1 - (\frac{\Phi_2(n_2)}{n_2} - c_2) \\ p_2 &= c_2 - (\frac{\Phi_1(n_1)}{n_1} - c_1) \end{aligned} \tag{7}$$

Assuming an interior solution, eqs. (7) characterize the optimal prices. Since n_2 depends on p_2 and n_1 depends on p_1, the equations constitute a simultaneous-equations system, the solution of which is the pair of optimal prices, p_1 and p_2.

[10] The corresponding second-order conditions for welfare maximization are $W_{n_1 n_1} = \phi_1'(n_1) n_2 < 0$, $W_{n_2 n_2} = \phi_2'(n_2) n_1 < 0$ and $W_{n_1 n_1} W_{n_2 n_2} > [W_{n_1 n_2}]^2$, where $W_{n_1 n_2} = \phi_1(n_1) + \phi_2(n_2) - c_1 - c_2$ and $W_{n_i n_j} = \frac{\partial^2 W}{\partial n_i \partial n_j}$.

Note that because of the network effects, the solution depends on the global properties of the two demand functions, rather than the marginal properties around the equilibrium (or optimum) point.

In words, the socially optimal cardholder price equals the marginal cost that one transaction gives rise to on the issuing side of the market, minus the difference between the *average* valuation that merchants that accept card payments assigns to a card transaction and the marginal cost of one transaction at the acquiring side. Similarly, the optimal merchant fee equals marginal costs on the acquiring side minus the difference between the cardholders' average valuation and marginal costs on the issuing side. The terms within parenthesis account for the two-sided network effect between cardholders and merchants; the marginal customer on one side of the market will only consider his or her own benefit, not the additional benefit that inframarginal customers on the other side of the market derive from being able to make one additional transaction each.

The following proposition asserts that optimal (linear) prices will result in the card system earning negative profits.

> *Proposition 1.* In welfare-generating card systems where the optimum corresponds to the first-order conditions, the socially optimal linear transaction prices will be such that the combined revenues of the acquirers and the issuers will be lower than their combined costs.[11]
>
> *Proof.*
>
> The socially optimal prices and the corresponding number of cardholders and merchants that accept cards are defined by eqs. (1), (2) and (7). Let n_1^* and n_2^* be the number of cardholders and card-accepting merchants in optimum. There will be $n_1^* n_2^*$ transactions and each transaction will give an issuer a net profit of $-(\Phi_2(n_2^*)/n_2^* - c_2)$ and an acquirer a net profit of $-(\Phi_1(n_1^*)/n_1^* - c_2)$. Possibly, the net profit will be positive on one of the two sides. However, aggregating over the two sides of the market and multiplying by the number of transactions, total profits will be
>
> $$\begin{aligned} & -\left((\frac{\Phi_2(n_2)}{n_2} - c_2) + (\frac{\Phi_1(n_1)}{n_1} - c_1) \right) n_1 n_2 \\ = \quad & -(\Phi_1(n_1) - c_1 n_1) n_2 + (\Phi_2(n_2) - c_2 n_2) n_1 = -W \end{aligned} \tag{8}$$
>
> where the last equality comes from eq. (4). It follows that for any card system that is able to generate positive welfare, the optimal linear transaction prices will not allow issuers and acquirers to recover their combined costs. ◊

It follows from the propostition that optimal prices are below costs. Only if a card system is unable to generate a net surplus should there be no subsidies

[11] Similar results have been derived in Bold and Tieman (2003, 2004) and Armstrong (2004).

- but then there should be no card system either.[12] On the other hand, if the subsidies are funded by measures that distort allocative efficiency elsewhere, then the welfare gains must be weighed against the welfare losses that are caused by the mechanism used for raising funds. Note that it follows directly from the proposition that the sum of the two prices must be lower than total marginal costs.

5 Second-best interchange fees

Given that all cardholders pay the same per-transaction price, p_1, and similarly for the merchants, given price taking by issuers and acquires and given that the number of transactions is, by definition, the same on both sides of the market, the second-best (or Ramsey) price structure (the optimal price structure in the absence of subsidies) can be achieved with an interchange fee. The first-order condition for welfare maximization is then[13]

$$\frac{dW}{da} = \frac{\partial W}{\partial n_1}\frac{dn_1}{dp_1}\frac{dp_1}{da} + \frac{\partial W}{\partial n_2}\frac{dn_2}{dp_2}\frac{dp_2}{da} = 0 \tag{9}$$

Using eq. (6) and the fact that $dp_1/da = -1$ and $dp_2/da = 1$ under perfect competition, we have that

$$\begin{aligned} 0 \quad = \quad & -\left((p_1 - c_1)n_2 + \Phi_2(n_2) - c_2 n_2\right)\frac{dn_1}{dp_1} + \\ & + \left((p_2 - c_2)n_1 + \Phi_1(n_1) - c_1 n_1\right)\frac{dn_2}{dp_2} \end{aligned} \tag{10}$$

Multiplying both the denominator and the numerator with $p_1 p_2 / n_1 n_2$ and introducing the price elasticities of (quasi) demand, $\varepsilon_i = -\frac{\partial n_i}{\partial p_i}\frac{p_i}{n_i}$ for $i = 1, 2$,

[12] If two-part tariffs can be used, efficiency can be improved relative to a situation with only a fixed-fee. For example, if the (inverse) demand curve ϕ does not represent variations in willingness-to-pay between different consumers, but rather an individual (representative) consumer's (cardholder or merchant) willingness-to-pay for additional transactions, then the variable (per-transaction) fee can be set equal to the socially optimal price, while the fixed fee can be set high enough to cover the producers' deficit. In this case, the fixed fees for cardholders and merchants, f_1 and f_2, could be set such that $n_1 f_1 + n_2 f_2 = W$, leaving the banks with zero profit and the consumers sharing all surplus, also equal to W.

However, if the valuations differ between consumers, a fixed fee will discourage some consumers from adopting cards, even though it would be socially optimal for them to do so. Hence, when determining the optimal fee structure, there will be a trade-off between low marginal fees, so as to encourage cardholders to use the card every time it is efficient to do so, and low fixed fees, so as to encourage low-demand consumers to become cardholders.

For merchants, whether to accept cards or not is arguably an all-or-nothing choice. I.e., a merchant cannot prevent some customers from paying with cards, while allowing others to do so. Hence, given that a merchant has decided to accept cards, a high variable merchant fee will not result in sub-optimal card use in that merchant's facilities. Ceteris paribus, it appears that the merchants' lack of discretion in this respect will lead to lower fixed fees and higher variable fees.

A further analysis of two-part tariffs is beyond the scope of this article. In the following, linear pricing will be assumed.

[13] Cf. proposition 2 in Rochet and Tirole, 2003, and the discussion in Section 7 below.

the expression can also be written

$$\frac{p_1 - c_1 - c_2 + \frac{\Phi_2(n_2)}{n_2}}{p_1}\varepsilon_1 = \frac{p_2 - c_1 - c_2 + \frac{\Phi_1(n_1)}{n_1}}{p_2}\varepsilon_2 \tag{11}$$

This expression resembles the optimality conditions for a price discriminating monopolist, $(p_i - c)/p_i = 1/\varepsilon_i$ or $\varepsilon_1(p_1 - c)/p_1 = \varepsilon_2(p_2 - c)/p_2$. Instead of the normal Learner index, the expression includes an "extended" Learner index that includes the average per-transaction net value created on the "other" side. As expected, this "extended" Learner index will be higher on the less elastic side of the market.

Alternatively, noting that $p_1 - c_1 = -a$ and $p_2 - c_2 = a$, equation (10) can be rewritten as

$$(an_2 - \Phi_2(n_2) + c_2 n_2)\frac{dn_1}{dp_1} + (an_1 + \Phi_1(n_1) - c_1 n_1)\frac{dn_2}{dp_2} = 0 \tag{12}$$

Solving for a gives

$$a = \frac{(\Phi_2(n_2) - c_2 n_2)\frac{dn_1}{dp_1} - (\Phi_1(n_1) - c_1 n_1)\frac{dn_2}{dp_2}}{n_2\frac{dn_1}{dp_1} + n_1\frac{dn_2}{dp_2}} \tag{13}$$

Note that this is not a closed expression, since n_1 and n_2 depends on prices and hence on a. In principle, eq. (13) can be solved for explicit demand functions, but the solutions are likely to be highly nonlinear. However, introducing once again elasticities, the above equation can be rewritten as

$$a = \frac{p_2\left(\frac{\Phi_2(n_2)}{n_2} - c_2\right)\varepsilon_1 - p_1\left(\frac{\Phi_1(n_1)}{n_1} - c_1\right)\varepsilon_2}{p_2\varepsilon_1 + p_1\varepsilon_2} \tag{14}$$

According to eq. (14), subsidies will tend to flow towards the issuing side of the market if cardholders' demand is more elastic than merchants' demand and if the difference between the average valuation of the merchants and the acquirers' cost in equilibrium is high relative to the difference between the average valuation of the cardholders and the issuers' cost - also evaluated in equilibrium.[14]

Under the assumptions of perfect competition on both sides of the market and in the context of the present model, there is no point trying to use the interchange fee as a collusion device: issuers and acquirers will still make zero profit. In fact, as noted by Schmalensee (2002), if both acquirers and issuers are perfectly competitive, there are no reasons for them to have any particular preferences over the interchange fee at all.

[14] This result can be contrasted with Proposition 2 in Rochet and Tirole, 2003, which states that, in a proprietary system, prices will be higher in the *elastic* market. See also the discussion below, in Section 7 and footnote 17.

6 Bilateral monopolies

In this section, it will be assumed that the market is controlled by two monopolies: one monopoly issuer and one monopoly acquirer. This is the simplest way to introduce market power in the model except, possibly, for assuming a single monopoly provider of both issuing and acquiring services. The latter alternative, however, also implies that a single entity can internalize network effects between the two sides of the market.

If the two monopolies take a as given, they will maximize the following profit functions

$$\begin{aligned} \pi_1 &= [\phi_1(n_1) - c_1 + a]n_1 n_2 \qquad (15) \\ \pi_2 &= [\phi_2(n_2) - c_2 - a]n_2 n_1 \end{aligned}$$

by choosing n_1 (the issuer) and n_2 (the acquirer). The monopoly issuer will take n_2 as given, while the monopoly acquirer will take n_1 as given. Using $\phi_1(n_1) = p_1$ and $\phi_2(n_2) = p_2$, the first-order conditions will be

$$\begin{aligned} p_1 &= c_1 - a - \phi_1'(n_1)n_1 \qquad (16) \\ p_2 &= c_2 + a - \phi_2'(n_2)n_2 \end{aligned}$$

or

$$\begin{aligned} \frac{p_1 - c_1 + a}{p_1} &= \frac{1}{\varepsilon_1} \qquad (17) \\ \frac{p_2 - c_2 - a}{p_2} &= \frac{1}{\varepsilon_2} \end{aligned}$$

Hence, the two monopolies will price above their marginal costs, in contrast to the below-cost pricing required for achieving the social optimum. Again, the formulaes look like standard monopoly pricing conditions: the Learner index for one side of the market will equal that market side's inverse demand elasticity.

As noted by Schmalensee (2002), the interchange fee does not resolve the problem of double marginalization, since it only shifts marginal costs from one side of the market to the other. However, it can mitigate problems that are due to differences (in terms of elasticities and average valuations of inframarginal customers) between cardholders and merchants.[15]

Assume further that two monopolists set a so as to maximize their combined profits and that, subsequently, each firm will maximize its own profit, taking a as given. The problem can be modelled as a two-stage game. In the second stage, each of the two firms maximize its own profit. In the first stage, they jointly set the interchange fee so that the sum of their profits is maximized. (I.e., it is assumed that equal weights are given for both firms' profits.) The latter is given by

$$\Pi = [\phi_1(n_1) - c_1 + \phi_2(n_2) - c_2)n_1 n_2 \qquad (18)$$

[15] Schmalensee provides further results on interchange fees under bilateral monopoly.

Hence, the maximization problem is given by

$$\underset{n_1,n_2}{Max}\Pi \tag{19}$$

$$\text{st } c_1 + c_2 - \phi_1(n_1) - \phi_2(n_2) - \phi_1'(n_1)n_1 - \phi_2'(n_2)n_2 = 0$$

where the constraint is derived by adding the two first-order conditions for the second-period maximization, eqs. (16).

Let the associated Lagrangian be $\mathcal{L}$, then the first-order conditions for constrained maximization are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial n_1} &= \phi_1'(n_1)n_1n_2 + [\phi_1(n_1) + \phi_2(n_2) - c_1 - c_2]n_2 - \lambda[2\phi_1'(n_1) + \phi_1''(n_1)] = 0 \\
\frac{\partial \mathcal{L}}{\partial n_2} &= \phi_2'(n_2)n_1n_2 + [\phi_1(n_1) + \phi_2(n_2) - c_1 - c_2]n_1 - \lambda[2\phi_2'(n_2) + \phi_2''(n_2)] = 0 \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= c_1 + c_2 - \phi_1(n_1) - \phi_2(n_2) - \phi_1'(n_1)n_1 - \phi_2'(n_2)n_2 = 0 \qquad (20)
\end{aligned}$$

Eliminating λ from the first two conditions gives

$$\phi_1'(n_1)n_1 + \frac{1}{2}\phi_1''(n_1)(n_1)^2 = \phi_2'(n_2)n_2 + \frac{1}{2}\phi_2''(n_2)(n_2)^2 \tag{21}$$

Together with the constraint, the above equation characterizes pricing by a bilateral monopoly that maximizes combined profits when setting a, but where each firm then maximizes its own profit, taking a as given.

7 A proprietary system (two-sided monopoly)

A two-sided monopoly - e.g., a proprietary system, such as American Express, although in a monopoly position - would maximize the sum of its profits generated on the two sides of the market, i.e.

$$\pi = [\phi_1(n_1) - c_1]n_1n_2 + [\phi_2(n_2) - c_2]n_1n_2 \tag{22}$$

Obviously, an interchange fee would play no role for a proprietary system. The first-order conditions for profit maximization, obtained by differentiating eq. (22) with respect to n_1 and n_2, will be

$$\begin{aligned}
[\phi_1'(n_1)n_1 + \phi_1(n_1) - c_1] + [\phi_2(n_2) - c_2] &= 0 \qquad (23)\\
[\phi_2'(n_2)n_2 + \phi_2(n_2) - c_2] + [\phi_1(n_1) - c_1] &= 0
\end{aligned}$$

Substituting $p_1 = \phi_1(n_1)$ and $p_2 = \phi_2(n_2)$ into the above equations, we have

$$\begin{aligned}
\phi_1'(n_1)n_1 + p_1 - c_1 + p_2 - c_2 &= 0 \qquad (24)\\
\phi_2'(n_2)n_2 + p_2 - c_2 + p_1 - c_1 &= 0
\end{aligned}$$

Or, alternatively,

$$p_1 + p_2 - c_1 - c_2 = -\phi_1'(n_1)n_1 = -\phi_2'(n_2)n_2 \tag{25}$$

The sum of the mark-up above costs on the two sides of the market should, hence, be equal to the revenue increase on inframarginal units on *one* side of the market, if output is reduced by one unit on that side.[16] Alternatively, one could say that when the monopolist considers increasing production by one unit on one side of the market, its marginal (per-transaction) costs will be $c_1 + c_2$, since each transaction necessarily involves both the issuing and the acquiring side. At the same time, its marginal revenues will include revenues generated from the additional sales on both sides of the market, *but only losses from price reductions on inframarginal units on one side of the market.* For example, the monopoly may consider lowering the price on the issuing side of the market. This will generate additional sales on the issuing side of the market, but also on the acquiring side, since each transaction involves both issuing and acquiring. Additional sales implies additional costs and revenues on both sides of the market, but the price reduction will reduce its revenues from inframarginal customers on the issuing side of the market only. Naturally, the monopoly should optimize against both sides of the market, so that the revenue increase on inframarginal customers that would result if the number of customers were reduced should be equal on the two sides of the market.

Eq. (25) can be re-written

$$\begin{aligned} \frac{p_1 + p_2 - c_1 - c_2}{p_1} &= \frac{1}{\varepsilon_1} \\ \frac{p_1 + p_2 - c_1 - c_2}{p_2} &= \frac{1}{\varepsilon_2} \end{aligned} \tag{26}$$

Again, the expression resembles the standard expression for monopoly pricing, except that the Learner index is extended to include the net profit generated on the other side of the market.[17] (In contrast, the welfare maximum incorporates the average net value created on the other side, as explained in Section 4.) Following Rochet and Tirole (2003, their Proposition 1), the above expression implies that

$$\frac{p - c}{p} = \frac{1}{\varepsilon} \tag{27}$$

i.e., the classical Learner formula, where $p = p_1 + p_2$, $c = c_1 + c_2$ and $\varepsilon = \varepsilon_1 + \varepsilon_2$. Less intuitively, it also follows directly from eq. (26) that

$$\frac{p_1}{\varepsilon_1} = \frac{p_2}{\varepsilon_2} \tag{28}$$

i.e., that the prices will be proportional to the elasticities, rather than the price-cost margin being proportional to the *inverse* of the elasticities. The

[16] See also Proposition 1 in Rochet and Tirole, 2003.

[17] An alternative interpretation is that the perceived effective marginal cost of selling an additional unit in, e.g., market 1 is $\widetilde{c}_1 = c_1 + c_2 - p_2$. Given $\widetilde{c}_1$, a standard monopoly-pricing condition holds: $(p_1 - \widetilde{c}_1)/p_1 = 1/\epsilon_1$.

above results hold, given that the optimal solution is interior. Rochet and Tirole assume that the (quasi) demand functions are log concave, an assumption which guarantees an interior solution. As noted above, the "extended" Learner index (or price-cost margin, including the profit margin generated on the other side) will be higher on the less elastic market, in accordance with intuition.[18]

A possibility is of course that the second-order conditions are violated for "natural" demand functions. For example, it is easy to show that this is the case for constant-elastic demand functions $n_1 = 10p^{-1}$ and $n_2 = 10p^{-2}$, with $c = c_1 + c_2 = 1$. Bolt and Tieman, 2004, argue in this direction and show that for constant-elasticity demand, both a monopolist and a social planner will choose corner solutions, with higher prices on the more elastic side of the market.

In the following, however, it is assumed that the first-order conditions characterize the optimum, as they will for log-concave quasi-demand functions (see Rochet and Tirole, 2003). In particular, linear demand functions are log-concave. Adding the respective first-order conditions of the two sellers in a bilateral monopoly, eqs. (16), we find that $p_1 + p_2 - c_1 - c_2 = -(\phi_1'(n_1)n_1 + \phi_2'(n_2)n_2)$. In other words, for given values of n_1 and n_2 the combined mark-up in a bilateral monopoly will be exactly twice as large as that in a proprietary system. However, n_1 and n_2 will in general *not* be the same under the two market configurations. For example, if ϕ_i' is constant for $i = 1, 2$ and if prices are lower under a proprietary system, then n_i will be higher. Consequently, the terms $-\phi_i'(n_i)n_i$ will increase relative to the bilateral monopoly situation, implying that the sum of the price-cost margins under a proprietary system will be greater than one half of the sum of the price-cost margins under a bilateral monopoly.

> *Proposition 2.* With linear demand curves, the combined price-cost margins of the two sides under a proprietary system will equal 2/3 of the combined price-cost margins under bilateral monopoly.
>
> *Proof.*
>
> Let the inverse demand functions ϕ_i be linear, such that $p_i = \phi_i(n_i) = a_i - b_i n_i$. Straight-forward calculations will show that prices under a proprietary system will be given by $p_1 = (2a_1 - b_2 + c_1 + c_2)/3$ and $p_2 = (2a_2 - a_1 + c_1 + c_2)/3$, while under bilateral monopoly the prices will be $p_1 = (a_1 + c_1 - \mathbf{a})/2$ and $p_2 = (a_2 + c_2 + \mathbf{a})/2$, with

[18] Rochet and Tirole provide no intuition as to why prices will be higher on the most elastic market. For the linear case, an intuitive explanation is the following.

Assume that at prices equal to marginal costs on the two sides, the demand elasticity is lower on market A and higher on market B. In accordance with standard intution, prices will be raised the most on the less elastic market, i.e., market A. Higher prices on market A has two effects: as one slides up the (linear) demand curve, the elasticity increases, while the effective marginal cost on the other side, side B, falls. Since prices will be higher on market A, the effective marginal cost will be lower on market B. Therefore, at the optimum, the price in market A will have increased to a point where elasticity is actually *higher* than in market B.

In other words, sliding upwards along the (initially inelastic) demand curve, demand on side A will become more elastic. This is very similar to the observation that a monopoly sets prices on the elastic segment of the demand curve, while subsidies drives prices down to the inelastic segment of the demand curve.

the interchange fee, **a**, written in bold to avoid confusion. Hence, the combined price-cost margin under a proprietary system will be

$$p_1 + p_2 - c_1 - c_2 = \frac{a_1 + a_2 - c_1 - c_2}{3}$$

while the combined price-cost margin under bilateral monopoly will be

$$p_1 + p_2 - c_1 - c_2 = \frac{a_1 + a_2 - c_1 - c_2}{2}$$

The former price-cost margin is 2/3 of the latter. $\Diamond$

It is evident that total profits will be at least as high under a proprietary system as under bilateral monopoly. A proprietary system can replicate the pricing structure of a bilateral monopoly, but it can also avoid double marginalization.

8 One-sided monopoly

An alternative assumption is that there is market power on *one* of the two sides of the market - e.g., on the issuing side - but not on the other. I assume that the acquiring side of the market is competitive, while the issuing side is controlled by a monopolist[19], but that the interchange fee is again restricted to be zero. Now the issuer will want to maximize the following profit function:

$$\pi_1 = [\phi_1(n_1) - c_1]n_1 n_2 \tag{29}$$

As under bilateral monopoly, the issuing monopoly will consider the number of merchants, n_2, as fixed (as long as an interchange fee is not available; see below). Hence, using $p_1 = \phi_1(n_1)$, the first-order condition for profit maximization is

$$p_1 = c_1 - \phi_1'(n_1)n_1 \tag{30}$$

Note that the first-order condition is identical to that of the issuer under bilateral monopoly when a is restricted to be 0. It follows that welfare is higher than under bilateral monopoly (and zero interchange fee), but lower than under competition (and no interchange fee).

Note also that the mark-up on the issuer side now equates revenue loss on inframarginal units from increasing sales with one unit, while in a proprietary system the sum of the mark-ups on the two sides of the market were optimally set equal to the loss on inframarginal units from increasing sales one unit (on one side of the market).

[19] Rochet and Tirole (2003) argue that the acquiring side is more competitive than the issuing side. In contrast to the present paper, they assume that the issuing side is characterized by a symmetric oligopoly. My assumption of a monopolized issuing market could perhaps be justified on the grounds that bank customers - at least non-business debit-card customers - are captive. I.e., bank customers will not switch bank in order to get a better price on card transactions.

If the issuer holds monopoly power vis-à-vis the cardholders, while the acquiring side is competitive, there appears to be good reasons to think that the issuer holds monopoly power also with respect to the acquirers. If this is the case, it can set an interchange fee that will determine the price on the acquiring side of the market; given that the acquires are competitive, they will set $p_2 = c_2 + a$. Since the interchange fee is paid to a monopolist, however, it will have no effect on the issuing side of the market. Therefore, the one-sided monopoly will, arguably, have the same power to set two prices as a proprietary monopolist has (unless multilateral agreements face stronger regulatory resistance). Accordingly, the issuer will maximize eq. (22) and the price structure will be identical to that of the proprietary system.

9 Welfare comparisons

So far, I have analysed price-setting under six different regimes: perfect competition without an interchange fee, perfect competition with a socially optimal interchange fee, the first-best solution, bilateral monopoly (with and without an interchange fees), one-sided monopoly (without an interchange fee) and a single monopoly/proprietary system. As argued above, a one-sided monopoly which can also determine the interchange fee will be able to replicate the outcome chosen by the proprietary monopoly. Consequently, there is no need to analyse this case separately.

> *Proposition 3.* Given that the optimum is characterized by the first-order conditions, the welfare effects of the ownership structure of the card industry will be such that:
>
> a) Welfare under socially optimal prices is higher than welfare under competition, also for a socially optimal interchange fee.
>
> b) Welfare under competition and with an optimal interchange fee is higher than under either of the following: a proprietary system or competition without an interchange fee.
>
> c) Welfare under competition and without an interchange fee could be either higher or lower than that under a proprietary system.
>
> d) Assume linear demand functions. If so, welfare is lower under bilateral monopoly, where the two firms set the interchange fee so as to maximize combined profits on the two sides of the market, than under a proprietary system.
>
> e) In the absence of interchange fees, welfare is higher under a one-sided monopoly than under bilateral monopoly, and higher still under competition.
>
> f) Given the existence of a monopoly on one side of the market and perfect competition on the other side, introducing an access fee can either increase or decrease welfare.

Proof.

a) From eq. (7) we know that the optimal issuing price is $p_1^{opt} = c_1 - (\phi_2(n_2)/n_2 - c_2)$. Under competition, the issuing price will be $p_1 = c_1 - a$. Using eq. (13), the issuing price with a socially optimal interchange fee will be

$$p_1^{int} = c_1 - \omega(\frac{\phi_2(n_2)}{n_2} - c_2) + (1-\omega)(\frac{\phi_1(n_1)}{n_1} - c_1) \qquad (31)$$

where

$$\begin{aligned} \omega &= \frac{n_2 \frac{dn_1}{dp_1}}{n_2 \frac{dn_1}{dp_1} + n_1 \frac{dn_2}{dp_2}} \qquad (32) \\ 1-\omega &= \frac{n_1 \frac{dn_2}{dp_2}}{n_2 \frac{dn_1}{dp_1} + n_1 \frac{dn_2}{dp_2}} \end{aligned}$$

Given that the card system generates welfare, the expressions $(\phi_2(n_2)/n_2 - c_2)$ and $(\phi_1(n_1)/n_1 - c_1)$ are both positive. From eq. (32) it is obvious that ω and $(1-\omega)$ are positive. Let the values of n_1 and n_2 corresponding to the prices p_1^{opt} and p_1^{int} be given by p_i^{opt} and p_i^{int}, respectively. Assuming that $\phi_i(n_i) < 0$ and subtracting p_1^{opt} from p_1^{int} yields

$$p_1^{int} - p_1^{opt} = (1-\omega)[(\frac{\phi_1(n_1^{int})}{n_1^{int}} - c_1) + (\frac{\phi_2(n_2^{int})}{n_2^{int}} - c_2)] + [\frac{\phi_2(n_2^{opt})}{n_2^{opt}} - \frac{\phi_2(n_2^{int})}{n_2^{int}}]$$

The first term on the right-hand side is positive, while the second term is negative if $p_1^{int} - p_1^{opt} > 0$ and positive if $p_1^{int} - p_1^{opt} < 0$. Assume that $p_1^{int} - p_1^{opt} < 0$. Then both terms on the right-hand side are positive. This is a contradiction; it follows that $p_1^{int} - p_1^{opt} > 0$. Since the price under competition differs from the optimal price, it follows that welfare is lower.

b) Comparing a proprietary system with a competitive system with an optimal interchange fee, we know that in the former, $p_1 + p_2 = c_1 + c_2 - \phi_1'(n_1)n_1 = c_1 + c_2 - \phi_2'(n_2)n_2$, while in the latter $p_1 + p_2 = c_1 + c_2$. That is, the total price-cost margin will be higher in a proprietary system. Since the optimal interchange fee is set so that welfare is maximized given that p_1 and p_2 sum to $c_1 + c_2$, welfare cannot be higher when p_1 and p_2 sum to something larger than that. It follows that welfare is higher under an optimal interchange fee than under a proprietary system (or one-sided monopoly with competition on the other side, where the monopoly can use an interchange fee).

Comparing a competitive system with or without an (optimally set) interchange fee, it is obvious that a system without an interchange fee will result in lower welfare, except when the optimal interchange fee is zero.

c) From b) and from the fact that the optimal interchange fee can be zero, it follows that welfare can be higher under competition without an interchange fee than for a proprietary system. To demonstrate the possibility that the opposite claim can be true, it is sufficient to consider a market where the marginal valuation of cardholders is constant and ε below c_1, where ε is a value close to zero, while the marginal valuation of merchants is downward sloping and higher than c_2 for some merchants. In the absence of an interchange fee and under competition, there will be no card system, while a proprietary system will find it profitable to subsidize cardholders with ε per transaction.

d) Using the same notation as in the proof of *Proposition 2* and from the first-order conditions for a proprietary firm's profit-maximization problem, it follows that $n_i = (a_1 + a_2 - c_1 - c_2)/(3b_i)$. If instead the linear demand functions are inserted in the constrained maximization problem of the bilateral monopolists, eq. (**??**), the resulting system of three equations can be solved to yield $n_i = (a_1 + a_2 - c_1 - c_2)/(4b_i)$. Since quantities are sub-optimal and since welfare is strictly increasing in quantities below the optimum, it follows that welfare is higher under a proprietary system.[20]

e) Under competition, the first-order conditions that determine prices and quantities will be $p_1 = c_1$ and $p_2 = c_2$. From eqs. (30) and (16) the corresponding first-order conditions for one-sided and bilateral monopolies will be $p_1 = c_1 - \phi_1'(n_1)n_1$ for the issuing side and $p_2 = c_2$ and $p_2 = c_2 - \phi_2'(n_2)n_2$, respectively, for the acquiring side. From eq. (7) we know that the optimal price is lower than c_i, but since $\phi_i'(n_i) < 0$ we see that monopolies will set prices higher than c_i. It follows that welfare is lower under a one-sided monopoly than under competition, and lower still in a bilateral monopoly.

f) Using the linear demand system representations $p_1 = a_1 - b_1 n_1$ and $p_2 = a_2 - b_2 n_2$ it can be shown that welfare under a one sided monopoly without an interchange fee will be

$$W = n_1 n_2 [a_1 - \frac{1}{2} b_1 n_1 - c_1 + a_2 - \frac{1}{2} b_2 n_2 - c_2] \qquad (33)$$

Inserting the values for n_1 and n_s, which can be calculated from the first-order conditions, gives

$$W_{1-sided} = \frac{1}{8b_1 b_2}\left(a_1 - c_1\right)\left(a_2 - c_2\right)\left(3a_1 + 2a_2 - 3c_1 - 2c_2\right) \qquad (34)$$

[20] A reasonable conjecture is that for all or almost all functional forms of the demand curves, welfare is higher for a proprietary system than for a bilateral monopoly.

while welfare under a proprietary system will be

$$W_{prop} = \frac{2}{27b_1b_2}(a_1 + a_2 - c_1 - c_2)^3 \quad (35)$$

It can easily be demonstrated that neither of the two regimes dominates the other. For example, with parameter values $a_1 = a_2 = 10$, $b_1 = b_2 = 1$ and $c_1 = c_2 = 2$, a one-sided monopoly is preferable to a proprietary system. On the other hand, if $a_2 = 3$ and the other parameters are unchanged, welfare will be higher under a proprietary system.$\diamondsuit$

According to the proposition, there is no unique ranking between a proprietary system and competition (without an interchange fee). If the two sides of the markets are perfectly symmetric - so that the optimal level of the interchange fee is $a = 0$ - then a competitive market will perform just as well as a market with an optimal interchange fee and competition. It follows that a competitive market (without an interchange fee) outperforms a proprietary system in a symmetric setting.

On the other hand, if the two sides of the market are highly asymmetric, then welfare will be higher if one side of the market subsidizes the other. That means that a proprietary system (with an explicit or implicit interchange fee) may be preferable to a competitive system (without an interchange fee). One example would be a situation where demand on the issuing side of the market, say, is such that all potential cardholders have a valuation that is marginally below the issuers' cost, while merchants have a relatively inelastic demand for card services, such that their average valuation is much higher than acquires' marginal costs. Under perfect competition, price would equal marginal costs and, hence, no one would become a cardholder. However, it would be in a monopolist's interest to set the interchange fee such that most or all potential cardholders would find it in their interest to become actual cardholders. This would generate a large surplus on the *acquiring* side of the market. Even though a monopolist would be able to extract a relatively large share of that surplus, welfare would still be higher than if there were no card transactions at all.

In contrast, if an interchange fee were not available, there *would* be a unique ranking that puts competition ahead of a one-sided monopoly, and the latter ahead of a bilateral monopoly (at least for linear demand). Without an interchange fee, firms with market power could not cross-subsidize between the two sides of the market, but they would still set prices with a positive mark-up above costs.

10 Oligopoly markets

In the previous sections, the polar cases of monopoly and perfect competition have been analysed. Although these cases provide basic intuition on the functioning of two-way network markets, it may also be desirable to allow for intermediate levels of market power. There appears to be at least three segments of the market where an assumption of either perfect competition or monopoly may be too simplistic. First, there can be oligopoly interaction between banks, either between issuers or between acquirers (or both). Second, the merchants can have market power vis-à-vis their customers. Finally, more than one card system may compete for customers - so-called platform competition. Literature relating to these three alternatives will be briefly surveyed below.

Merchant marker power

In the literature that analyses the effect of merchant market power, a typical assumption is that merchants have market power (a positive price-cost margin) in relation to their customers, while being price takers in relation to acquiring banks. As mentioned in the introduction, the result may be overprovision of payment cards. The reason is that, due to a business-stealing effect, the merchants' *private* valuation of card services will exceed the social value and that acquirers will be able to extract part of that value, which may then be used to induce (excessive) card adoption on the issuing side.

More specifically, the above mechanism will be effective if at least some consumers are aware of which merchants that accept cards and are influenced in their choice of merchant by this aspect of service quality. Alternatively, some consumers may search until they find a merchant that accepts cards. Then, a merchant will have a *strategic* interest in accepting card payments, in addition to the intrinsic benefits of receiving card payments rather than cash. Accepting card payments will increase the merchant's sales, at the expense of other merchants. This strategic interest can be exploited by acquiring banks with market power, or by issuing banks through an interchange fee, as show by Rochet and Tirole (2002). The acquirers will be able to charge a higher price than otherwise. However, the gains the merchants will be able to make at the expense of one another will not correspond to net welfare gains. The additional surplus the acquirers are able to extract will, therefore, lead to higher consumer prices. In effect, assuming that the merchants cannot price discriminate between card customers and non-card customers, the latter category of consumers will be subsidizing the former.[21] The welfare effect of this subsidy, however, can be either positive or negative. Because of network effects, it may be welfare-increasing to subsidize cardholders, but the subsidies may also be excessive.

Rochet and Tirole (2002), Gans and King (2003) and Wright (2004) all find that under certain conditions, the profit-maximizing interchange fee may deviate from the optimal one, while under other conditions they may coincide. (For a discussion of these results, see Chakravorti, 2003, Section 3.2.)

Platform competition

Although a discussion of platform competition falls outside the scope of this paper, a few references will be provided. Rochet and Tirole (2003) provide a general framework for analysing platform competition. Guthrie and Wright (2003) find that competition between platforms is ineffective, unless cardholders

[21] Rochet and Tirole derive this result under the assumption of price competition in differentiated products. Under the assumption of price competition in homogenous products, non-card customers would go to merchants that do not accept cards, while card customers would patronize merchants that do.

hold just one card, while Chakravorti and Roson (2004) find that competition unambiguously increases consumer and merchant welfare.

Oligopolistic banks

The analysis becomes less tractable under oligopoly interaction. An intuitively plausible result is that the interchange fee will be used to transfer profit from the more competitive side of the market to the less competitive side. This is formally shown by Manenti and Somma (2003), in a setting with one not-for-profit platform and one proprietary system. They assume that the not-for-profit platform maximizes the sum of issuers' and acquirers' profits and that competition between banks is such that there will be a fixed mark-up above issuers' and acquirers' marginal costs. For example, given issuers' and acquirers marginal costs c_1 and c_2, prices will be αc_1 on the issuing side and βc_2 on the acquiring side, where $\alpha, \beta > 1$. With an interchange fee a, equilibrium prices will instead be $\alpha(c_1 - a)$ and $\beta(c_2 + a)$. A more surprising result is that the banks' total profit will not depend on α or β. In particular, if competition increases on one side of the market, i.e., if α or β falls, then the interchange fee will be rebalanced so that total profit remains unchanged.

Assuming fixed mark-ups can be seen as a reduced-form analysis. In the following, the behaviour of the oligopoly firms will instead be modelled explicitly. In order to simplify the analysis, the acquiring side of the market is assumed to be competitive and the merchants are assumed to have no strategic reasons for accepting cards (the "Baxter case"). With the exception that there are now two issuers that compete in quantities, the set-up is the same as in the previous sections. Demand is given by eqs. (1) and (2), but with the sum of the two firms' quantities as the argument in ϕ_1. The two issuers are assumed to have identical and constant marginal costs, equal to c_1. It follows that the profit function of the issuers will be

$$\pi_{1i} = [\phi_1(n_{11} + n_{12}) - c_1 + a]n_{1i}n_2 \tag{36}$$

where n_{1i} now denotes the number of card customers that bank $i = 1, 2$ has, such that $n_1 = n_{12} + n_{12}$.[22] The corresponding first-order condition for issuer i is

$$\phi_1'(n_{11} + n_{12})n_{1i} + \phi_1(n_{11} + n_{12}) - c_1 + a = 0 \tag{37}$$

Exploiting the fact that the two issuers are symmetric, we have that

$$\phi_1'(2n_{11})n_{11} + \phi_1(2n_{11}) - c_1 + a = 0 \tag{38}$$

In order to proceed, a more specific model will be used in the following. Assume now that demand is linear and given by inverse demand function

$$p_1 = L - n_{11} - n_{12} \tag{39}$$

[22] Note that when there are two subindices, the first index refers to market or "side", while the second index refers to firm (on the issuing side). If there is only one index, it refers to market.

on the issuing side of the market, while (inverse) demand on the competitive acquiring side of the market is given by

$$p_2 = M - n_2 \tag{40}$$

where n_2, as before, is the total number of customers (merchants) of all acquiring banks. As before, the marginal costs are denoted c_1 and c_2. It is straight-forward to show that the equilibrium quantities, for a given level of a, will be $n_{11} = n_{12} = (L - c_1 + a)/3$, which will result in $p_1 = L/3 + 2(c_1 - a)/3$. This, in turn, will yield profits of

$$\pi_{11} = \pi_{12} = ((L - c_1 + a)/3)^2 n_2 = ((L - c_1 + a)/3)^2 (M - c_2 - a) \tag{41}$$

where the last equality comes from the merchants' inverse demand function and from the assumption of perfect competition between the acquirers. For the particular case of $a = 0$, prices will be $p_1 = \frac{1}{3}(L + 2c_1)$ and $p_2 = c_2$.

If the duopolists can jointly set the interchange fee, they will set it so as to maximize the above expression, i.e., such that

$$a = \frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 - \frac{2}{3}c_2 = \frac{1}{3}(2M - L - c_1 - 2c_2) \tag{42}$$

This will result in prices

$$\begin{aligned}
p_1 &= L/3 + 2(c_1 - \left(\frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 - \frac{2}{3}c_2\right))/3 = && (43)\\
&= \frac{5}{9}L - \frac{4}{9}M + \frac{4}{9}c_1 + \frac{4}{9}c_2 && (44)\\
p_2 &= c_2 + \frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 - \frac{2}{3}c_2 = &&\\
&= \frac{2}{3}M - \frac{1}{3}L + \frac{1}{3}c_1 + \frac{1}{3}c_2 && (45)
\end{aligned}$$

and quantities

$$\begin{aligned}
n_1 &= 2n_{11} = 2n_{12} = \frac{4}{9}(L + M - c_1 - c_2) && (46)\\
n_2 &= \frac{1}{3}(L + M - c_1 - c_2) &&
\end{aligned}$$

To simplify further, assume that $M = L$ and that $c = c_1 = c_2$. Then

$$\begin{aligned}
a &= \frac{1}{3}(L - c) && (47)\\
p_1 &= \frac{1}{9}(L + 8c) &&\\
p_2 &= \frac{1}{3}(L + 2c) &&
\end{aligned}$$

This means that, in a market with competitive acquisition and an issuing duopoly, the two-firm Cournot price will be set in the *acquiring* market, while price in the issuing market will be identical to the *eight*-firm Cournot price.

In the general (linear) case, corresponding to prices given by eq. (43) and interchange fee given by eq. (42), prices will fall in the issuing market and rise in the acquiring market if $(L-c_1) < 2(M-c_2)$, while if $(L-c_1) > 2(M-c_2)$ prices will *rise* in the issuing market and *fall* in the acquiring market (relative to a situation with $a = 0$). The interpretation is that if approximately equal amounts of consumer surplus are generated in both markets, the duopoly issuers will set the interchange fee so as to extract profit from the acquiring market. This will lower the marginal cost in the issuing market. Hence, as a side-effect, prices will fall on that market. However, if much more consumer surplus is generated in the issuing market, then the duopolists should in fact set the interchange fee so as to *subsidize* the acquiring market (and tax themselves). This will reduce the merchant discount and increase the number of merchants. This, in turn, increases the number of transactions per cardholder, with the ultimate effect of increasing the issuers' revenues.

The welfare effect of the interchange fee can be evaluated by comparing welfare when the interchange fee is set according to eq. (42) and welfare when the interchange fee is set to zero, as the following proposition shows.

> *Proposition 4.* Welfare under a one-sided duopoly, where the firms compete in quantities, where there is competition on the other side and where the duopolists can set the interchange fee so as to maximize their combined profits, can be either higher or lower than welfare under the same market structure with the interchange fee restricted to zero.
>
> *Proof.*
>
> Simple calculations show that when the interchange is fixed at zero, quantities will be given by
>
> $$\begin{aligned} n_1 &= \frac{2}{3}(L-c_1) \\ n_2 &= M - c_2 \end{aligned} \tag{48}$$
>
> i.e., Cournot quantities in the issuing market and the competitive quantity in the acquiring market. Welfare for the two cases can be calculated using eq. (33), but with L taking the place of a_1, M replacing a_2 and $b_1 = b_2 = 1$. Inserting quantities, given by eqs. (46) and (48), respectively, into eq. (33) gives
>
> $$\begin{aligned} W^{IF} &= \frac{22}{243}(L+M-c_1-c_2)^3 \\ W^0 &= \frac{1}{9}(L-c_1)(M-c_2)(4L+3M-4c_1-3c_2) \end{aligned} \tag{49}$$
>
> where W^{IF} denotes welfare when the duopolists set a profit-maximizing interchange fee, according to eq. (42) and W^0 denotes welfare when

there is no (or zero) interchange fee. It is easily shown that neither of the two cases dominates the other, for example by using the numerical example from Section 9 (see the proof of Proposition 3f).◊

In line with intuition, when the two sides of the markets are symmetric, welfare is higher if the interchange fee is fixed at zero, while if demand is much smaller on the acquiring side, a profit-maximizing interchange fee will increase welfare.

A third regulatory strategy, in addition to allowing the profit maximizing interchange fee or to set the interchange fee to zero, is to cap the interchange fee at $c_1 + c_2$. If this constraint is not binding, the firms can be assumed to set the profit maximizing interchange fee, but for some parameter values the cap will be effective and will hold the interchange fee down. Some numerical calculations suggest that welfare may be improved from imposing such a cap.

11 Conclusions

One ambition of this study was to develop a simple bench-mark model, with which the welfare consequences of different market structures - with or without an interchange fee - can be analysed. The analysis suggests that network effects in two-sided payment markets will result in quantities being too small (prices too high), also when the banks act as price takers. This is so, since the buyers do not (cannot) incorporate the positive external effect of buying another unit, while the sellers do not (cannot) incorporate the positive external effect of reducing the price. If the banks have market power, matters may be even worse - i.e., prices may then be even further from the social optimum.

In competitive markets, an interchange fee can potentially be used to improve welfare. However, the banks will not have incentives to set the interchange fee at any particular level, since they will make zero profit regardless of its level. For this reason, the interchange fee cannot be used to *create* market power. In contrast, if the banks already have market power, they will in general have an interest in setting a non-zero interchange fee. Issuing banks will want to set the interchange fee so as to extract profit from the acquiring side of the market, and vice versa. However, banks with market power will also have incentives to set the interchange fee with an eye to system-wide network effects. Therefore, if there is a monopoly on one side of the market, the welfare effect of introducing an interchange fee is indeterminate, given that it is set at the profit-maximizing level. The reason is not that there will be over-provision of card services (Cf. the results of Rochet and Tirole, 2002, referred to in Section 10), but that the market power on, say, the issuing side will spill over to the acquiring side of the market.

From a policy perspective, there appears to be (at least) three possible regulatory strategies. First, the banks could be prohibited from using interchange fees. (Or, equivalently, the interchange fee can be restricted to be zero.) Second, the banks could be free to set interchange fees at the profit-maximizing level. Finally and third, the regulatory authorities (the bank regulator or the competition authorities) could try to approximate the optimal fee structure by allowing interchange fees at some intermediate level. The results presented in this paper, in particular *Proposition 3*, shows that in a comparison between the first and the second alternative, neither is welfare dominant. Trivially, an optimal interchange fee in combination with perfect competition will always be preferrable to less competitive market structures and to non-optimal fees (including a fee of zero). However, setting an optimal fee is a non-trivial task.

The European Commission's competition directorate has opted for a version of the third strategy, when it decided to allow positive interchange fees - but only such fees as could be justified by costs. One way to interpret the decision is that the Commission accepted interchange fees that were no higher than the issuers' costs (corresponding perhaps most closely to $c_1 + c_2$ in this article).[23] Indirectly, a similar regulatory regime has been established in the US, although this is the result of an out-of-court settlement following private litigation.[24]

From a welfare point-of-view, the optimal interchange fee has no simple direct relation to the banks' cost structure, although the EU Commission in its decision has linked the two. However, we do know that the *sum* of prices chosen by profit-maximizing firms will tend to be too high, assuming no fixed costs and constant marginal costs. In such a setting, the sum of prices in the two markets should be below total marginal (or average variable) costs. We also know that in order to achieve optimal network effects across the two markets, it will typically be necessary that one side of the market subsidizes the other.

The regulatory policy adopted by the EU Commission appears to strice a reasonable balance between the two possible sources of inefficiences: prices above costs and lack of efficient cross subsidies. While allowing quite substantial cross subsidies, it has the potential of limiting the extent to which market power can be exerted between the two sides of the market. Some numerical calculations, not reported in the paper, suggests that capping the interchange fee at the level of marginal costs may be welfare improving, relative to a policy of non-intervention. Of course, the interchange fee can still be excessive at this level, but it appears that finding the optimal level would require a much more sophisticated modelling approach, perhaps one that attempted to simulate the actual market characteristics.

12 References

Armstrong, Mark 2004, *Competition in Two-Sided Markets*, mimeo, University College London.

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

[23] See the EU Commissions decision 24.07.2002 in the *Visa International* case, COMP/29.373, published in the Official Journal L 318, 22.11.2002, p. 17-36 .

[24] See the Economist, May 3, 2003, p. 66-67.

Bolt, Wilko and Alexander F. Tieman, 2003, *Pricing Payment Services: An IO Approach*, Working Paper No. 202, International Monetary Fund.

Bolt, Wilko and Alexander F. Tieman, 2004, *Skewed Pricing in Two-Sided Markets*, mimeo, De Nederlandsche Bank and IMF.

Chakravorti, Sujit, 2003, Theory of Credit Card Networks: A Survey of the Literature, *Review of Network Economics*, 2, 50-68.

Chakravorti, Sujit and Roberto Roson, 2004, *Platform Competition in Two-Sided Markets: The Case of Payment Networks*, mimeo, Federal Reserve Bank of Chicago and Department of Economics, Universita' Ca' Foscari di Venezia.

Gans, J.S. and S.P. King, 2003, The Neutrality of Interchange Fees in the Payment Systems, *Topics in Economic Analysis & Politics*, 3, 1, Article 1. (See http://www.bepress.com.)

Guthrie, Graeme and Julian Wright, 2003, *Competing Payment Schemes*, Working Paper No. 245, University of Auckland.

Hausman, Jerry A., Gregory K. Leonard and Jean Tirole, 2003, On Nonexclusive Membership in Competing Joint Ventures, *Rand Journal of Economics*, 34, 43-62.

Katz, Michael, 2001, *Reform of Credit Card Schemes in Australia II*, Reserve Bank of Australia, Sydney, Australia.

Manenti and Somma, 2003, *Plastic Clashes: Competition among Closed and Open Systems in the Credit Card Industry*, mimeo, Università di Padova and Università degli Studi di Bari.

Rochet, Jean-Charles and Jean Tirole, 2002, Cooperation Among Competitors: Some Economics of Payment Card Associations, *Rand Journal of Economics*, 33, 549-570.

Rochet and Tirole, 2003, Platform Competition in Two-Sided Markets, *Journal of the European Economic Association*, 1, 990-1029.

Schmalensee, Richard, 2002, Payment Systems and Interchange Fees, *Journal of Industrial Economics*, 50, 103-122.

Wright, Julian, 2004, The Determinants of Optimal Interchange Fees in Payment Systems, *Journal of Industrial Economics*, 52, 1-26.

Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se