Credit card interchange fees *

Jean-Charles Rochet[†] Julian Wright[‡]

December 23, 2008

Abstract

We build a model of credit card pricing that explicitly takes into account credit functionality. We show that a monopoly card network always selects an interchange fee that exceeds the level that maximizes consumer surplus. If regulators only care about consumer surplus, a conservative regulatory approach is to cap interchange fees based on retailers' net avoided costs from not having to provide credit themselves. In the model, this always raises consumer surplus compared to the unregulated outcome, often to the point of maximizing consumer surplus.

1 Introduction

Even through payment cards are gradually becoming the most popular and most efficient means of payments in many countries, there is a growing suspicion surrounding the pricing of credit cards. Retailers complain that the fees they have to pay to accept credit card transactions are out of proportion with the costs incurred by banks. Some competition authorities and central banks have suggested banks provide consumers with exaggerated incentives to use their credit cards, to the detriment of other means of payments like cash and debit cards which they believe to be more efficient. The usual suspects are the interchange fees, the transfer fees paid by the bank of the retailer to the banks of the cardholders, which are often considerably higher than those for debit

*We thank conference participants at the 'Conference on Payments' at the Central Bank of Norway for helpful comments.

[†]Toulouse School of Economics. Email: rochet@cict.fr

[‡]Department of Economics, National University of Singapore: E-mail: jwright@nus.edu.sg.

card.[1] In the past several years, there have been more than 50 lawsuits concerning interchange fees filed by merchants and merchant associations against card networks in the United States, while in about 20 countries public authorities have take regulatory actions related to interchange fees and investigations are proceeding in many more (Bradford and Hayashi, 2008).

Given the obvious importance of understanding how interchange fees should be set, this article analyzes credit card interchange fee determination to see whether there are grounds for regulatory intervention, and if so, in what form. The point of departure from the existing literature is to model *credit cards* explicitly. An existing literature models price determination in payment cards networks, initiated by Schmalensee (2002), Rochet and Tirole (2002) and Wright (2003).[2] The models in this literature have essentially focused on the choice between payment cards (which could just as well be debit cards) and cash (or checks). We contribute to this literature by extending the models to allow a separate role for the credit functionality of credit cards, thereby allowing us to discuss credit card interchange fees specifically.[3]

In our model, credit cards can be used for two types of transactions — "ordinary purchases" for regular convenience usage for which cash (or a debit card) are assumed to provide identical benefits, and for "credit purchases" where credit is necessary for purchases to be realized. Credit purchases include a range of different types of purchases (such as unplanned purchases, impulse purchases and large purchases) for which the consumer does not have the cash or funds immediately available to complete the purchase. For regular convenience usage, we assume credit cards are inefficient given we assume there are additional costs of transacting with credit cards. As a result, card networks which maximize profit by maximizing the number of card transactions have an incentive to encourage over-usage of credit cards by convenience users (even when these consumers do not need the credit facility) provided merchants still accept such credit card transactions. A card network does this by setting interchange fees high enough to induce issuers to offer rewards and cash back bonuses (equivalent to negative fees). On the other hand, the alternative to using credit cards for credit purchases is the direct provision of credit by retailers or "store credit", which is assumed to be relatively inefficient. Since consumers do not internalize

[1]Unregulated credit card interchange fees are typically between 1% to 2% of transaction value, whereas debit card interchange fees are typically between 0% and 1%. See, for instance, Charts 2 and 3 in Weiner and Wright (2005).

[2]See also Baxter (1983) for a much earlier treatment, and Rochet (2003) for a survey of the literature.

[3]Among the few papers to model explicitly the credit functionality are Chakravorty and To (2007) and more recently Bolt and Chakravorty (2008), although these papers do not focus on the determinant of interchange fees for credit cards, nor the regulation of these fees.

retailers' cost savings from avoiding direct provision of credit and since merchants cannot distinguish the type of consumer they face, there is also a case for setting a relatively high interchange fee so that consumers needing to use credit are induced to use credit cards when it is efficient for them to do so. For this reason, to maximize consumer surplus (including the surplus of cash customers) may require setting an interchange fee which induces some excessive usage of credit cards.

Taking into account both types of transactions, the card network always sets its interchange fee too high in our setting. Thus, if regulators only care about (short-run) consumer surplus, our theory can provide a rationalization for placing a cap on interchange fees.[4] The theory suggests one of two possible caps will maximize consumer surplus. Depending on the relative costs and benefits of the different instruments, the cap should either be based on the issuers' costs (to avoid excessive usage of cards for convenience purchases) or on retailers' net avoided costs from not having to provide credit directly (so that consumers use their cards efficiently for credit purchases). Since evaluating which of the two options gives higher consumer surplus is informationally very demanding, a conservative regulatory approach would be to cap interchange fees using the maximum of these two levels, which we show is the latter option. In our model, this always raises consumer surplus compared to the unregulated outcome, and will sometimes result in the best outcome for consumers. In contrast, using issuer costs to regulate interchange fees gives the lower bound of possible interchange fees that maximize consumer surplus.

2 The Model

There is a continuum of consumers (of total mass normalized to one) with quasi-linear preferences. They spend their income on a composite good taken as a numéraire and on retail goods costing γ to produce. There are two payment technologies: cash (which could also capture cheques and debit cards) and credit cards. Credit cards are assumed to be more costly but allow consumers to purchase on credit. They are held by a fraction x of consumers, where x is initially taken as given.[5] Consumers purchase one unit of the retail good (what we call "ordinary purchases") providing them with utility u_0 with $u_0 > \gamma$. In addition, with probability θ, they also receive

[4] In focusing on consumer surplus, we ignore the need for issuers to recover fixed costs and the effect this has on entry incentives (and therefore, on long-run consumers surplus). We also ignore the need to get consumers to internalize the effect of their decisions on the profit of issuers so as to maximize total welfare. As Rochet and Tirole (2008) show, taking these effects into account justifies higher interchange fees.

[5] In section 4 we extend the model to allow consumers to decide whether to join the card network or not.

utility u_1 from consuming another unit of the retail good (what we call "credit purchases"). We assume that retailers cannot bundle the two transactions and also cannot distinguish between "ordinary" and "credit" purchases. We assume that each consumer has always enough cash to pay for his ordinary purchases, but must rely on credit for credit purchases.

Each retailer can directly provide credit to the consumer ("store credit"), but this entails a cost c_B for the consumer (buyer) and cost c_S to the retailer (seller). The cost c_S is the same for all credit purchases from a given seller, but c_B is transaction specific, and is observed by the consumer only when he is in the store. c_B is drawn from a continuous distribution with the cumulative distribution function H. We assume the distribution has full support over some range $(\underline{c}_B, \overline{c}_B)$ where $\underline{c}_B$ is sufficiently negative, such that cardholders will sometimes choose to use store credit[6], and $\overline{c}_B$ is positive but not too high, such that consumers will always prefer to buy each of the goods than not buy at all. H is also assumed to be well behaved, so that the objective functions analyzed below are concave over the relevant ranges.

Given the costs of store credit, accepting credit cards is a potential means for retailers to reduce their transaction costs of accepting credit purchases and to increase the quality of service to buyers. The cost of a credit transaction is c_A for the bank of the retailer (an acquirer) and c_I for the bank of the cardholder (an issuer). The total cost of a credit card transaction is thus $c = c_A + c_I$. Bank fees for credit card transactions are denoted f for consumers and m for merchants. When $f < 0$ (cash back bonuses) consumers prefer to also use their credit cards for ordinary purchases, which given our assumptions is socially wasteful.[7] For each credit card transaction, an interchange fee a is paid by the bank of the merchant (the acquirer) to the bank of the consumer (the issuer).

For simplicity, we assume that acquiring merchants is perfectly competitive for banks, which implies that the merchant fee m is equal to the sum of the acquiring cost c_I and the interchange fee a:

$$m = c_I + a. \tag{1}$$

By contrast, we assume that issuers are imperfectly competitive: the cardholder fee f is equal to the net issuer cost $c_I - a$ plus a profit margin π, assumed for simplicity to be constant. We

[6]For example, store credit may allow cardholders to temporarily access more credit, so preserving their credit card balance for some other anticipated use. This assumption will also be relaxed in section 4.

[7]We assume consumers and merchants face no costs of using cash (which can be thought of as a normalization). For convenience transactions, any costs or benefits of using cash are assumed to be the same as those obtained from using credit cards (which can be thought of as an approximation).

thus have:

$$f(a) = c_I - a + \pi. \tag{2}$$

The level of the interchange fee a has an impact on retailer cost and thus on the retail price[8] $p(a)$, which results from competition between retailers. To model competition between retailers, we use the standard Hotelling model: consumers are uniformly distributed on an interval of unit length, with one retailer ($i = 1, 2$) located at each extremity of the interval. Transport cost for consumers is t per unit of distance. We consider the case where credit card service are provided by a single network (monopoly).[9] Its objective is to maximize total profits of its member banks, which is proportional to total volume of credit card transactions. This simplification is due to our assumption that banks' profit margins are constant (0 for acquirers, π for issuers).

The timing of our model is as follows:

- The card network sets the interchange fee a so as to maximize banks' total profit.
- Banks set their fees: $f(a) = c_I - a + \pi$ for cardholders and $m(a) = c_A + a$ for retailers.
- Retailers independently choose their card acceptance policies: $L_i = 1$ if retailer i accepts credit cards, 0 otherwise.
- After observing (L_1, L_2), retailers independently set retail prices p_1, p_2.
- Consumers observe retail prices and card acceptance policies and select one retailer to patronize.
- Once the consumer is in the store, nature decides whether he has an opportunity for a credit purchase (this occurs with probability θ) and in this case, the cost c_B of using store credit for the buyer is drawn according to the c.d.f. H. Cardholders then select their mode of payment.

In the framework above we assumed that consumers sometimes will obtain a negative draw of c_B (i.e. $\underline{c}_B < 0$). This was one way to ensure store credit is not dominated by credit cards. Despite this, we also assumed store credit is never used for ordinary purchases. We did this by assuming that for ordinary purchases consumers can only choose between cash and credit cards. This helps to simplify the analysis. Subsequently, we will show our results continue to hold

[8] We assume that retailers cannot, or do not want to, charge different retail prices for cash and card payments (i.e. no surcharging).

[9] We will discuss the generalization to the case with competing networks.

even without this restriction so that consumers can choose between cash, credit cards and store credit for ordinary purchases. Another reason why store credit may not be dominated by credit cards is that some consumers will not want to hold credit cards unless the costs of doing so are sufficiently subsidized (or they receive sufficient rewards for doing so). We, thus, also analyze an alternative specification in which we endogenize the choice by consumers of whether to hold a credit card in the first place but then assume the cost to consumers of using store credit c_B is always positive (so there is no problem of consumers wanting to use store credit for ordinary purchases). These two extensions are analyzed in Section 4.

3 Analysis and policy implications

We first analyze the model to derive optimal interchange fees. This involves determining when retailers will accept credit cards so as to determine what interchange fees a card network will set. We then compare this to the interchange fees that maximize consumer surplus and derive some policy implications.

3.1 When Do Retailers Accept Credit Cards?

This section derives the equilibrium behavior of retailers as a function of the fundamental policy variable in our model, namely the interchange fee a. We first construct a retailer's profit function.

We will use the notation $L_c = 1$ if $f < 0$, $L_c = 0$ if $f \geq 0$ to distinguish whether credit cards are used for ordinary purchases or not. A fraction $1 - xL_i$ of the time, consumers cannot use credit cards. For ordinary purchases, such consumers must therefore use cash. For credit purchases, which happen with probability θ, these consumers will always use store credit. A fraction xL_i of the time, consumers can also use credit cards. For ordinary purchases, these cardholders will use cash if $f \geq 0$ (i.e. if $L_c = 0$) and credit cards if $f < 0$ (i.e. if $L_c = 1$). For credit purchases, which happen with probability θ, these cardholders will use store credit if $c_B < f$ but otherwise will use credit cards. Collecting together a retailer's margins associated with each of these different possibilities, retailer i's expected margin per-customer is therefore equal to

$$\begin{aligned} M_i &= (1 - xL_i)(p_i - \gamma + \theta(p_i - \gamma - c_S)) \\ &\quad + xL_i(p_i - \gamma - L_c m + \theta(H(f)(p_i - \gamma - c_S) + (1 - H(f))(p_i - \gamma - m))) \end{aligned}$$

or after simplifying

$$M_i = (1+\theta)(p_i - \gamma) - \theta c_S - x\Gamma(a) L_i,$$

where

$$\Gamma(a) = L_c m + \theta (m - c_S)(1 - H(f))$$

is the retailer's expected net cost per-cardholder from accepting cards as a function of the interchange fee.

Corresponding to each of the different types of possible transactions, retailer i offers an expected surplus (ignoring transportation costs) of

$$\begin{aligned} U_i &= (1 - xL_i)(u_0 + \theta u_1 - (1+\theta) p_i - \theta E(c_B)) \\ &\quad + xL_i \left(u_0 + \theta u_1 - (1+\theta) p_i - L_c f - \theta \left(\int_{-\infty}^{f} c_B dH(c_B) + \int_{f}^{\infty} f dH(c_B) \right) \right) \end{aligned}$$

or after simplifying

$$U_i = u_0 + \theta u_1 - (1+\theta) p_i - \theta E(c_B) + xS(a) L_i,$$

where

$$S(a) = -L_c f + \theta \int_{f}^{\infty} (c_B - f) dH(c_B)$$

is the expected net consumer surplus per-cardholder from credit card usage as a function of the interchange fee. Determining retailer i's market share by finding the indifferent consumer in the normal way, we find

$$s_i = \frac{1}{2} + \frac{(1+\theta)(p_j - p_i)}{2t} + xS(a)\frac{L_i - L_j}{2t}. \tag{3}$$

Retailer i's profit function can then be written $\pi_i = M_i s_i$. Let us denote by $\phi(a)$ the difference between $S(a)$ and $\Gamma(a)$, which we call total user surplus:

$$\phi(a) = S(a) - \Gamma(a) = \theta \int_{c_I + \pi_I - a}^{\infty} (c_B + c_S - c - \pi_I) dH(c_B) - (c + \pi_I) L_c. \tag{4}$$

The first term represents total expected surplus of the two users (consumers and retailers) from using credit cards for consumers' credit purchases. The second term is a deadweight loss associated with convenience usage for ordinary purchases: the gain for cardholders is more than offset by the loss of retailers. We are now ready to derive the equilibrium choices of retailers. We first derive equilibrium resulting from price competition between retailers for given card acceptance decisions L_1, L_2.

Proposition 1 *For any couple L_1, L_2 of card acceptance decisions, price competition between retailers leads to retail prices such that*

$$(1+\theta)p_i^* = t + \gamma(1+\theta) + \theta c_S + x\Gamma(a)L_i + \frac{x}{3}\phi(a)(L_i - L_j). \tag{5}$$

Retailers' profits are $\pi_i^* = 2t(s_i^*)^2$, where retailer i's equilibrium market share s_i is

$$s_i^* = \frac{1}{2} + \frac{x\phi(a)(L_i - L_j)}{6t}. \tag{6}$$

Proof of Proposition 1: See the Appendix.

So as to maximize profit, each retailer chooses to accept cards if it increases its market share, which will be the case whenever $\phi(a) \geq 0$. An immediate consequence of Proposition 1 is:

Proposition 2 *Retailers accept credit cards at equilibrium if and only if expected consumer surplus from card transactions exceeds expected retailer cost. That is, $L_1^* = L_2^* = 1$ if and only if $S(a) \geq \Gamma(a)$.*

3.2 Analysis of Retailer Prices and Consumer Surplus

This section considers the impact of the interchange fee a on consumer surplus. Obviously this question only matters in the region where $S(a) \geq \Gamma(a)$, i.e. when credit cards are accepted.

Proposition 1 allows us to compute retail prices and consumer surplus as a function of a. Considering the case where $L_1 = L_2 = 1$ and thus $p_1(a) = p_2(a) = p(a)$, formula (5) gives

$$(1+\theta)p(a) = t + \gamma(1+\theta) + \theta c_S + x\Gamma(a). \tag{7}$$

Since $\Gamma(a)$ (the expected net retailer cost from card transactions) increases in a we obtain an immediate corollary of Proposition 1.

Corollary 1 *The equilibrium retail price is an increasing function of the interchange fee a.*

The expected surplus of cash consumers (those not holding credit cards) is

$$U_{\text{cash}}(a) = u_0 + \theta u_1 - \frac{t}{4} - (1+\theta)p(a) - \theta E(c_B)$$

and of consumers holding credit cards is

$$U_{\text{credit}}(a) = u_0 + \theta u_1 - \frac{t}{4} - (1+\theta)p(a) - L_c f - \theta\left(\int_{-\infty}^{f} c_B dH(c_B) + \int_{f}^{\infty} f dH(c_B)\right).$$

If we aggregate the surplus of all consumers (cash consumers and cardholders) and take into account equilibrium prices from (7), we obtain

$$\begin{aligned} CS(a) &= xU_{\text{credit}}(a) + (1-x)\,U_{\text{cash}}(a) \\ &= u_0 + \theta u_1 - \gamma\,(1+\theta) - \frac{5t}{4} - \theta\,(E\,(c_B) + c_S) + x\phi(a). \end{aligned} \tag{8}$$

Thus, aggregate consumer surplus is equal, up to an additive and a positive multiplicative constant, to total user surplus $\phi(a)$.

The following three figures show how total user surplus ϕ varies with the level of the interchange fee a. Note

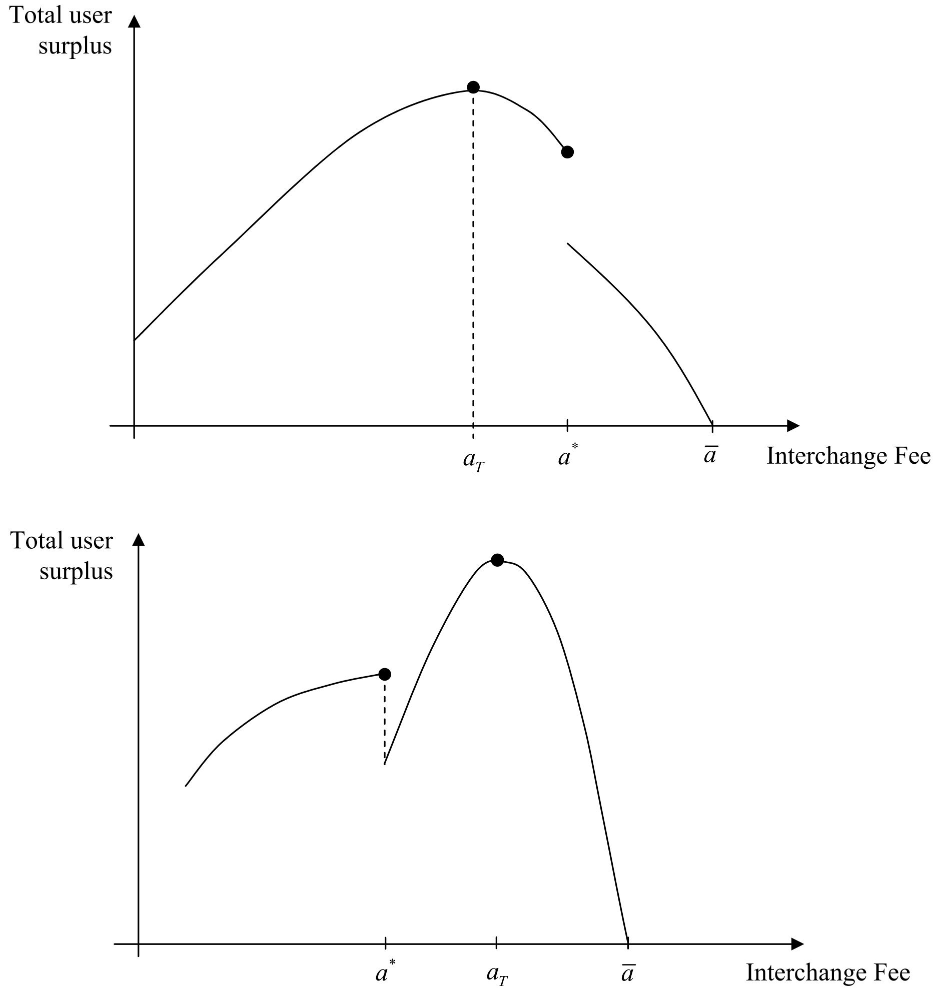
$$\frac{d\phi\,(a)}{da} = \theta\,(-a + c_S - c_A)\,h\,(c_I + \pi_I - a)\,,$$

so $\phi\,(a)$ obtains a local maximum at $a = a_T \equiv c_S - c_A$ provided H is well behaved so that $\phi\,(a)$ is concave over the relevant ranges (as illustrated in the figures). This may also be a global maximum. In the first regime, total user surplus is maximized for $a = a_T \equiv c_S - c_A$, which is less than $a^* = c_I + \pi_I$. This corresponds to the unrealistic situation where credit card transactions are more costly to provide than retailer provided store credit:

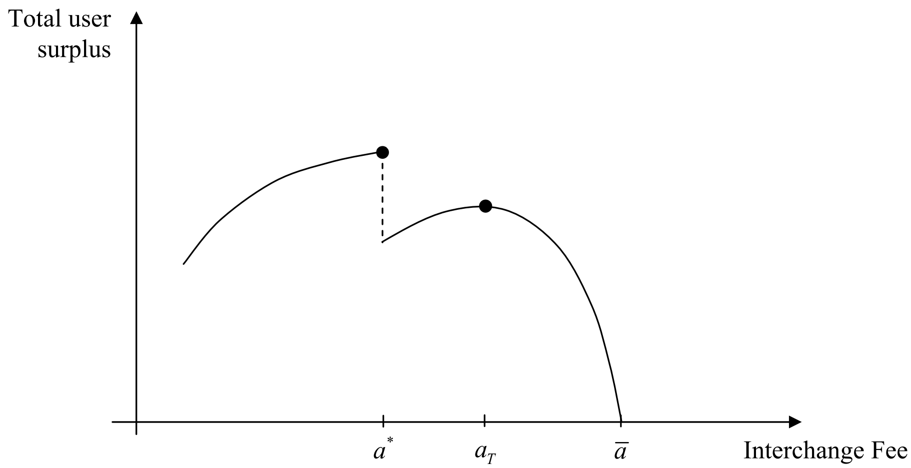
$$c_S - c_A \leq c_I + \pi_I \quad \Leftrightarrow \quad c_S \leq c + \pi_I.$$

In regime 1, regulation to maximize consumer surplus would require making sure credit cards were priced to be *more* expensive for consumers than using cash (or debit cards). We do not consider this case as relevant.

The more realistic cases are captured by regimes 2 and 3, which correspond to the reverse situation. In regime 2 the incidence of convenience users is not so large (so that a_T is still the maximum of total user surplus). By contrast in regime 3, the maximum of ϕ is obtained at a^*.



Second regime: total user surplus is maximum for $a = a_T > a^*$.



Third regime: total user surplus is maximum for $a = a^*$

In all regimes, $\bar{a}$ denotes the interchange fee selected by a monopoly network.

3.3 Policy Implications

Recall that the volume of credit card transactions, and thus the profit of banks, increase with a. This implies that a monopoly card network will set an interchange fee equal to the maximum level $\bar{a}$ that is compatible with merchant acceptance of payment cards. This level is always greater than both a_T and a^*.[10] Therefore, if competition authorities aim at maximizing (short-term) consumer surplus (or equivalently $\phi(a)$), they will always find the privately optimal interchange fee excessive.[11] A regulatory cap on interchange fees can therefore be justified but the appropriate level of the cap is not always the same: it is a_T in regime 2 and a^* in regime 3. These results are recapitulated in the next proposition:

Proposition 3 *If regulatory authorities aim at maximizing (short-term) consumer surplus, privately optimal interchange fees are too high. A regulatory cap on interchange fees can therefore raise (short-term) consumer surplus, but two cases must be considered:*

- ***a)*** *If* $c_S \leq c + \pi_I$ *or if* $c + \pi_I \leq \theta \int_{c+\pi_I - c_S}^{0} (c_B + c_S - c - \pi_I)\, dH(c_B)$*, the regulatory cap should be* $a_T = c_S - c_A$*.*

[10] By adapting the arguments of Guthrie and Wright (2007) one can show that Bertrand competition between two card networks providing perfectly substitutable credit cards leads to interchange fees that are also (weakly) higher than the level that maximizes consumer surplus.

[11] This is not necessarily true anymore if competition authorities aim at maximizing long-term consumer surplus or social welfare, which includes banks' profits (see Rochet and Tirole, 2002, 2008) or if merchants are heterogenous (see Wright, 2004).

b) *If* $c_S > c + \pi_I > \theta \int_0^{c_S - c - \pi_I} (c_B + c_S - c - \pi_I) dH(c_B)$*, the regulatory cap should be* $a^* = c_I + \pi_I$*.*

The proposition implies that the interchange fee $a_T = c_S - c_A$ achieves the global maximum of $\phi(a)$ provided (i) $a_T \leq a^* = c_I + \pi_I$ (i.e provided $L_c = 0$ at a_T) or (ii) if $a_T > a^*$ (i.e. $L_c = 1$) provided that $\theta \int_{c+\pi_I - c_S}^{0} (c_B + c_S - c - \pi_I) \, dH(c_B) \geq c + \pi_I$. Regime 2 arises when the drop in surplus caused by inefficient use of credit cards for ordinary purchases is not too great. Regime 3 arises when the deadweight loss from excessive use of credit cards by convenience users dominates. In either case, lowering interchange fees from the private maximum to a_T unambiguously raises consumer surplus. Further lowering of interchange fees (towards a^*) may raise or lower consumer surplus (and always initially lowers it). Therefore, given it would be extremely difficult in practice for a regulator to determine which of these two regimes prevails, a conservative approach to regulation would be to cap the interchange fee at a_T. This cap is relatively easy to construct. It is based on a retailer's net avoided costs. This is the retailer's cost saving from not having to provide credit directly, or in other words, the cost of providing credit itself less the cost of the acquiring service offered by its bank (i.e. the costs incurred by the retailer's bank in providing the retailer with the ability to accept credit cards, including any such costs that are passed onto the merchant other than the interchange fee).[12] In contrast, the interchange fee a^* (which is based on the issuers' cost) represents a lower bound for desirable interchange fees since any interchange fee lower than a^* unambiguously lowers consumers surplus and welfare in our setting. This could be used as a check to make sure the calculated a_T is not too low. Thus, existing regulations which base interchange fees on issuer costs (such as those adopted in Australia) can be rationalized in our framework, although they represent a lower bound on what might be desirable for consumers.

4 Alternative specifications

In reaching the policy implications above, we adopted a model in which cardholders sometimes prefer using store credit for credit purchases. This ensures store credit was not dominated by credit cards. Despite this, consumers were *assumed* to never use store cards for ordinary purchases. In this section, we show this simplifying restriction can be relaxed. First, we allow that consumers can choose between cash, store credit, and credit cards (if they hold credit cards)

[12]Such costs could be calculated for specific industries or classes of retailers, if required.

when making ordinary purchases. This relaxation of assumptions turns out not to change our main results. Second, we consider the case in which the cost of using store credit is always positive ($\underline{c}_B \geq 0$) so that consumers would never want to use store credit for ordinary purchases (cash would always be preferred). Instead, we endogenize the choice by consumers of whether to hold a credit card in the first place. We show this framework still leads to similar policy implications to our benchmark case.

4.1 Allowing store credit for ordinary purchases

We follow the same steps as in section 3. The main impact of the additional types of possible payment transactions is to make the definitions of the underlying terms M_i, $\Gamma(a)$, $S(a)$ and $\phi(a)$ more complicated, while leaving the structure of the analysis largely unaffected. As we will see, allowing consumers to choose store credit when making ordinary purchases only strengthens the policy implications of section 3. Essentially, it increases the number of situations in which it is desirable to have consumers internalize the cost to merchants of providing store credit, and thereby makes it more likely that the interchange fee $a_T = c_S - c_A$ maximizes consumer surplus.

We start by considering the retailer's margins and demand arising from each type of transaction. A fraction $1 - xL_i$ of the time, consumers cannot use credit cards. For ordinary purchases, these consumers will use store credit if $c_B \leq 0$, which happens with probability $H(0)$; otherwise, with probability $1 - H(0)$ they will prefer to use cash. For credit purchases, which happen with probability θ, these consumers will always use store credit. A fraction xL_i of the time, consumers can also use credit cards. For ordinary purchases, these cardholders will use store credit if $c_B < L_c f$, and otherwise either credit cards (if $L_c = 1$) or cash (if $L_c = 0$). For credit purchases, these cardholders will use store credit if $c_B < f$ and otherwise credit cards. Collecting together a retailer's margins associated with each of these different possibilities, retailer i's expected margin per-customer is therefore equal to

$$\begin{aligned} M_i \;=\; & (1 - xL_i)\left((H(0) + \theta)(p_i - \gamma - c_S) + (1 - H(0))(p_i - \gamma)\right) \\ & + xL_i\left(H(L_c f)(p_i - \gamma - c_S) + (1 - H(L_c f))(p_i - \gamma - L_c m)\right) \\ & + xL_i\theta\left(H(f)(p_i - \gamma - c_S) + (1 - H(f))(p_i - \gamma - m)\right) \end{aligned}$$

or after simplifying

$$M_i = (1 + \theta)(p_i - \gamma) - (H(0) + \theta)c_S - x\Gamma(a)L_i,$$

where

$$\Gamma(a) = (H(L_c f) - H(0)) c_S + (1 - H(L_c f)) L_c m + \theta (1 - H(f)) (m - c_S).$$

Similarly, we can rewrite consumers' expected utility and retailer i's market share taking into account the additional types of payment transactions that are possible. Retailer i offers an expected surplus (ignoring transportation costs) of

$$\begin{aligned}
U_i &= (1 - xL_i) \left(u_0 + \theta u_1 - (1+\theta) p_i - \int_{-\infty}^{0} c_B dH(c_B) - \theta E(c_B) \right) \\
&\quad + xL_i \left(\begin{array}{c} u_0 + \theta u_1 - (1+\theta) p_i - \int_{-\infty}^{L_c f} c_B dH(c_B) - \int_{L_c f}^{\infty} L_c f dH(c_B) \\ -\theta \left(\int_{-\infty}^{f} c_B dH(c_B) + \int_{f}^{\infty} f dH(c_B) \right) \end{array} \right)
\end{aligned}$$

or after simplifying

$$U_i = u_0 + \theta u_1 - (1+\theta) p_i - \int_{-\infty}^{0} c_B dH(c_B) - \theta E(c_B) + xS(a) L_i,$$

where

$$S(a) = \int_{L_c f}^{0} (c_B - L_c f) dH(c_B) - \int_{0}^{\infty} L_c f dH(c_B) + \theta \int_{f}^{\infty} (c_B - f) dH(c_B).$$

Total user surplus can then be written

$$\begin{aligned}
\phi(a) &= S(a) - \Gamma(a) \\
&= \int_{L_c(c_I + \pi_I - a)}^{0} (c_B + c_S - (c + \pi_I) L_c) dH(c_B) \\
&\quad + \theta \int_{c_I + \pi_I - a}^{\infty} (c_B + c_S - c - \pi_I) dH(c_B) - \int_{0}^{\infty} (c + \pi_I) L_c dH(c_B).
\end{aligned}$$

The last two terms in $\phi(a)$ are similar to before. The first term is new, reflecting that there may be cost savings from using credit cards for ordinary purchases if this avoids the use of more costly store credit. Together, the first two terms represent total expected cost savings of the two users (consumers and retailers) from using credit cards for consumers' purchases (as opposed to using store credit). The last term is the deadweight loss associated with convenience usage of credit cards for ordinary purchases, although this now arises for fewer transactions since store credit is now used for some ordinary transactions.

Aside from these new surplus definitions, the analysis of retailers' decisions remains unchanged. That is, retailer i's market share and profit expressions remain the same, as do propositions 1 and 2. Retailer i's equilibrium price is now given by

$$(1+\theta) p_i^* = t + \gamma (1+\theta) + (H(0) + \theta) c_S + x\Gamma(a) L_i + \frac{x}{3} \phi(a) (L_i - L_j),$$

where the right-hand-side includes the additional term $H\left(0\right)c_{S}$ arising from the use of store credit for some ordinary purchases by cash customers.[13] Consumer surplus is similarly adjusted for this additional term, and so now equals

$$CS\left(a\right)=u_{0}+\theta u_{1}-\gamma\left(1+\theta\right)-\frac{5t}{4}-\int_{-\infty}^{0}\left(c_{B}+c_{S}\right)dH\left(c_{B}\right)-\theta\left(E\left(c_{B}\right)+c_{S}\right)+x\phi(a).$$

Thus, aggregate consumer surplus is still equal, up to an additive and a positive multiplicative constant, to total user surplus $\phi(a)$.

As before $\phi\left(a\right)$ has a jump (down) at $a=a^{*}$. However, the total user surplus function is now more likely to be maximized for a higher interchange fee. The derivative $\phi'\left(a\right)$ is the same as before except it is multiplied by $\left(L_{c}+\theta\right)/\theta$ which exceeds one for $a>a^{*}$. This also implies a_{T} which maximizes $\phi\left(a\right)$ for $a>a^{*}$ is identical to before. That is, as before $a_{T}=c_{S}-c_{A}$. Proposition 3 now becomes

Proposition 4 *If regulatory authorities aim at maximizing (short-term) consumer surplus, privately optimal interchange fees are too high. A regulatory cap on interchange fees can therefore raise (short-term) consumer surplus, but two cases must be considered:*

- ***a)*** *If $c_{S}\leq c+\pi_{I}$ or if $\int_{0}^{\infty}\left(c+\pi_{I}\right)dH\left(c_{B}\right)\leq\left(1+\theta\right)\int_{c+\pi_{I}-c_{S}}^{0}\left(c_{B}+c_{S}-c-\pi_{I}\right)dH(c_{B})$, the regulatory cap should be $a_{T}=c_{S}-c_{A}$.*
- ***b)*** *Otherwise, the regulatory cap should be $a^{*}=c_{I}+\pi_{I}$.*

The proposition implies that the interchange fee $a_{T}=c_{S}-c_{A}$ achieves the global maximum of $\phi\left(a\right)$ provided (i) $a_{T}\leq a^{*}=c_{I}+\pi_{I}$ (i.e provided $L_{c}=0$ at a_{T}) or (ii) if $a_{T}>a^{*}$ (i.e. $L_{c}=1$) provided

$$\left(1+\theta\right)\int_{c+\pi_{I}-c_{S}}^{0}\left(c_{B}+c_{S}-c-\pi_{I}\right)dH\left(c_{B}\right)\geq\int_{0}^{\infty}\left(c+\pi_{I}\right)dH\left(c_{B}\right). \tag{9}$$

Notice the trade-off here. The left-hand-side of (9) measures the cost saving from using credit cards rather than store credit when credit cards have a negative fee as opposed to a zero fee. The right-hand-side of (9) measures the additional costs to users from the use of credit cards rather than cash for ordinary purchases when credit cards have a negative fee (and so are used whenever $c_{B}>0$). Compared to before, the condition for case (ii) to apply is more easily satisfied since

[13] It is no longer the case that $\Gamma(a)$ is necessarily increasing in a since $\Gamma'(a)=\left(L_{c}+\theta\right)\left(h\left(f\right)\left(m-c_{S}\right)+\left(1-H\left(f\right)\right)\right)$ so that for m much below c_{S}, higher interchange fees could actually lower prices. However, for m not too much below c_{S} (i.e. a not too much below $c_{S}-c_{A}$), then prices will still be increasing in interchange fees.

the left-hand-side integral is the same as before except it is multiplied by $1+\theta$ instead of θ and the right-hand-side integral is the same as before except it is multiplied by $1-H(0)$ instead of 1. In other words, allowing for consumers to choose store credit when making ordinary purchases only strengthens the previous policy implications.

4.2 Allowing for card membership decisions

In this section, we consider an extension of the existing model in which consumers make prior membership decisions, on whether to hold a credit card or not. Prior membership decisions are potentially important since if interchange fee are set too low (for instance, at zero), then consumers can be expected to face higher fees (for instance, the full cost of issuing cards) at which point many consumers may choose to no longer hold credit cards. In order to simplify the resulting model, we assume that consumers always view using store credit as costly ($\underline{c}_B = 0$). With this assumption there is no issue about whether consumers can use store credit for ordinary purchases; even if they could use store credit for these purchases, as was the case in section 4.1, they would always prefer to use cash for this purpose. This assumption on c_B will also imply that, whenever $f < 0$, cardholders will *always* want to use cards if they are accepted. In this case, cardholders will never use store credit. However, as we will see, setting an interchange fee such that $f < 0$ may still be optimal for consumers (as opposed to setting an interchange fee such that $f = 0$) since it induces more people to hold a credit card, thereby reducing the use of expensive store credit for credit purchases by consumers who otherwise would not hold a credit card.

The model is the same as before except that (i) there is an additional choice for consumers as to whether to hold a card or not and (ii) we set $\underline{c}_B = 0$. Specifically, at the same time as retailers choose their card acceptance policies (i.e. stage 3), consumers receive a random draw v of the benefit of holding a credit card and decide whether to hold the card or not. v is drawn according to the distribution function Ψ over the support $(-\infty, \overline{v}]$ for some $\overline{v} > 0$. Let $x(a)$ be the measure of consumers who join given an interchange fee a. Some consumers draw $v < 0$, so they would prefer not to hold a credit card other things equal (e.g. if they didn't expect to use it). This could also capture that there are significant per-customer costs to issuers associated with managing a cardholder which are passed through to cardholders. For other consumers, cards may offer more than just the ability to make transactions at retailers, in which case $v > 0$ is possible.

We follow the same steps as in section 3. The assumption $\underline{c}_B \geq 0$ is consistent with the analysis in section 3 which assumes consumers never use store credit for ordinary purchases. Therefore, the analysis of retailers' decisions (which treat x as given), and the various surplus expressions are identical to before. Note, however, that when $f < 0$, then $1 - H(f) = 1$ since all cardholders will use credit cards rather than store credit, which is just a special case of the more general analysis detailed in section 3. This implies

$$\phi(a) = \theta \int_{(1-L_c)f}^{\infty} (c_B + c_S - c - \pi_I) dH(c_B) - (c + \pi_I) L_c,$$

which is now independent of a if $a > a^*$ (so that $L_c = 1$). In order to make the analysis interesting, we assume

$$\theta(E(c_B) + c_S) > (1 + \theta)(c + \pi_I), \tag{10}$$

so that the surplus created from credit cards (the cost saving) is higher than their additional cost to users when they are always used (i.e. when $f < 0$). Without making the assumption in (10), retailers would reject cards if $a > a^*$. Taking this into account, the card network would set the interchange fee at a^* and there would be no rationale for any regulation. With (10), the card network will want to set $a > a^*$, after which the interchange fee is neutral in terms of usage decisions. Retailers will always accept cards no matter how high the fees and cardholders will always use them. From the point of view of users (in aggregate), interchange fees beyond a^* just represent pure transfer fees. It will therefore be optimal for the card scheme to set interchange fees to the point where all consumers hold cards. Not surprisingly, like before, this will lead interchange fees to be set too high.

Using the expressions for $U_{\text{credit}}(a)$ and $U_{\text{cash}}(a)$ from section 3 but taking into account that $\underline{c}_B = 0$, the expected additional utility to a consumer of holding a card is

$$U_{\text{credit}}(a) - U_{\text{cash}}(a) = \theta \int_{(1-L_c)f}^{\infty} (c_B - f)\, dH(c_B) - L_c f$$

so consumers will hold cards if

$$v > L_c f - \theta \int_{(1-L_c)f}^{\infty} (c_B - f)\, dH(c_B)$$

implying $x(a) = 1 - \Psi\left(L_c f - \theta \int_{(1-L_c)f}^{\infty} (c_B - f)\, dH(c_B)\right)$ which is increasing in a. Taking into account the additional utility from holding a card, total consumer surplus is equal to (up to an additive and multicative constant)

$$\int_{L_c f - \theta \int_{(1-L_c)f}^{\infty} (c_B - f) dH(c_B)}^{\overline{v}} (v + \phi(a))\, d\Psi(u).$$

When $a > a^*$, this simplifies to

$$\int_{(1+\theta)f-\theta E(c_B)}^{\overline{v}} (v + \theta (E(c_B) + c_S) - (1+\theta)(c+\pi_I)))\, d\Psi(u).$$

As before, total surplus has a jump (down) at $a = a^*$. However, the total surplus function behaves slightly differently for $a > a^*$. The interchange fee a_T which maximizes total user surplus for $a > a^*$ is now

$$a_T = \frac{\theta}{1+\theta} c_S - c_A. \tag{11}$$

This is lower than $a_T = c_S - c_A$ in the benchmark model. Here the purpose of the optimal interchange fee is to induce consumers to hold cards even when they otherwise would not want to, so that they internalize retailers' surplus from being able to accept their credit cards rather than rely on store credit for credit purchases. Note the retailers' surplus $c_S - c_A$ only arises a fraction $\theta/(1+\theta)$ of the time, whereas a fraction $1 - \theta/(1+\theta)$ of the time, the retailer is actually worse of by c_A due to the excessive usage of cards. Following this interpretation, a_T can also be written:

$$a_T = \frac{\theta}{1+\theta}(c_S - c_A) - \left(1 - \frac{\theta}{1+\theta}\right) c_A.$$

Proposition 3 now becomes

Proposition 5 *If regulatory authorities aim at maximizing (short-term) consumer surplus, privately optimal interchange fees are too high. A regulatory cap on interchange fees can therefore raise (short-term) consumer surplus, but two cases must be considered:*

a) *If $c_S \leq c + \pi_I$ or if*

$$\int_{(1+\theta)(c+\pi_I)-\theta(E(c_B)+c_S)}^{\overline{v}} (v + \theta(E(c_B) + c_S) - (1+\theta)(c+\pi_I))\, d\Psi(u) \geq \tag{12}$$
$$\int_{-\theta E(c_B)}^{\overline{v}} (v + \theta(E(c_B) + c_S - c - \pi_I))\, d\Psi(u),$$

the regulatory cap should be $a_T = \frac{\theta}{1+\theta} c_S - c_A$.

b) *Otherwise, the regulatory cap should be $a^* = c_I + \pi_I$.*

The proposition implies that the interchange fee $a_T = \frac{\theta}{1+\theta} c_S - c_A$ achieves the global maximum of $\phi(a)$ provided (i) $a_T \leq a^* = c_I + \pi_I$ (i.e provided $L_c = 0$ at a_T) or (ii) if $a_T > a^*$ (i.e. $L_c = 1$) provided (12) holds. Given it may be hard to measure θ with any confidence or to evaluate whether (12) holds or not, a conservative regulatory approach could again be to use

$a_T = c_S - c_A$ as the regulatory cap since according to the model this (i) increases consumer surplus relative to the privately set interchange fee; (ii) possibly maximizes consumer surplus; and (iii) yet is never too low from the point of view of maximizing (short-term) consumer surplus. Thus, in this alternative framework, policy implications remain remarkably similar to our benchmark case in section 3.

5 Conclusions

Much of the existing literature on interchange fees treats payment cards as though they were debit cards. This paper provides a new theory of interchange fees that is specifically applicable to credit card networks. The model we provide captures a trade-off that can arise between the excessive usage of credit cards for convenience purchases and the importance of getting cardholders who do need credit to internalize retailers' avoided costs arising from their credit card usage. In terms of this trade-off, we find that an unregulated card network always sets the interchange fee too high. Consumer surplus can be increased by imposing a cap on interchange fees which equals the retailers' net avoided costs from not having to provide credit themselves. Further lowering interchange fees from this level towards issuing costs may either increase or decrease consumer surplus, although initially consumer surplus always falls in our framework.

We conclude by noting some other implications of our theory, which may be able to explain real-world observations that have previously defied theoretical explanation. If credit is more likely to be needed by customers for large purchases, then the optimal interchange fee should be ad valorem in our setting (thereby better targeting the transfer to cardholders for the types of transactions where credit is needed). Thus, the model potentially provides a justification for the widespread use of ad valorem credit card interchange fees. It also explains why merchants may want to reject credit cards for small transactions (where people are more likely to be able to purchase anyway using cash). Most importantly, it explains why interchange fees are typically lower for debit cards than for credit cards. Finally, the theory suggests large retailers that are able to gain a competitive advantage over smaller rivals from being able to offer their own store-credit to customers, may have an interest in opposing the widespread use of general purpose credit cards.

Appendix

Proof of Proposition 1:

From the text we have

$$\pi_i = \left(\left(1+\theta\right)\left(p_i-\gamma\right)-\theta c_S - x\Gamma\left(a\right)L_i\right)\left(\frac{1}{2}+\frac{(1+\theta)(p_j-p_i)}{2t}+xS\left(a\right)\frac{L_i-L_j}{2t}\right)$$

Differentiating we get

$$\frac{2t}{1+\theta}\frac{\partial\pi_i}{\partial p_i} = t+(1+\theta)(p_j-p_i)+xS(a)(L_i-L_j)-(1+\theta)(p_i-\gamma)+\theta c_S + xL_i\Gamma(a).$$

At a Nash equilibrium, we have for $i, j = 1, 2$:

$$(1+\theta)(2p_i-p_j) = t+\gamma(1+\theta)+\theta c_S + x\left(S(a)(L_i-L_j)+L_i\Gamma(a)\right).$$

Solving for p_i, we obtain formula (5):

$$(1+\theta)p_i = t+\gamma(1+\theta)+\theta c_S + x\Gamma(a)L_i + \frac{x}{3}\phi\left(a\right)\left(L_i-L_j\right). \tag{13}$$

Substituting (13) into (3) and using (4) implies formula (6).

References

BAXTER, W.P. (1983) "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics*, 26: 541-88.

BOLT, W. AND S. CHAKRAVORTY (2008) "Consumer Choice and Merchant Acceptance of Payment Media," mimeo.

BRADFORD, T. AND F. HAYASHI (2008) "Developments in Interchange Fees in the United States and Abroad," Payments System Research Briefing, Federal Reserve Bank of Kansas City. April.

CHAKRAVORTY, S. AND T. TO (2007) "A Theory of Credit Cards," *International Journal of Industrial Organization*, 25: 583-95.

GUTHRIE, G. AND J. WRIGHT (2007) "Competing Payment Schemes," *Journal of Industrial Economics*, 55: 37-67.

ROCHET, J.-C. (2003) "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *Review of Network Economics*, 2: 97-124.

ROCHET, J.-C. AND J. TIROLE (2002) "Cooperation among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics*, 33: 549-70.

ROCHET, J.-C. AND J. TIROLE (2008) "Must-Take Cards: Merchant Discounts and Avoided Costs," mimeo, Toulouse School of Economics.

SCHMALENSEE, R. (2002) "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, 50: 103-22.

WEINER, S. AND J. WRIGHT (2005) "Interchange Fees in Various Countries: Developments and Determinants," *Review of Network Economics*, 4: 290-323.

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

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