Approaches to Regulating Interchange Fees in Payment Systems

JOSHUA S. GANS*

Melbourne Business School, University of Melbourne

STEPHEN P. KING

Melbourne Business School, University of Melbourne

Abstract

Significant attention worldwide has been paid to the regulation of credit card interchange fees. In part, this attention has followed concerns expressed about the level of these fees in Europe, the U.S. and Australia. The Reserve Bank of Australia recently conducted an extensive inquiry into the interchange fees associated with credit cards and has moved to regulate those fees. At the same time, research economists have considered determinants of the socially optimal interchange fee. In this paper, we use the Australian experience to highlight alternative methods of regulating interchange fees in payments systems. We use a simple model to derive a socially optimal interchange fee when merchants cannot freely set different prices for different payment instruments. We compare the socially optimal interchange fee from this model with those presented in the economics literature and show that most analyses capture a simple externality within the optimal fee. Credit card usage for a specific transaction is determined by the customer. But the customer does not bear the costs or receive the benefits that card usage imposes on the merchant. The optimal interchange fee internalises this externality. We then compare the theoretical optimal interchange fee with the approaches proposed in Australia, and show that the regulatory approach adopted by the Reserve Bank of Australia may be viewed as economically conservative in certain situations. Finally, we consider additional issues that will impinge on the regulation of interchange fees.

1 Introduction

In a relatively short period of time, credit card associations have moved from being largely unregulated, albeit subject to antitrust scrutiny, to being part of an industry that faces far reaching and intrusive price regulation. Leading the regulatory charge has been the Reserve Bank of Australia (RBA). In the 1990s, the RBA was given new powers by the Federal Australian government to regulate standards in payment systems, including technical and entry standards as well as some pricing powers. In 2002, it used these powers to eliminate prohibitions on surcharging that were imposed by the three major credit card

* Contact author. Mailing address: Melbourne Business School, 200 Leicester Street, Carlton Victoria 3053 Australia. E-mail: j.gans@unimelb.edu.au Parts of this paper were previously circulated under the title "Regulating Interchange Fees in Payment Systems." We appreciate the helpful comments of an anonymous referee and Julian Wright, the editor, on an earlier version of this paper. All errors are our own.

associations in Australia – MasterCard, Visa and Bankcard – and to specify a methodology that will regulate interchange fees in those associations from late 2003.[1]

Not surprisingly, this move towards regulation has provoked a response from the MasterCard and Visa associations that question the legitimacy and application of the RBA's approach. In particular, they question whether the RBA made an economic case for regulation and for its proposed methodology for setting interchange fees.

While the domestic Australian regulations apply to a relatively small part of Visa's and MasterCard's operations, the RBA decision has potential international ramifications. For example, regulators in Europe and Israel are currently considering the potential regulation of interchange fees and the RBA approach is likely to influence any regulatory approach adopted by overseas authorities.

For this reason, we believe that it timely to review the alternative approaches for the setting of interchange fees. While there has been much research in recent times as to whether interchange fees should be regulated, less attention has been paid to the methodology that would be appropriate to use if it is determined that card associations should be regulated. There is considerable divergence between the methodologies that have arisen informally within the context of the Australian debate. We review these informal methodologies in Section 2. Nonetheless, as we will demonstrate in this paper, there is a substantial level of implicit agreement about the form of the socially optimal interchange fee embedded within formal economic models of payment systems. We provide our own simple model of payment systems in Section 3 and show how this relates to existing models (Section 4).

Finally, in Section 5, we consider some of the important practical issues that need to be considered in regulating interchange fees and how they impact upon the benchmark that might be chosen by regulatory authorities.

2 Suggested approaches to interchange fee regulation

In the course of the Australian debate several approaches were put forward as to how interchange fees might be set. Those approaches are (i) the MasterCard approach; (ii) the balancing approach; (iii) the Frontier Economics approach; and (iv) the ABA approach. In the end, the RBA adopted its own approach – the RBA approach. In this section, we review these alternatives.

First, however, it is useful to specify a common notation to describe the various cost and revenue components. Let c_I, C_I and R_I be the marginal costs, average fixed costs and average revenue of issuers while c_A, C_A and R_A are those for acquirers. Finally, let a denote the interchange fee itself. We follow convention by defining a positive interchange fee as a payment from acquirers to issuers.

We now briefly summarise the alternative approaches.

2.1 The MasterCard approach

The RBA (2001) report that MasterCard use a methodology that views the interchange fee as a means by which card issuers can recover the cost of the services provided to acquirers

[1] Bankcard is a domestic Australian credit card that commenced operations in the 1970s. It is jointly owned by Australian banks.

and, through them, to merchants. The MasterCard approach involves setting $a = c_I^+$ where c_I is comprised of the costs of providing a payment guarantee (including fraud and credit write offs) and the costs of processing transactions from acquirers. The '+' refers to the addition of the cost of funding the interest free period. While some proportion of this might be part of c_I as there is a minimum practical interest free period, in reality, issuers also set the length of the interest free period. As such the costs associated with any discretionary interest free period could be viewed as a negative revenue for issuers. Notably, fixed costs and acquirer costs are not part of the MasterCard approach.

2.2 The balancing approach

The RBA/ACCC (2000) defined an alternative approach that was employed in practice for setting interchange fees.[2] This approach involved a comparison of the costs of issuing and acquiring with the revenues obtained by each side of the credit card market. The basic idea is that the interchange fee will offset any shortfall in revenue by either issuers or acquirers; thereby, restoring balance between these parties.[3]

What this appears to mean is that if $R_I < c_I + C_I$, then a would be set so that $a = c_I + C_I - R_I$. In this situation, if both $R_I > c_I + C_I$ and $R_A > c_A + C_A$, then $a = 0$ while if there was a shortfall in acquiring, then $a = R_A - c_A - C_A$. In this situation, the interchange fee would be negative. If there was a shortfall in both, the association would not be financially viable.

2.3 The frontier economics approach

The consulting firm, Frontier Economics (2001), were asked to come up with an interchange fee methodology that could accompany a proposal by member banks to authorise the behaviour of credit card associations and permit them to operate without further antitrust scrutiny from the Australian Competition and Consumer Commission (ACCC). Their approach was directed at one of the principal concerns raised by the RBA/ACCC (2000) joint study; namely, that issuers appeared to be earning excessive margins on credit card transactions.

The Frontier Economics approach was simply to eliminate these excessive margins. That is, they proposed to set $a = c_I + C_I - R_I$. In contrast to the balancing approach, this would only allow a reduction of interchange fees to ensure issuers were not earning excessive profits. However, acquirers would continue to earn any relevant economic profits. The Frontier Economics regulatory rule includes revenue elements as well as cost elements. In effect, issuers would be allowed a certain rate of return on their capital investments. As such, this approach could lead to the standard incentive problems that arise with rate-of-return style regulation. In practice, the Frontier Economics proposal could be applied as an initial cap with a declining price thereafter. However, in that case,

---

[2] For a detailed discussion of the approaches identified by the RBA/ACCC (2000), see Gans and King (2001).

[3] The balancing approach is similar to the approach adopted by Visa. According to Wright (2000), the interchange fee is set so that $(R_I + a - c_I - C_I)/(R_I + a) = (R_A - a - c_A - C_A)/(R_A - a)$ or $a = (R_A(c_I + C_I) - R_I(c_A + C_A))/(2(R_A - R_I) - (c_A + C_A - c_I - C_I))$. However, as Wright (2000) emphasises, the precise calculation requires resolution of a fixed point problem; something that is also an issue for the balancing and the Frontier Economics approaches. However, in contrast to the Visa approach, those approaches are in a "corner solution" form that makes comparison with other proposals easily. For this reason, we do not explicitly consider the Visa approach below.

while there might be pressure on issuers to reduce costs, that pressure would not be present for acquirers. Ultimately, the ACCC rejected this approach and it was not submitted to the RBA (2001) for its consideration.

2.4 The avoidable cost approach

The Australian Banking Association (ABA), on behalf of nine member banks in Australia, proposed an "avoidable cost" methodology to the RBA. That methodology was based on a simple question: "What costs would be unavoidable if an issuer were to provide (on a sustainable basis) only credit card payment services?" (ABA, 2001, p.51). The basic idea was that interchange fees should be capped at the stand-alone cost of providing payment functionality (in the absence of credit services) and should be no lower than the incremental cost of that functionality.

The ABA methodology included costs to issuers associated with fraud costs, credit losses from the interest free period, operating costs, marketing, promotion and retention costs (including loyalty program costs) and the cost of equity capital and sunk costs. In essence, it capped the interchange fee at $a = c_I^+ + C_I$. Thus, it potentially allowed all issuer costs associated with payment functionality to be passed through to acquirers and merchants.

In theory, the ABA cap considers the maximum interchange fee such that merchants just do not have an incentive to backwards integrate into card provision by issuing store cards. Of course, we would expect this constraint to exist in the market for card services regardless of any regulation. In this sense, the ABA maximum would not be a binding regulatory price.

2.5 The RBA approach

The RBA were dissatisfied with all of the above approaches. Interestingly, their main concern was not the focus on issuer costs but on which costs were included. First, they argued that only issuer costs associated with payment functionality should be included. Costs associated with say credit losses on unpaid credit card debt related to the line of credit and could be recovered through interest payments by "revolvers" who utilised those services.

Second, the interest free period was seen by the RBA (2001) to be a cost associated with providing a benefit purely to cardholders and, therefore, should be recovered by fees on cardholders. The RBA favoured the inclusion of costs only to the extent that they were associated with an unambiguous benefit to merchants. All of the above methodologies included these costs, but also included other costs such as the costs associated with the interest free period. This said, the RBA ultimately conceded that a minimum interest free period was a defining characteristic of credit cards as opposed to debit cards.

Third, the RBA criticised the avoidable cost approach (and implicitly the Frontier Economics approach) for the inclusion of loyalty points. These points are effectively a discount to cardholders. Being a price discount rather than a cost element, the RBA believed they should not be part of interchange fees as this would create an undue incentive for issuers to accelerate such programs at effectively no cost to the issuers.

Finally, the RBA did not believe that the cost of capital should be included in interchange fees. The RBA accepted an argument put forward by Gans and King that if interchange fees for issuers were reduced because their sunk investment costs were not taken into account, issuers would be able to recover those costs from other revenue sources (such as annual card fees). Essentially, all issuers receive a common interchange fee and so a change in that fee would, for the most part, not impact on their profits in competition with one another.

In the end, the RBA specified classes of issuer costs that it believed should be included in the interchange fee. These were the costs of processing card transactions, fraud and fraud prevention, authorisation costs and costs associated with funding interest-free periods. In effect, the RBA proposed that $a = c_I^-$. That is, the interchange fee was to be based on marginal issuing costs less costs associated with credit losses and costs associated with maintaining credit card accounts (hence the superscript minus).

The RBA's methodology identified issuer costs and allowed their inclusion if (i) they were incremental or marginal; (ii) they related to payment functionality rather than a line of credit; and (iii) they were attributable to a merchant benefit. Interestingly, they did end up deviating from this broad principle by including the costs of funding the interest free period; which may be for the most part a discretionary benefit to cardholders (and hence a revenue item) rather than a cost item. We will discuss the merits of this in Section 5 below when we address some practical issues.

2.6 Summary

In summary, all of the above methodologies focus on the recovery of issuer costs in one form or another. They differ as to whether they include issuer revenues to offset those costs, whether they include fixed issuer costs and also in the components that comprise marginal issuer costs. Nonetheless, the avoidable cost approach was by far the weakest cap with the RBA's approach being the most stringent if $R_I > C_I$.

We turn now to consider a formal economic model of a payment system. As will be seen, that model and also existing models in the literature do not support the inclusion of fixed issuer costs or issuer revenues in the interchange fee and, in general, suggest the inclusion of acquirer cost elements. Finally, no distinction between payment functionality and the line of credit is generated with the key criterion being whether merchants conceivably receive some benefit from the particular function.

3 Socially optimal interchange fees without network effects

We begin by providing a simple model of a payment system without network effects. One reason to do this is the suggestion of Katz (2001) that such effects are no longer relevant in mature payment systems.[4] Regardless, this approach provides a way to consider the determinants of a socially optimal interchange fee and a basis for the regulation of interchange fees. We will also be able to relate the interchange fee derived here to those contained in alternative and more elaborate models of payment systems (see Section 4).

Our model is of a "four-party" credit card association. In such systems, *issuers* encourage *customers* to use credit cards when purchasing goods and services from *merchants*. They can only do this if merchants themselves have agreed to process credit

[4] Network effects are built explicitly into a number of models; requiring some coordination between merchants and consumers in their card adoption decisions (see, most notably, Schmalensee, 2002). Our model here does not build those effects in but continues to have the "fundamental usage externality" (Rochet, 2003) common to most models in this literature.

card transactions on the terms given to them by *acquirers*. Thus, issuers and acquirers are joint suppliers of card services while customers and merchants are both consumers of them; potentially deriving value from such transactions in return for payments they might make to issuers and acquirers respectively. Finally, relations between issuers and acquirers are governed by interchange arrangements that set rules for who bears costs and risks arising from disputes or fraud and any payments between issuers and acquirers. In credit card associations such as MasterCard or Visa, interchange arrangements are determined collectively by all issuing and acquiring participants.

We begin by modeling the interaction between a representative customer and a representative merchant. The representative customer has demand curve $Q(p)$, takes merchant prices and bank fees as given, and seeks to minimise the total cost of their purchases. The customer can use cash or credit card or any combination to pay for total purchases. We consider the costs and benefits of credit card use relative to cash. Using the credit card involves an additional fixed charge of F_c and a fee of f per unit purchased. The banks that issue credit cards set these customer fees.

The customer can save transaction expenses by using a credit card rather than cash. We denote these savings by $\int_0^{Q^c} b_c(q)dq$ where Q^c is the customer's total credit card purchases and $b_c(.)$ is their marginal benefit from credit card usage. We assume that $b_c(0) > f$ in all relevant situations and $b_c' < 0$. In other words, if cash and credit card retail prices are identical, the customer has paid the fixed fee F_c and has a credit card, then it always pays the customer to make some credit purchases. However, the relative benefits of credit purchases over cash decline as the total amount of purchases rises. Let s be the difference between M's price if a credit card is used and its price if cash is used, where $s > 0$ if the credit price exceeds the cash price. If the customer makes both cash and credit card purchases, this implies that they will purchase on credit card until $s + f - b_c(Q^c) = 0$. To avoid trivial outcomes, we only consider situations where, once they pay the fixed charge, the customer chooses to make both cash and credit purchases.

The representative customer can be interpreted in two ways. First, consider the purchases of an individual. Often credit card purchases are "higher value" items. For example, a customer might use credit card for the weekly grocery shopping at a supermarket, but might use cash to purchase just milk or bread at the same supermarket. Our representative customer model captures this effect to the extent that credit card purchases are inframarginal while cash purchases are marginal. Alternatively, credit cards tend to be used more by higher income customers. Such customers will tend to have higher levels of willingness to pay for an item and this is captured in our framework.

The merchant may also receive benefits from credit card sales relative to cash sales. The marginal merchant benefit is denoted by $b_m(Q^c)$ where $b_m' \leq 0$. The merchant pays a fixed charge of F_m before they can accept any card transactions and then pay a merchant-services fee of m for each unit sold to a customer using a credit card. The bank that provides credit card acquiring services to the merchant sets these fees. Total profit for a merchant who accepts credit cards is given by

(1) $$\pi = Q^c(s-m) + \int_0^{Q^c} b_m(q)dq + pQ - c(Q) - F_m$$

where Q refers to total sales by both cash and credit and $c(.)$ is the merchant's cost function. We assume the standard restrictions on both $Q(p)$ and $c(Q)$ for a solution to the merchant's profit maximisation problem to be both well defined and unique.

Consideration of the fees for credit cards requires a formal model of issuer and acquirer banks. Issuing, acquiring or both functions could be characterised by imperfect competition. We assume, however, that both functions are characterised by competition in two-part pricing. In other words, issuing and acquiring banks compete by setting both fixed charges to their customers or merchants, F_c and F_m, and by setting per transaction fees f and m.[5] Merchants and customers will choose their banks according to the total benefit that they receive. Profit maximising behavior in such circumstances will lead banks to set transaction fees that reflect the true marginal cost of credit card transactions. In other words, banks have no incentive to distort marginal prices but rather seek to maximise profits by capturing surplus from merchants and customers.

For simplicity, suppose that the per-transaction cost to an issuing bank is constant and given by c_I while the per-transaction cost to an acquiring bank is constant and given by c_A. The fixed costs to issuers and acquirers can be set at zero without significantly affecting our analysis. There might also be an interchange fee between issuers and acquirers. We denote the per-transaction interchange fee by a and adopt the convention that this fee is paid by acquirers to issuers (although that fee may be positive or negative). Thus, competition between different issuers and acquirers will lead to credit card fees $f = c_I - a$ and $m = c_A + a$.

As banks compete through non-linear prices, with fees set at the relevant (constant) marginal costs for issuers and acquirers, all bank profit is generated by the fixed merchant and customer fees. These fees will be determined by issuer and acquirer competition. At this stage, we only need to make a relatively weak assumption about these fees. We assume that, regardless of the degree of competition in issuing and acquiring, the equilibrium fixed fees for the banks will not be set at a level greater than the total merchant and customer benefits.[6]

The timing of behavior in the market is as follows. First, banks set credit card charges. The merchant and the customer then independently decide whether to pay the fixed charge and avail themselves of the ability to make credit card transactions. The merchant then simultaneously sets both the cash and credit prices and the customer chooses both their total purchases and how to divide their purchases between each payment instrument.[7]

---

[5] We relax this assumption below and consider competition in linear fees.

[6] This is consistent with most models of issuer and acquirer competition. It simply rules out the possibility of an "odd" equilibrium where neither issuers nor acquirers make any profit because each sector is setting fixed charges above their respective customer's total benefits from card usage.

[7] The merchant and the customer engage in a co-ordination game when they choose whether or not to avail themselves of card facilities. The surplus to both parties from using the card can be positive, but neither party will wish to pay their fixed fee unless they believe that the other party is also going to pay the fee and avail themselves of card services. This "network effect" has been analysed by a number of other papers (e.g. Rochet and Tirole, 2002a; Schmalensee, 2002). However, we concentrate here on the case where both the merchant and the customer choose to avail themselves of card facilities. Given our assumption on bank competition, in the equilibrium of the subgame where both the customer and the merchant choose to avail themselves of card facilities, both parties will always receive non-negative surplus from card purchases. Hence, it is always a Nash equilibrium for both the merchant and the customer choosing to avail themselves

Finally, we require a social benchmark to evaluate the effects of the credit card system. Our concern here is with the efficiency of the payments system rather than the general economic surplus generated by all transactions. Consequently, we use transactions costs generated by the payments system as our point of evaluation.[8] The total transaction cost is given by:

$$T = \left(c_I Q^c - \int_0^{Q^c} b_c(q)dq \right) + \left(c_A Q^c - \int_0^{Q^c} b_m(q)dq \right).$$

This is minimised when the customer divides total purchases between cash and credit so that $c_I - b_c(Q^{c*}) + c_A - b_m(Q^{c*}) = 0$. If the consumer makes both cash and credit card purchases then the optimal split of total purchases between payment instruments occurs when the total marginal benefit of credit card purchases to both the customer and the merchant equals total marginal cost of a credit transaction to the banks. The socially optimal quantity of credit card transactions in this situation is Q^{c*}. This is the same condition as identified by Baxter (1983).

3.1 Frictionless surcharging

It is worthwhile beginning, as a point of illustration, with the case where surcharging for credit card use is both allowed and frictionless; that is, p and p^c can be separately set by merchants. In numerous countries (most recently, Australia) card associations are prevented from restricting any surcharging by merchants although transaction costs may still limit the practice. Here we suppose that surcharging is costless to achieve and see what this means for the socially optimal interchange fee.

We first analyse the retail market outcome. The merchant will set both the cash and credit card price and will seek to divide cash and credit sales to maximise profit. However, the merchants' desired split of total sales between cash and credit must be consistent with the customer's choice of payment instrument. Thus, the merchant will set p^c, Q^c, and p to maximise π subject to $s + f - b_c(Q^c) = 0$. From (1), the first order conditions for the merchant's profit maximisation problem with respect to p^c, Q^c, and p respectively are given by:

(2) $Q^c + \lambda = 0$

(3) $s + b_m(Q^c) - m - \lambda b_c'(Q^c)$

(4) $Q - Q^c + Q'(p)p - c'(Q)Q'(p) - \lambda = 0$

---

of card facilities. How the existence of network effects changes the socially optimal interchange fee is discussed in a later section.

[8] The model here has been constructed so that this benchmark is consistent with standard welfare analysis. As will be shown below, the fees established for credit cards and the rules of the card association do not affect total customer purchases. Thus, these rules and fees have no effect on standard allocative efficiency in the retail market. The only role of credit card fees and rules in this model is to determine the division between cash and credit purchases, and hence, the total transactions costs.

where λ is the Lagrange multiplier on the constraint imposed by the customer's choice of cash and credit purchases.

Substituting $\lambda = -Q^c$ from (2) into (4), the optimal value for *p* is simply the standard profit-maximising price for a monopoly seller. This reflects that when a customer makes both cash and credit purchases, the cash price determines total purchases while the difference between the cash and credit prices determine the infra-marginal split between cash and credit sales. We denote this profit maximising cash price by p^m.

Equation (3) determines the relationship between the cash and credit prices. By substitution, $s + (b_m(Q^c) - m) + Q^c b_c'(Q^c) = 0$. But, by the customer's optimal choice of payment instruments, we know that $b_c(Q^c) - f = s$. Thus, the merchant will set the credit card price so that $(b_c(Q^c) - f) + (b_m(Q^c) - m) + Q^c b_c'(Q^c) = 0$. Substituting for credit card fees, this becomes

(5) $(b_c(Q^c) - c_I) + (b_m(Q^c) - c_A) + Q^c b_c'(Q^c) = 0$

Comparing (5) with the socially optimal rule, and noting that $b_c' < 0$, the merchant will set a credit card price that results in too few credit card sales from a social perspective. While transactions costs are minimised when $(b_c(Q^c) - c_I) + (b_m(Q^c) - c_A) = 0$, the merchant will set prices so that $(b_c(Q^c) - c_I) + (b_m(Q^c) - c_A) > 0$. Proposition 1 immediately follows.

Proposition 1 *A profit-maximising merchant monopolist will not minimise total transactions cost and will have credit card sales Q^c strictly less than the socially optimal level Q^{c*}.*

In the absence of any pricing restriction, the merchant will use credit cards as a form of second-degree price discrimination. Credit cards are more likely to be used by either customers with a relatively high willingness-to-pay or for relatively high value purchases. By setting a relatively high credit card price, the merchant is able to discriminate between these high value sales and other sales. To maximise profits, the merchant will trade off the transactions cost benefits of increased credit card sales, as measured by $(b_c(Q^c) - f) + (b_m(Q^c) - m)$, and the benefit from raising profits by raising credit card prices. The ability to raise credit card prices is limited by the ability of the customer to switch to cash purchases at the margin if credit card prices are too high. This is captured by the term $Q^c b_c'(Q^c)$.

Unlike other models (e.g., Rochet and Tirole, 2002a and Wright, 2001), the (socially undesirable) tendency of merchants to limit credit card sales here does not depend on any network externality or other externality. Rather it is simply a device for price discrimination. The merchant will tend to lower credit card sales and raise the credit card price because this allows them to identify high value customers and high value transactions.

It follows directly from (5) that the merchant's desire and ability to price discriminate against card purchases is unaffected by the interchange fee *a* set by the card association – this fee does not enter equation (5). The interchange fee, however, will determine the relative cash and credit prices. The customer will divide purchases between cash and credit so that $s+c_I=a+b_c(Q^c)$. Given s and the merchant's profit maximising level of credit card transactions, there will be a one-to-one relationship between the credit card price and the interchange fee. A rise in the interchange fee will lead to an equal rise in the credit price, leaving the customer indifferent. The change in the interchange fee also leaves the merchant unaffected. While the rise in the interchange fee leads to a rise in the merchant service charge, this is just "passed through" to customers. While the interchange fee will affect relative prices it will have no real effects.

The neutrality of the interchange fee has also been noted in other specific models, for example, by Carlton and Frankel (1995), and Rochet and Tirole (2002a). It is a general property of payments systems when merchants can set separate cash and credit prices (Gans and King, 2003a). In our framework here it means that the merchant's discrimination against credit card use is only reflected in the relative cash and credit card prices. It does not mean that the credit card price is either higher or lower than the cash price in absolute terms. The exact relationship between the two prices will depend on the interchange fee.

To see this, consider when a merchant will set identical cash and credit prices. If cash and credit prices are identical then the customer will choose credit purchases $\tilde{Q}^c$ so that $f-b_c(\tilde{Q}^c)=0$. By substitution into (5), the credit and cash prices only coincide in the absence of the no surcharge rule if $\left(b_m(\tilde{Q}^c)-c_A-a\right)+\tilde{Q}^c b_c'(\tilde{Q}^c)=0$. Rearranging, the cash and credit card prices will only coincide in the absence of the no surcharge rule if the interchange fee is $\tilde{a}$, which is implicitly defined by $\tilde{a}=\left(b_m(\tilde{Q}^c)-c_A\right)+\tilde{Q}^c b_c'(\tilde{Q}^c)$ and $\tilde{Q}^c=b_c^{-1}(c_I-\tilde{a})$.[9] If the actual interchange fee exceeds $\tilde{a}$ then the credit card price will exceed the cash price. In contrast, if the interchange fee is less than $\tilde{a}$ then the credit card price will be less than the cash price. In this latter case, the merchant is still price discriminating against credit card holders – at the effective credit card price for customers, $s+p+c_I-a$, the customer still chooses an inefficiently low level of credit card purchases.

In summary, when there is frictionless surcharging, interchange fees may change the nominal prices and fees in payment systems but not relative prices or real decisions being made. In this sense, there is neither a socially optimal interchange fee as welfare variables are independent of *a*, nor is there a privately optimal interchange fee as profit variables are independent of *a*. At first blush, this suggests that policies that remove restrictions on surcharging are a substitute for policies that directly regulate interchange fees. However, neutrality has some "knife edge properties". For instance, if there is a single sector that finds surcharging prohibitive for practical reasons,[10] the interchange fee will have real effects on the choice of payment instruments in that sector. In this situation, a regulated interchange fee will be effective as a policy instrument for altering payment instrument choice.

[9] Note that by our assumptions, $b_c^{-1}(.)$ is a well defined, continuous, monotonically decreasing function.

[10] And at the same time the sector is characterised by imperfect competition (Gans and King, 2003a).

3.2 Prohibited surcharging

We now turn to consider the situation where surcharging is prohibited, either by the card association or because merchant pricing contingent on payment instrument is prohibitively costly for some reason. If the merchant can only set a single price, then the division of sales between cash and credit card will be determined completely by the customer. The merchant will set the price p to maximise $\pi = \int_0^{\tilde{Q}^c} b_m(q)dq - m\tilde{Q}^c + pQ - c(Q)$ where, as noted above, the quantity of credit card sales $\tilde{Q}^c$ is chosen by the customer so that $f - b_c(\tilde{Q}^c) = 0$. Assuming that the merchant continues to make both cash and credit card sales, the merchant will simply set the single profit-maximising price at the same level as the cash price in the absence of a no surcharge rule, $s = 0$.

In the absence of surcharging, the interchange fee is no longer neutral and there will be a socially optimal interchange fee that minimises total transactions costs T. At this fee, $\tilde{Q}^c = Q^{c*}$. But, without surcharging, $f - b_c(\tilde{Q}^c) = 0$. Noting that $f = c_I - a$ and that $c_I - b_c(Q^{c*}) + c_A - b_m(Q^{c*}) = 0$, this means that the socially optimal interchange fee is given by $a^* = b_m(Q^{c*}) - c_A$.[11]

The socially optimal interchange fee is intuitive. In the absence of surcharging, the customer chooses the level of credit card transactions according to their own marginal costs and benefits. They ignore the marginal costs and benefits of credit card purchases to the merchant. Thus, the customer's choice of an extra credit card purchase imposes an externality on the merchant. This externality is positive if $c_A < b_m(Q^c)$ and negative if $c_A > b_m(Q^c)$. The interchange fee acts to internalise this externality. The fee is positive if there is a marginal benefit to the merchant from an additional credit card transaction at the socially optimal level of transactions. The interchange fee is negative otherwise.

The optimal interchange fee will depend on the relative marginal benefits from additional credit card transactions to merchants and customers. It is convenient to define a variable α to capture these relative benefits. Thus, at the socially optimal level of credit transactions, $\alpha b_c(Q^{c*}) = (1-\alpha)b_m(Q^{c*})$. Proposition 2 calculates the socially optimal interchange fee.

Proposition 2 *The socially optimal interchange fee is* $a = \alpha c_I - (1-\alpha)c_A$.

Proof: $a^* = b_m(Q^{c*}) - c_A$. By substitution, $a^* = \frac{\alpha}{1-\alpha} b_c(Q^{c*}) - c_A$. But from the first order condition for transaction cost minimisation, $b_c(Q^{c*}) = (1-\alpha)(c_A + c_I)$. By substitution, $a^* = \alpha c_I - (1-\alpha)c_A$.

[11] Wright (2003) obtains essentially the same socially optimal interchange fee in a model where issuers and acquirers are perfectly competitive and there are many heterogeneous merchants. In his specification, however, the optimal interchange fee is set equal to the average benefit of merchant's accepting cards (less acquirer costs) rather than the marginal benefit as arises in our case. Rochet (2003) also notes that this fee is the optimal fee obtained by Baxter (1983) when customers do not have information or do not take into account whether a merchant offers cards when choosing a merchant. Rochet notes that, in this case, this fee would be socially optimal if issuers were imperfectly competitive.

If authorities wish to regulate the interchange fee to the socially optimal level, then they need to estimate the *relative* marginal benefits from additional credit card transactions to merchants and customers at the socially desirable level of credit card transactions. This will often be a difficult (if not impossible) task. In such circumstances, a reasonable starting assumption is that these marginal benefits will be relatively similar. Under this assumption, the socially optimal interchange fee takes a particularly simple form, $a^* = \frac{1}{2}(c_I - c_A)$. In other words, if merchant and customer marginal benefits are relatively symmetric, the socially optimal interchange fee results in equal per transaction credit card fees for both merchants and customers with $m = f = \frac{1}{2}(c_I + c_A)$. Again, this accords with intuition. The interchange fee leads to merchant service charges and customer charges that reflect the marginal benefits of an additional credit card transaction to each of these parties.

It is interesting to note that a term similar to $\frac{1}{2}(c_I - c_A)$ appears in many formulations of interchange fees in the literature. We defer a discussion of these until the following section.

3.3 Linear credit card fees

The analysis above assumes that issuers and acquirers compete through setting two-part tariffs for credit card users. This assumption accords with reality as many credit card schemes do allow members to charge such tariffs. However, it is interesting to consider how the optimal interchange fee might alter if issuers and acquirers can only set linear credit card fees.[12]

To analyse this situation we need modify our assumptions on bank competition. If issuers and acquirers are limited to setting linear card fees, *f* and *m*, without (or with restricted) fixed fees, then competition will not in general force these fees down to marginal cost. Rather, we would expect $f > c_I - a$ and $m > c_A + a$, with *f* decreasing in *a* and *m* increasing in *a*. To capture the effects of issuer competition, let $f = c_I - a + M_I$ where M_I reflects the mark-up over marginal cost charged by issuers. M_I will increase as the level of issuer competition decreases. Similarly, let $m = c_A + a + M_A$ where the mark-up M_A increases when acquirer competition declines.

We continue our assumption that merchant prices cannot be made contingent on payment instrument and consider the socially optimal interchange fee. As before, the interchange fee will not determine the price set by the merchant as card transactions are infra-marginal, but the fee will determine the split of total purchases between payment instruments. Given the merchant price, the customer will make credit purchases until $f = b_c(Q^c)$. The level of card purchases that minimise transactions costs, Q^{c*}, solves $c_I - b_c(Q^{c*}) + c_A - b_m(Q^{c*}) = 0$. Hence, the socially optimal interchange fee is given by $a^* = b_m(Q^{c*}) - c_A + M_I$.[13] Note that this is simply a generalisation of the optimal

[12] Alternatively, banks might have a legal limit or another constraint on the level of fixed fees. If this constraint binds then any further ability of banks to raise profits will be reflected by increasing card fees above marginal costs.

[13] Rochet (2003) provides a more general framework for considering issuer competition based on the formulation in Rochet and Tirole (2002a). In his case, $f = f(c_I – a)$, a general function which collapses to c_I –

interchange fee under two-part tariffs – if $M_I = 0$ then the optimal interchange fee collapses to the fee presented Proposition 2. Similarly, if we define α such that $\alpha b_c(Q^{c*}) = (1-\alpha) b_m(Q^{c*})$, then the optimal interchange fee is $a^* = \alpha c_I - (1-\alpha) c_A + M_I$. Again this collapses to the fee presented in Proposition 2 when $M_I = 0$. Proposition 3 follows from the solution for the socially optimal interchange fee.

Proposition 3 *The socially optimal interchange fee under the no surcharge rule when issuers and acquirers can only set linear fees (a) is independent of the degree of acquirer competition and (b) decreases when issuer competition increases.*

Proof: Part (a) follows immediately from $a^* = b_m(Q^{c*}) - c_A + M_I$. As this equation does not depend on M_A, the optimal interchange fee cannot depend on the level of acquirer competition. Part (b) also immediately follows. An increase in issuer competition will lead to a fall in M_I and a concomitant fall in a^*.

The intuition behind Proposition 3 is straightforward. For part (a), acquirer competition has no effect on the optimal interchange fee because it is customers, not merchants, who determine the mix of cash and credit purchases. A reduction in acquirer competition will make the representative merchant worse off, but it will not change either the socially optimal mix of transactions or the customer's transaction choice. In contrast, issuer competition directly affects the actual mix of cash and card transactions. A reduction in issuer competition raises the mark-up of card fees over true marginal transactions costs and discourages credit card transactions. To offset this tendency towards insufficient use of credit when the issuer segment is not competitive, it is desirable to raise the interchange fee. Raising this fee lowers issuers' costs and, for any level of competition, tends to reduce customer charges (see also Rochet and Tirole, 2002a).

In the discussion on competition with two-part tariffs, we could ignore issues of network effects. We simply made the reasonable assumption that, given marginal fees were set at marginal cost, fixed fees were not set so high as to drive either consumer or merchant surplus from credit card transactions to less than zero. Thus, at both the socially optimal interchange fee and under bank competition, both the merchant and the customer made non-negative surplus from availing themselves of the ability to make card transactions. With linear fees and imperfect competition, this may no longer be the case. While, at the socially optimal interchange fee, the customer gains positive surplus from card transactions (and so will make those transactions if the merchant accepts the credit card) the merchant may not make positive surplus.

To see this, note that the merchant's surplus from accepting cards as well as cash, relative to just accepting cash, is given by $S_m(Q^c) = \int_0^{Q^c} b_m(q)dq - (c_A + a + M_A)Q^c$ when there are Q^c card transactions. At the socially optimal interchange fee, the merchant's surplus simplifies to $S_m(Q^{c*}) = \int_0^{Q^{c*}} b_m(q)dq - b_m(Q^{c*})Q^{c*} - (M_I + M_A)Q^{c*}$. This is always

$a + M_I$ in our case. He finds that the optimal interchange fee is defined implicitly by $c_A + c_I - b_m = f(c_I - a)$. While Rochet's formulation does not allow a closed form statement of the optimal interchange fee, substituting our equilibrium f into his model yields the same interchange fee.

positive when marginal fees are set at marginal cost so $M_I = M_A = 0$. However, when there is imperfect competition and banks are restricted to linear fees, then it may not be possible to implement the socially optimal interchange fee. In particular, if competition in issuing and acquiring is sufficiently weak, so that $S_m(Q^{c^*}) < 0$, then the merchant will refuse to accept credit cards even at the socially efficient interchange fee. Thus, while the socially optimal interchange fee only depends on the degree of acquirer competition, both the degree of issuer and acquirer competition affect the incentive for the merchant to participate in the card scheme at the socially optimal interchange fee. Acquirer competition has a direct effect on the surplus gained by the merchant when they are only able to set a single cash and credit price. Issuer competition indirectly affects the merchant's surplus through the optimal interchange fee. As issuer competition declines, the socially optimal interchange fee rises and this reduces the merchant's surplus from accepting cards.

Our focus in this paper is on the socially optimal interchange fee. In particular, we presumed that the regulator had made a case for regulation and the issue was the benchmark interchange fee. This presumption however invites a simple question: will the profit maximising interchange fee for the banks differ from the socially optimal interchange fee? In the appendix we provide a more formal model of bank competition that is consistent with the above analysis to consider this question. We demonstrate that when surcharging is costly or prohibited, there will likely be a divergence between the socially optimal fee and the fee that is privately set.

3.4 Comparison with suggested approaches

How does the socially optimal fee here compare with the suggested approaches listed in Section 2? Notice that, in our case, $a^* = \alpha c_I - (1-\alpha) c_A$; that is the optimal fee is based on both issuer and acquirer marginal costs as well as α, the relative proportion of marginal benefits flowing to merchants. In the linear fee case, it also depends positively on issuer mark-ups.[14]

The upper bound of this fee is $a^* = c_I$ that arises when $\alpha = 1$ (i.e., there are no cardholder benefits). If $R_I > C_I$, the RBA approach is the only one where the interchange fee lies below this upper bound. In all other cases, the interchange fee will exceed the socially optimal level and it is certain that the cost of transacting will not be at its minimum.

The RBA fee, being strictly less than c_I, may be above or below the social optimum. If it is above the optimal fee, then the RBA fee is strictly preferable to the other suggested methodologies. If it is below it, then a ranking is ambiguous. Of course, note that the upper bound is consistent with an assumption that all of the benefits of the payment system flow the merchants. This is a reason why the RBA chose to focus purely on issuer marginal costs. According to our specification, that approach is consistent with the RBA presumption.

It is also interesting to note that the analysis of this section suggests that the current interchange fee in Australia exceeds the social optimum. In Australia, competition by issuers does take place using two part tariff pricing. There cardholders face fixed charges but, because of the existence of loyalty points, $f \le 0$. This implies that $f = c_I - a$ and,

[14] Notice there that as $M_I = R_I - c_I$ the optimal fee can be re-written as $a^* = R_I - (1-\alpha)(c_I + c_A)$.

consequently, that $a \geq c_I$.[15] Thus, it must be the case that interchange fees are higher than the social optimum as it exceeds the upper bound of a^*. In this respect, subject to some qualifications in the next section, the RBA policy of focusing on issuer costs is a conservative approach to placing a price cap on interchange fees.

4 Other factors

Of course, our model as specified above neglects two potentially important determinants of interchange fees: network effects and strategic merchant benefits. Each of these might imply that the socially optimal interchange fee exceeds c_I.

Network effects arise when merchants do not adopt card services because there are insufficient cardholders. Of course, this only has meaning when there are many different merchants making card adoption decisions. Schmalensee (2002) analyses this case where issuers and acquirers compete in linear prices. Wright (2001) generalises this model to consider other models of issuer and acquirer competition. Indeed, when issuers and acquirers compete in two part tariffs, and if customers and merchants are not heterogeneous, then Wright's model implies that the socially optimal interchange fee is: $a^* = \frac{1}{2}\left(b_m - c_A - (b_c - c_I)\right)$ so long as merchants choose to adopt card services at this interchange fee. In our notation, this fee becomes: $a^* = \alpha c_I - \left(1-\alpha\right)c_A$. As such, at least for this special case, the existence of network effects generates the same interchange fee as our model that assumes no network effects (at least as an upper bound).

Strategic merchant benefits arise when merchants engage in imperfect competition and their adoption of payment instruments plays an important role in attracting customers over their competitors. This situation was analysed by Rochet and Tirole (2002a) and was generalised by Wright (2001) to consider merchant heterogeneity. Wright (2003) demonstrates that in this case, $a^* = \frac{1}{3}\left(2(b_m - c_A) - (b_c - c_I)\right)$; a formula that he derives for perfect issuer and acquirer competition but that would also arise if they competed in two part tariffs. Notice that if $b_m - c_A > (<) b_c - c_I$, this fee exceeds (is less than) the fee from our model. Thus, in the absence of additional information, it is not possible to say whether the existence of strategic merchant benefits would loosen the upper bound on interchange fees that is implied by our model. However, it is true that if strategic reasons for the merchant adoption of credit cards are strong, this tends to favour interchange fees that are relatively high.

Admittedly, the analysis here is based on a fairly specific model of payment systems. However, it is the case that there is substantial agreement in the findings of various models – Schmalensee (2002), Wright (2001, 2003), Gans and King (2003b) and Rochet (2003) – that the socially optimal interchange fee will, at a minimum, ensure that customers internalise the externalities they impose in the choice of payment instrument. In general, when marginal cardholder fees reflect marginal costs facing the issuer, this means that the interchange fee should be set equal to a relevant measure of merchant benefits less the marginal cost of acquirers. Accounting for the potential endogeneity of merchant benefits,

---

[15] Note that if cardholder fixed fees cannot be perfectly set, there is an implicit mark-up in *f*. In this case, $a \geq c_I + M_I$.

this implies a fee that is a weighted sum of the marginal issuing cost and the negative of marginal acquiring cost. Thus, in contrast to the suggested approaches in Section 2, the inclusion of marginal acquiring costs as an offset factor for the interchange fee is socially desirable.[16] However, a true benchmark would likely be based on a more general model; at present, the model of Wright (2001).[17]

5 Conclusion: practical issues in regulation

While there has been some discussion of the relevant benchmark for setting a regulated interchange fee, there is been little discussion in policy debates regarding the practical issues associated with regulating interchange fees. We conclude by offering thoughts on two issues that are of clear practical importance but have not really been analysed. The first concerns the functions of credit cards and whether regulators should make a distinction between these functions in deriving an interchange methodology. The second concerns how regulation will proceed and how revisions might take place.

5.1 Functions of credit cards

The RBA (2001, 2002) consistently made a distinction between the payment functionality aspect of credit cards (something it shares with other payment instruments such as cash, checks and debit cards) and the line of credit function (something that is closer to functions that charge cards provide). They chose to base their methodology purely on payment functionality and for this reason did not include cost components that related to the line of credit function.

The difficulties of doing this can be seen in the RBA's treatment of credit losses and the interest free period. The RBA decided that credit losses – that is, losses that arise from non-payment of credit card debt to issuers – should not be part of the interchange fee. The reason is that issuers currently mark-up the interest rate on credit card debt to recover those losses. The RBA argued that to include these losses in the interchange fee would be, in effect, double counting. This, in turn, would lead to low standards when extending credit to cardholders as banks strategically manipulate their credit losses.

First of all, it is not entirely clear that double counting currently occurs. To be sure, it may be that part of credit losses is built into the interest rate premium and part comes from interchange fees. Given the lack of transparency of existing interchange arrangements, it is simply difficult to tell.

But, more importantly, if there is to be "single counting" imposed, are cardholders the right agents to face the cost of credit losses? If the cost of unpaid debt is covered by putting a premium on interest rates, all debt-paying cardholders are forced to fund the debt

[16] This again suggests that the RBA's approach might be considered conservative as it does not offset the interchange fee with acquirer costs while focusing, for the most part, on issuer marginal costs.

[17] The model would also take into account the regulation of competing payment systems. This situation is analysed by Rochet and Tirole (2002b) and, more completely, by Guthrie and Wright (2003). That paper suggests that when there is competition between payment systems and no coordination in the setting of their interchange fees, those fees tend to be lower than the social optimum. Hence, the existence of competing payment systems is something that questions the need for regulation although Guthrie and Wright (2003) also consider the difficulties that arise if only one payment system is regulated.

defaulters. Alternatively, if credit losses are recovered through interchange fees, merchants are paying for those losses. Merchants, as a group, benefit from the pooling of risk of unpaid debt and, as such, would appear to be the natural party who should pay for this benefit.[18] For this reason, a socially optimal interchange fee would take such losses into account.

The RBA (2002) with some reluctance decided to include the costs of funding the interest free period as a cost component for the interchange fee. However, as noted earlier, the interest free period is determined by card issuers and, indeed, its length is a pricing rather than a pure cost component for them. To include it would not reflect economic costs, and would also open the interchange fee up to strategic manipulation.

The cost of funding the interest free period in part is an artefact of the inability to separate *ex ante* those customers who will or who will not use the associated line of credit. Issuers cannot easily identify which customers will be transactors – paying off credit card debt before interest has occurred – and which will be revolvers – who incur some interest bearing debt at certain times. The very nature of credit cards means that both types of customers will exist. Moreover, allowing payment on extended credit terms relaxes cash-flow constraints facing individual customers and, by increasing the depth of the market, yields benefits to merchants.

For this reason, the nature of credit cards as opposed to any other payment instrument requires that included in the costs borne in part by merchants are the costs of funding the debt that occurs – whether cardholders are actually charged interest or not. That is, cost-based regulation requires that we separate out cost components from revenue components and set prices on the basis of the former only. If an accounting practice, such as the cost of funding the interest free period, confounds the two then the solution is not to eliminate those costs entirely but to create the right measure.

In this case, the right measure is the expected cost of funding credit card debt. To do this, one would have to measure the expected average duration of that debt and also the cost of funding it. This would capture the cost component driving a key benefit to merchants – the liquidity afforded by credit card payments.

The consequences of not including debt cost would be to lower the regulated interchange fee. As we explain below, this does not necessarily mean that credit cards will cease to have an interest free period. However, it may mean that there would be a reduction in the use of credit cards and consequently, a reduction in merchant benefits. This is because at certain times of the year – for example, Christmas – merchants derive benefits from having a liquid payment instrument with an unsecured line of credit and that is likely to lead to a greater level of consumption than might otherwise occur.[19]

5.2 Regulatory revisions

If our experience of price regulation has taught us anything over the last five decades it is that care must be taken as to how regulated prices change over time. In the past, a purely cost-based or rate of return revision gave rise to poor incentives for regulated firms to reduce costs and to avoid over-capitalisation. The modern alternative is to use price cap

---

[18] For a related view see Chakravorti and Emmons (2001).

[19] Katz (2001) argues that payment instruments shift around transactions amongst merchants but do not necessarily raise the volume of transacting. While this may be true in general and a reasonable assumption to motivate the goals of policy, potential seasonal credit constraints need to be taken into account.

regulation with an automated adjustment. A CPI-X (or equivalent) approach creates favourable cost reducing incentives as the regulated firm bears the marginal impact of any reduction or increase in costs.

But what of regulated interchange fees? A regulated fee would have to ensure that both issuers and acquirers have incentives to reduce costs. However, if it is to be based on a benchmark that in part reflects the relative benefits to merchants and customers, revisions in the interchange fee cannot simply be automatic. At the same time, if the regulated fee is based on both the costs of issuers and acquirers, revisions would have to take place in a manner where the actual costs of those issuers and acquirers do not bear significantly on the interchange fee they face. In concentrated and vertically integrated banking industries – such as Australia – how to achieve this is an open issue.

6 References

ABA (2001) “Credit Card Networks in Australia: An Appropriate Regulatory Framework,” *Submission to the RBA*.

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

Carlton, D. and A.S. Frankel (1995) “The Antitrust Economics of Credit Card Networks,” *Antitrust Law Journal*, 68: 643-668.

Chakravorti, S. and W. Emmens (2001) “Who Pays for Credit Cards?” Emerging Payments Occasional Paper Series, EPS-2001-1, Federal Reserve Bank of Chicago.

Frontier Economics (2001) *Report on Credit Card Interchange Fees to Review Banks*, Melbourne.

Gans, J.S. and S.P. King (2001) “The Role of Interchange Fees in Credit Card Associations: Competitive Analysis and Regulatory Options,” *Australian Business Law Review*, 29: 94-122.

Gans, J.S. and S.P. King (2003a) “The Neutrality of Interchange Fees in Payments Systems,” *Topics in Economic Analysis and Policy*, Vol 3, Article 1.

Gans, J.S. and S.P. King (2003b) “A Theoretical Analysis of Credit Card Regulation,” *Economic Record*, Vol.79, No.247, forthcoming.

Guthrie, G. and J. Wright (2003) “Competing Payment Schemes,” Working Paper, No. 245, Department of Economics, University of Auckland.

Katz, M. (2001) *Network effects, interchange fees and no-surcharge rules in the Australian credit and charge cards industry*, Reserve Bank of Australia, August.

RBA (2001) *Reform of credit card schemes in Australia I: A consultation document*, RBA: Sydney.

RBA (2002) *Reform of credit card schemes in Australia IV: Final reforms and regulation impact statement*, RBA: Sydney.

RBA/ACCC (2000) *Debit and Credit Card Schemes in Australia: A Study of Interchange Fees and Access*, RBA: Sydney.

Rochet, J-C. (2003) "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *Review of Network Economics*, this issue.

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

Rochet, J-C. and J. Tirole (2002b) "Platform Competition in Two-Sided Markets," mimeo, University of Toulouse.

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

Wright, J. (2000) "An Economic Analysis of a Card Payment Network," mimeo, University of Auckland.

Wright, J. (2001) "The Determinants of Optimal Interchange Fees in Payment Systems," Working Paper, No.220, Department of Economics, University of Auckland.

Wright, J. (2003) "Pricing in Debit and Credit Card Schemes," *Economics Letters*, forthcoming.

7 Appendix: profit maximising interchange fees

In this appendix, we consider a credit card association that wishes to maximise the total profit of participants, and calculate the profit maximising interchange fee. We then analyse when this fee will coincide with the socially optimal fee. As the interchange fee only matters for profits when frictionless surcharging cannot occur, we assume that there is no surcharging throughout this appendix. We assume initially that credit card fees comprise two part tariffs before turning to the linear fee case.

With non-linear tariffs, the marginal fees for card usage will be set at marginal cost, and all bank profits will be derived from fixed card fees. We capture the level of competition in issuing and acquiring by a parameter $\gamma_i \in [0,1]$ where $i = A, I$ refers to acquirers and issuers respectively. If $\gamma_i = 1$ then there is a single bank providing the relevant credit card service. This bank can act as a monopoly and can charge a fixed fee that reaps the entire surplus from using credit (relative to cash) from their customer. If $\gamma_i = 0$, then there is perfect competition in the relevant function and banks charge no fixed fees and make no economic profits in this activity. Intermediate parameter values refer to intermediate levels of competition.

The total profit accruing to the banks that are members of the card association is given by:

$$\Pi = \gamma_A \left(\int_0^{Q^c} b_m(q)dq - (c_A + a)Q^c \right) + \gamma_I \left(\int_0^{Q^c} b_c(q)dq - (c_I - a)Q^c \right)$$

The first bracketed term is the merchant's total surplus from credit card transactions relative to cash. The second bracketed term is the equivalent surplus for the representative customer.

A credit card association that wishes to maximise total profit will set *a* and Q^c to maximise total issuer and acquirer profits subject to the customer's decision about the payment instrument, $c_I - a = b_c(Q^c)$. Substituting for *a*, the first order condition with respect to Q^c is

(6) $\gamma_A \left(b_m(Q^c) - c_A + b_c(Q^c) - c_I \right) + \left(\gamma_A - \gamma_I \right) Q^c b_c'(Q^c) = 0$

To analyse the fee set by the association, first, suppose that the degree of issuer and acquirer competition is the same so that $\gamma_I = \gamma_A$. Then equation (6) is identical to the equation for the first-best interchange fee. Thus, if competition is symmetric in the issuing and acquiring segments, the credit card association will set the socially optimal interchange fee.

The intuition behind this result is straightforward. Suppose that $\gamma_I = \gamma_A = 1$. Then the issuers and acquirers are both monopolies and these banks are able to seize all the surplus created by the use of credit cards. As such, the association will maximise joint profits by maximising credit card surplus. This is achieved by minimising transactions costs. The association's objective is the same as the social objective. This continues to hold even if $\gamma_I = \gamma_A < 1$. The association captures a fixed fraction of the total social benefits created by the use of credit cards and it pays the association to maximise these benefits.

In contrast, suppose that $\gamma_I \neq \gamma_A$. If $\gamma_I < \gamma_A$, then competition is more intense in card issuing than in merchant acquiring. Noting that $b_c'(.) < 0$, at $Q^c = Q^{c*}$ the left-hand-side of (6) will be negative. The profit maximising level of credit card transactions for the card association will involve $Q^c < Q^{c*}$ so that the association will set an interchange fee *below* the socially optimal fee. Serving merchants is relatively more profitable for the banks than issuing cards. Thus, the association wishes to increase merchant surplus from credit card transactions. Lowering the interchange fee and hence lowering the merchant service charge achieve this, albeit at a cost of lowering total credit card transactions.

Conversely, if $\gamma_I > \gamma_A$ then competition is more intense in merchant acquiring than in issuing cards and at $Q^c = Q^{c*}$ the left-hand-side of (6) will be positive. The profit maximising level of credit card transactions for the card association will involve $Q^c > Q^{c*}$. The association will set an interchange fee *above* the socially optimal fee leading to a lower customer fee. This is profitable for the association members as they are able to capture relatively more of the surplus from customers than from merchants, albeit it also leads to excessive use of credit cards. Notice that, in contrast to similar conclusions that rely on the presence of network externalities (Rochet and Tirole, 2002a), here the potential for high interchange fees does not occur because of the potential exploitation of a cross subsidy from cash to card using customers.

In summary, under a no surcharge rule, a credit card association will only have an incentive to set the socially optimal interchange fee, and so minimise total transactions costs, if competition is “balanced” between issuers and acquirers. If competition is not balanced then the association will set an interchange fee that favors the sector that is relatively less competitive.

Finally, we can consider the interchange fee that a profit maximising card association will set when issuers and acquirers are restricted to setting linear fees. Total association profit is given by $\left(M_A + M_I\right)Q^c$, so the association always finds it profitable to encourage an increased volume of card transactions. But the association is constrained by two factors. First, given the retail price set by the merchant, the customer will divide purchases between card and cash so that $c_I - a + M_I = b_c(Q^c)$. The customer’s marginal benefit from credit relative to cash declines as total credit purchases rise. So, to encourage credit card transactions, the card association wishes to set the interchange fee as high as possible. Raising the interchange fee lowers the customer’s card fee and encourages the customer to make more card transactions. The ability of the association to raise the fee, however, is limited by the second requirement – that the merchant must expect to make a surplus from accepting credit cards. Thus, the association can only raise the interchange fee so long as $S_m(Q^c) > 0$. The profit maximising interchange fee will set $S_m(Q^c) = 0$ resulting in credit card purchases that solve $\int_0^{Q^c} b_m(q)dq - \left(c_A + M_A + c_I + M_I - b_c(Q^c)\right)Q^c = 0$.

Proposition 4 *The profit maximising interchange fee for the card association under the no surcharge rule when issuers and acquirers can only set linear fees is always at least as high as the (constrained) socially optimal interchange charge. Further, when the socially optimal interchange fee provides the merchant with positive surplus then the profit maximising interchange fee (a) is decreasing in the marginal cost of both issuing and acquiring and (b) increases as the degree of either issuing or acquiring competition increases.*

Proof: The first part of the proposition follows immediately from $S_m(Q^c) = 0$ under the profit maximising interchange fee, $S_m(Q^{c*}) \geq 0$ under the (constrained) socially optimal interchange fee and $c_I - a + M_I = b_c(Q^c)$ by the consumer’s decision about payment instruments. For part (a) and (b), totally differentiating $S_m(Q^c) = 0$ means that

$$\frac{da}{dM_i} = \frac{da}{dc_i} = \frac{Q^c}{\left(b_m(Q^c) + b_c(Q^c) - c_A - c_I - M_A - M_I + b_c'(Q^c)\right)\frac{\partial Q^c}{\partial a}} \text{ for } i = A, I .$$

The profit maximising interchange fee is always at least as large as the socially optimal fee and always drives merchant surplus to zero. Thus, if merchant surplus is positive under the socially optimal fee, then this fee is less than the profit maximizing interchange fee. By the consumer’s choice of payment instruments this means that at the profit maximising interchange fee, $Q^c > Q^{c*}$ and $b_m(Q^c) + b_c(Q^c) - c_A - c_I < 0$. Substituting in and noting that $\frac{\partial Q^c}{\partial a} > 0$ and $b_c'(.) < 0$ means that $\frac{da}{dM_i} = \frac{da}{dc_i} < 0$. (a) and (b) immediately follow.