

Credit card interchange fees ^{*}

Jean-Charles Rochet[†]

Julian Wright[‡]

October 2009

Abstract

We build a model of credit card pricing that explicitly takes into account credit functionality. In the model 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. This always raises consumer surplus compared to the unregulated outcome, sometimes to the point of maximizing consumer surplus.

1 Introduction

Even though payment cards are gradually becoming the most popular and most efficient means of payment 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 payment like cash and debit cards which they believe to be more efficient. The usual suspects are the credit card interchange fees set by MasterCard and Visa, the transfer fees paid by the banks of the retailers (acquirers) to the banks of the cardholders (issuers), and which are often

^{*}We benefited from the helpful comments of an anonymous referee, conference participants at the AEA January 2009 meeting in San Francisco (especially our discussant, Charlie Kahn) and two Conferences on Payments Economics at the Central Bank of Norway (November 2008) and the Central Bank of Canada (November 2008).

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

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

considerably higher than those for debit cards.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 taken 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 and regulation. The point of departure from the existing literature on interchange fees 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. 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.

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 could capture 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 or for purchases for which the deferment of payment facilitates the transaction. Thus, offering credit allows an individual merchant to make sales that they otherwise would not make. The ability to make these incremental sales is, we think, the major reason explaining why merchants accept credit cards and indeed are willing to pay higher fees to do so compared to the fees paid to accept debit cards, and why prior to the widespread use of credit cards, store credit was much more widely used (Evans and Schmalensee, 2005, pp. 48-51).

For ordinary purchases, 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

¹Unregulated 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).

²See Baxter (1983) for a much earlier treatment, and Rochet (2003) for a unified treatment.

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 merchants or “store credit”, which is assumed to be relatively inefficient. Since consumers do not internalize 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 who wish to rely on 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 excessive usage of credit cards for ordinary purchases.

Taking into account both types of transactions, a monopoly card network always sets its interchange fee too high in our setting. Thus, if regulators only care about (short-term) consumer surplus, our theory can provide a rationalization for placing a cap on interchange fees.³ 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 ordinary purchases) or on merchants’ 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 is likely to be 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 is realistically only likely to give a lower bound of possible interchange fees that maximize consumer surplus.

As we noted above, in the existing literature on interchange fees, the ability of merchants to make sales to consumers who would not buy if not for the availability of credit facilities is notably absent. Among the few related papers to explicitly model the exten-

³In 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 consumer 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.

sion of credit are Chakravorti and Emmons (2003), Chakravorti and To (2007) and Bolt and Chakravorti (2008). We view these studies as been complementary to ours. Rather than focusing on interchange fees and their regulation, the main focus in these papers is to provide a positive theory of the pricing and usage of credit cards. In particular, they are more explicit in modelling the different roles of credit, whereas we simply assume credit allows additional purchases to be made a fraction of the time. On the other hand, they do not consider store credit. The existence of store credit in our framework means credit cards will not have any effect on aggregate consumption.

For instance, Chakravorti and To show how the ability of a merchant to accept credit cards of liquidity constrained consumers leads to inter-temporal business stealing, in which merchants that attract additional sales today by accepting credit cards do so at the expense of rival businesses whom might have attracted the same sales tomorrow. This effect complements the intra-temporal business stealing effect that is present in our paper (and in earlier papers, starting with Rochet and Tirole, 2002). On the other hand inter-temporal business stealing is not considered in our setting. The existence of store credit means liquidity constrained consumers do not need to delay their purchase if a merchant does not accept their credit card.

Another difference between our setting and these other models of credit cards is that to simplify the model we do not incorporate the interest paid by consumers on their credit balances into our model, which would introduce another price to be determined (i.e. the interest rate on these balances). According to Chakravorti and Emmons, in the U.S., over 75 percent of U.S. card issuer revenue is derived from cash-constrained consumers. They explore the interesting possibility that issuers would want to entice convenience users with low cardholder fees in order that some of these users would end up paying high interest on credit card debt (as revolvers). By ignoring interest income, we assume interest income is not relevant for the setting of interchange fees. This assumption is appropriate if issuing is competitive, so that there is no excess profit from credit card debt, and consumers can easily transfer their balances across issuers to obtain the best interest rate. The assumption is also appropriate to the extent consumers are rational and take into account any excess interest charges in their initial choice of issuer so that an issuer cannot benefit by shifting its cardholder fees to interest charges. In any case, in our model, interchange fees are set by card network operators (e.g. MasterCard and Visa) to maximize the volume of their credit card transactions, so these payment networks are not motivated by interest income

as they would be in a three-party scheme (or an issuer controlled system).

Closer to our own work is a recent paper by Bolt and Chakravorti who do endogenously solve for the optimal price structure of a card scheme, albeit of a three-party scheme, which sets prices to cardholders and merchants directly.⁴ In particular, they model the insurance role of credit cards and how this affects the pricing structure of credit cards. In their model, consumers participate in credit cards to insure themselves against three types of shocks — income shocks, theft and merchant match, which can affect the possibility of consumption. In our setting, the availability of store credit would again be the relevant alternative, rather than a loss of consumption.

The rest of the paper proceeds as follows. Section 2 presents our benchmark theoretical model in which the proportion of consumers holding credit cards is taken as given. In section 3 we determine the equilibrium level of the interchange fee, and develop a regulatory cap on interchange fee that raises (short-term) consumer surplus. Section 4 considers an extension of our benchmark model to endogenize the fraction of consumers that hold credit cards. Finally, section 5 offers some concluding remarks and possible extensions.

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 debit cards⁵) and credit cards. Credit cards are assumed to be more costly than cash but allow consumers to purchase on credit. They are held by a fraction x of consumers, where x is initially taken as given.⁶ 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 utility u_1 from consuming another unit

⁴Their analysis of price structure therefore does not correspond to our analysis of interchange fees (which are assumed to maximize the volume of credit card transactions) if bank margins in their setting vary with cardholder fees. Rather, assuming acquiring is perfectly competitive (as we do), their approach would correspond to the setting of interchange fees by an issuer-controlled card system.

⁵Since the focus of this paper is on the provision of consumer credit, we do not introduce any differences between pure payment technologies. Such differences are discussed at length in the literature mentioned in the introduction, e.g. Schmalensee (2002), Rochet and Tirole (2002) and Wright (2003).

⁶In section 4 we extend the model to allow consumers to decide whether to join the card network or not.

of the retail good (what we call “credit purchases”). We assume that merchants cannot bundle the two transactions and also cannot distinguish between “ordinary” and “credit” purchases. We assume that each consumer always has sufficient cash (or money in his current account) 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 $(\underline{c}_B, \bar{c}_B)$ where \underline{c}_B is sufficiently negative, such that cardholders will sometimes choose to use store credit even if cash can be used instead, and \bar{c}_B is positive but not too high (in comparison with $u_1 - \gamma$), such that consumers will always prefer to make the credit purchase even if they have to pay with store credit rather than not buy at all.⁷ The draw c_B is interpreted as a transaction cost that arises equally for ordinary purchases and credit purchases, and represents the *net* cost of using store credit rather than credit cards (or cash, if applicable) for these purchases, which is why we allow the possibility of negative draws of c_B . For instance, a negative draw of c_B could represent a situation whereby a cardholder needs to preserve his cash or credit card balance for some other contingencies and so values the use of store credit. Thus, consumers will choose between cash, store credit, and credit cards (if they hold credit cards) when making ordinary purchases.

Given the costs of store credit, accepting credit cards is a potential means for merchants 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 merchant (which is called the acquirer of the transaction) and c_I for the bank of the cardholder (the issuer of the card). 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 will prefer to use their credit cards rather than cash

⁷Without this assumption, credit card acceptance would not just add value by reducing the transaction costs of arranging credit; it would also lead to an increase in the volume of credit purchases (as in Chakravorti and Emmons, 2003, Chakravorti and To, 2007 and Bolt and Chakravorti, 2008), which would provide an additional factor affecting efficient interchange fees.

for ordinary purchases, which given our assumptions is socially wasteful.⁸ 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 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 :

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

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 thus have⁹:

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

These assumptions are not made to necessarily capture any intrinsic asymmetry between the nature of competition in issuing and acquiring, the existence of which is an empirical matter. Rather, the main reason for making these particular assumptions on bank competition is that they provide a simple setting in which a card network that seeks to maximize its members' profit also maximizes its volume of card transactions. This ensures our model applies to the past situation in which card payment schemes were controlled by members and interchange fees may have been set to maximize their members' joint profits, as well as the present situation in most countries in which MasterCard and Visa, as network operators, set interchange fees to maximize their volume of card transactions (and thereby their own network profits given that they earn most of their profit from small network fees charged to members on each card transaction).

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 assume that retailers cannot, or do not want to, charge different retail prices for different types of payments (e.g. for cash versus card payments). This follows Frankel

⁸We assume consumers and retailers face no costs of using cash (which can be thought of as a normalization). For ordinary purchases, any costs or benefits to consumers and retailers of using cash are assumed to be the same as those obtained from using credit cards (which can be thought of as an approximation).

⁹Issuer fees per cardholder are set to zero. We view this as a natural assumption given interchange fees are transaction based. Moreover, fixed fees are no longer so common for credit cards. For instance, based on the January 2009 Federal Reserve System six month survey of major U.S. issuers, only about 23% of the 153 recorded credit card plans had annual fees.

(1998), who suggested there is a general tendency for merchants to adhere to the setting of a single price regardless of the form of payment. Part of the reason for this is the no-surcharge rules adopted by the credit card systems. However, even in countries that have forced credit card systems to allow merchants to surcharge, the evidence suggests the vast majority of merchants do not surcharge credit cards (Chang et al., 2005). Moreover, those merchants that do surcharge may be doing so to extract additional surplus out of cardholders or to impose a “hidden fee” on their customers. It is seldom the case that the relative costs of accepting different forms of payments will be accurately reflected in retail prices (as would be predicted in our Hotelling model of merchant competition if surcharging were allowed).

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, he buys a first unit of the retail good (“ordinary purchase”), and pays it by cash or a credit card (if he has one).
- Finally, 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 , with full support on $[c_B, \bar{c}_B]$. Cardholders then select their mode of payment.

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 preferred to cash for ordinary purchases or not. A fraction $1 - xL_i$ of the time, consumers cannot use credit cards at merchant i since they do not hold a credit card or the merchant does not accept credit cards. For ordinary purchases, these consumers will use store credit if $c_B < 0$, which happens with probability $H(0)$; otherwise, with probability $1 - H(0)$ they will use cash. For credit purchases, which happens 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 $c_B < L_c f$, which happens with probability $H(L_c f)$; otherwise they will use either credit cards (if $L_c = 1$) or cash (if $L_c = 0$). For credit purchases, these cardholders will use store credit if $c_B < f$; otherwise they 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

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

or after simplifying

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

where

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

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

Since purchases always take place (ordinary purchases with probability one and credit purchases with probability θ), consumer surplus only depends on retail prices and transaction costs. Corresponding to each of the different types of possible transactions, retailer

i offers an expected consumer surplus (ignoring transportation costs) of

$$U_i = (1 - xL_i) \left(u_0 + \theta u_1 - (1 + \theta) p_i - \int_{c_B}^0 c_B dH(c_B) - \theta E(c_B) \right) \\ + xL_i \left(u_0 + \theta u_1 - (1 + \theta) p_i - \int_{c_B}^{L_c f} c_B dH(c_B) - \int_{L_c f}^{\bar{c}_B} L_c f dH(c_B) \right) \\ - \theta \left(\int_{c_B}^f c_B dH(c_B) + \int_f^{\bar{c}_B} f dH(c_B) \right)$$

or after rearranging

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

where

$$S(a) = L_c \int_f^0 (c_B - f) dH(c_B) - L_c \int_0^{\bar{c}_B} f dH(c_B) + \theta \int_f^{\bar{c}_B} (c_B - f) dH(c_B)$$

is the expected net consumer surplus per-cardholder from credit card usage at retailer i as a function of the interchange fee. Determining retailer i 's market share by finding the indifferent consumer in the normal way, we find¹⁰

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

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:

$$\phi(a) = S(a) - \Gamma(a) \quad (4) \\ = L_c \int_f^0 (c_B + c_S - (c + \pi)) dH(c_B) \\ + \theta \int_f^{\bar{c}_B} (c_B + c_S - (c + \pi)) dH(c_B) - L_c \int_0^{\bar{c}_B} (c + \pi) dH(c_B).$$

The first term reflects that there may be cost savings from using credit cards for ordinary purchases if this avoids the use of more costly store credit (i.e. if $f < 0$ and $c_B + c_S > c + \pi$). 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, which arises if $f < 0$ since we assumed credit cards were less efficient than cash.

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 .

¹⁰We assume that t is large enough so that each retailer always has a positive market share at equilibrium ($0 < s_i < 1$ for all i). We also assume that $u_0 + \theta u_1$ is large enough so that all the market is served ($U_i > t s_i$ for all i).

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

$$(1 + \theta)p_i^* = t + \gamma(1 + \theta) + (H(0) + \theta)c_S + x\Gamma(a)L_i + \frac{x}{3}\phi(a)(L_i - L_j). \quad (5)$$

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

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

Proof of Proposition 1: See the Appendix.

So as to maximize profit, each retailer chooses to accept cards if it increases its equilibrium 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 $\phi(a) \geq 0$, i.e. $S(a) \geq \Gamma(a)$.*

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

$$(1 + \theta)p(a) = t + \gamma(1 + \theta) + (H(0) + \theta)c_S + x\Gamma(a). \quad (7)$$

Interestingly, the equilibrium retail price may decrease in the interchange fee. A higher interchange fee shifts more transactions to credit cards from store credit which could potentially outweigh the effect of a higher cost of accepting credit cards if store credit is particularly costly to accept. To see this, note that

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

so that for m sufficiently below c_S , a higher interchange fee could lower retail prices.

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

$$U_{\text{cash}}(a) = u_0 + \theta u_1 - \frac{t}{4} - (1 + \theta)p(a) - \int_{c_B}^0 c_B dH(c_B) - \theta E(c_B)$$

and of consumers holding credit cards is

$$\begin{aligned} U_{\text{credit}}(a) = & u_0 + \theta u_1 - \frac{t}{4} - (1 + \theta)p(a) - \int_{c_B}^{L_c f} c_B dH(c_B) - \int_{L_c f}^{\bar{c}_B} L_c f dH(c_B) \\ & - \theta \left(\int_{c_B}^f c_B dH(c_B) + \int_f^{\bar{c}_B} f dH(c_B) \right). \end{aligned}$$

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

$$\begin{aligned} CS(a) &= xU_{\text{credit}}(a) + (1 - x)U_{\text{cash}}(a) \\ &= u_0 + \theta u_1 - \gamma(1 + \theta) - \frac{5t}{4} - \int_{c_B}^0 (c_B + c_S) dH(c_B) - \theta(E(c_B) + c_S) + x\phi(a). \end{aligned}$$

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

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

$$\frac{d\phi(a)}{da} = (L_c + \theta)(-a + c_S - c_A)h(c_I + \pi - a), \quad (8)$$

so $\phi(a)$ obtains a local maximum¹¹ at $a = a_T \equiv c_S - c_A$ (as illustrated in the figures). Note, however, a_T is not necessarily a global maximum for consumer surplus. This is because consumer surplus has a jump (down) at $a = a^* \equiv c_I + \pi$. This downward jump in consumer surplus is due to the fact that when $a > a^*$, the cardholder fee f becomes negative, and cardholders find it convenient to use their credit card for ordinary purchases (thereby replacing some efficient use of cash). This change in consumer surplus, which equals $-x \int_0^{\bar{c}_B} (c + \pi) dH(c_B)$, complicates the analysis of consumer surplus maximization, and introduces three regimes.

In regime 1, illustrated in figure 1, $a_T = c_S - c_A$ is less than $a^* = c_I + \pi$, and therefore consumer surplus is maximized for $a = a_T$. This corresponds to the situation where credit card transactions are more costly to provide than retailer provided store credit:

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

¹¹Equation (8) shows that $d\phi/da$ is positive for $a < a_T$ and negative for $a > a_T$, implying a local maximum of ϕ for $a = a_T$.

In regime 1, regulation to maximize consumer surplus would imply making credit cards *more* expensive for consumers than using cash (or debit cards). This together with the idea that credit card transactions are less efficient than store credit does not seem plausible except perhaps with respect to the very largest retailers.

The more realistic cases are captured by regimes 2 and 3 (illustrated in figures 2 and 3), which correspond to the reverse situation in which $c_S > c + \pi$ and thus $a_T > a^*$. In regime 2, the incidence of convenience users is not so large (so that a_T is still the maximum of consumer surplus). By contrast in regime 3, the maximum of consumer surplus is obtained at a^* .

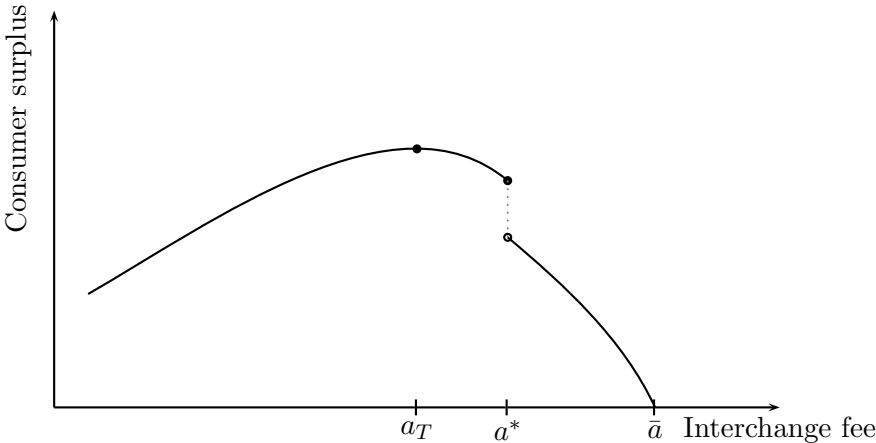


Figure 1: First regime. Consumer surplus is maximized at $a = a_T < a^*$.

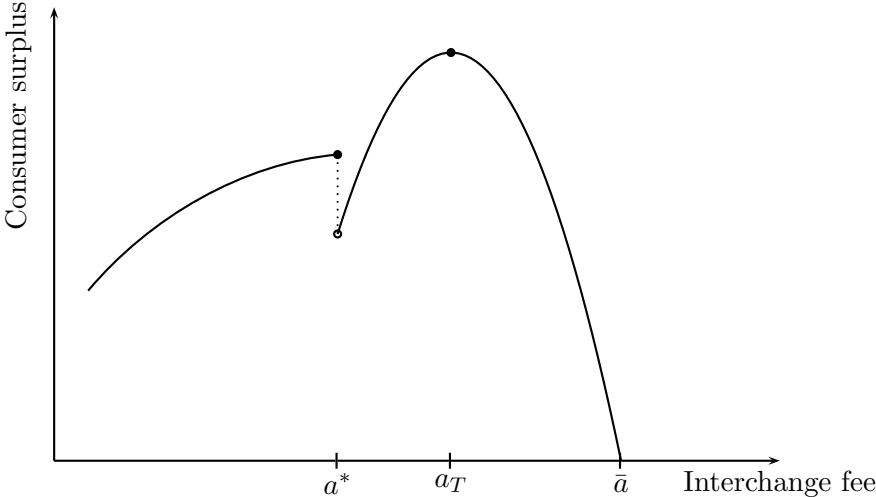


Figure 2: Second regime. Consumer surplus is maximized at $a = a_T > a^*$.

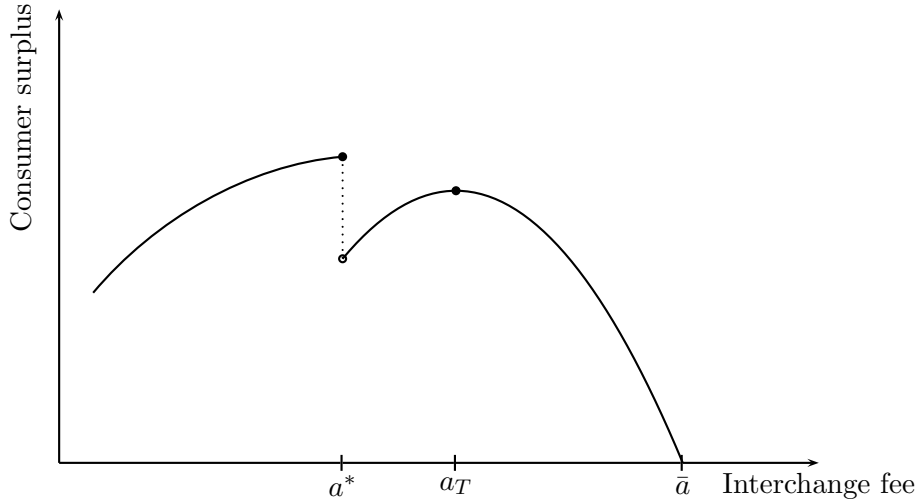


Figure 3: Third regime. Consumer surplus is maximized at $a = a^*$.

Recall that the volume of credit card transactions, and thus the profit of the monopoly card network, increases with a . This implies that a monopoly card network will set an interchange fee equal to the maximum level that is compatible with merchant acceptance of payment cards. This is the point where $\phi(a) = 0$, which we define as \bar{a} .

In order to keep the analysis interesting, we assume that $\phi(a) > 0$ for some $f < 0$ sufficiently close to 0 (i.e. corresponding to a just above a^*). From (4), this is equivalent to assuming

$$\theta \int_0^{\bar{c}_B} (c_B + c_S) dH(c_B) > (1 + \theta) \int_0^{\bar{c}_B} (c + \pi) dH(c_B), \quad (9)$$

so that the surplus created from credit cards (the cost saving) is higher than their additional cost to users in the limit as f tends to zero from below. If (9) is not assumed, then the monopoly card network would never want to set an interchange fee that leads to $f < 0$, which would be an uninteresting case from a regulatory point of view. In other words, without (9) we would have $\bar{a} \leq a^*$. With (9), \bar{a} is always greater than both a_T and a^* as illustrated in figures 1-3.

3.3 Policy Implications

As noted in the previous section, 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, which given (9), is always greater than both a_T and a^* . Therefore, if competition authorities aim at maximizing (short-run) consumer surplus (or equivalently $\phi(a)$), they

will always find the privately optimal interchange fee excessive.¹² A regulatory cap on interchange fees can therefore be justified in the setting of this paper but the appropriate level of the cap is not always the same: it is a_T in regime 1 and regime 2 and a^* in regime 3. Which of these regimes is relevant depends on the magnitude of the difference between c_S (the retailer's cost of providing store credit) and $c + \pi$ (the total user cost of credit transactions). We denote this difference by $\delta \equiv c_S - c - \pi$. 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, depending on the value of δ :*

- a) If $\delta < 0$ (regime 1) or if $(c + \pi)(1 - H(0)) \leq (1 + \theta) \int_{-\delta}^0 (c_B + \delta) dH(c_B)$ (regime 2), the regulatory cap should be $a_T = c_S - c_A$.*
- b) Otherwise (regime 3), the regulatory cap should be $a^* = c_I + \pi$.*

Proof of Proposition 3: See the Appendix.

The proposition implies that the interchange fee $a_T = c_S - c_A$ achieves the global maximum of consumer surplus provided (i) $\delta < 0$, which corresponds to regime 1 or (ii) if $\delta > 0$ provided that

$$(1 + \theta) \int_{-\delta}^0 (c_B + \delta) dH(c_B) \geq (c + \pi)(1 - H(0)), \quad (10)$$

which corresponds to regime 2. Recall regime 1 arises when credit card transactions are more costly to provide than retailer provided store credit, which does not seem realistic except perhaps for the very largest retailers. Regime 2 arises when the efficiency gains from promoting the use of credit cards over store credit dominate, which is most likely to occur for smaller retailers for which store credit is particularly costly or more generally when credit cards are relatively more efficient. The condition (10) comes from requiring consumer surplus at $a = a_T$ is higher than at $a = a^*$. In this case, setting $a = a_T$ implies $f < 0$ which induces some inefficient use of credit cards rather than cash for ordinary

¹²This 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).

purchases as compared to setting $a = a^*$. This explains the right-hand-side of (10). On the other hand, when $c_B < 0$, consumers will want to use credit cards rather than less efficient store-credit provided f is sufficiently negative. This is the cost saving from setting $a = a_T$ rather than $a = a^*$, as measured by the left-hand-side of (10). Regime 3 arises instead when the deadweight loss from excessive use of credit cards by convenience users dominates the efficiency gains from reducing the use of store credit.

In either case, lowering interchange fees from the private maximum to a_T unambiguously raises consumer surplus. The increase in consumer surplus from regulation equals

$$\Delta CS^T(\delta) = x \left(L_c \int_{-\delta}^0 (c_B + \delta) dH(c_B) + \theta \int_{-\delta}^{\bar{c}_B} (c_B + \delta) dH(c_B) - L_c \int_0^{\bar{c}_B} (c + \pi) dH(c_B) \right)$$

when a is regulated based on retailer avoided cost ($a = a_T$). In contrast, the increase in consumer surplus equals

$$\Delta CS^*(\delta) = x\theta \int_0^{\bar{c}_B} (c_B + \delta) dH(c_B)$$

when a is regulated based on issuer cost ($a = a^*$). The assumption in (9) ensures both regulations unambiguously increase consumer surplus.

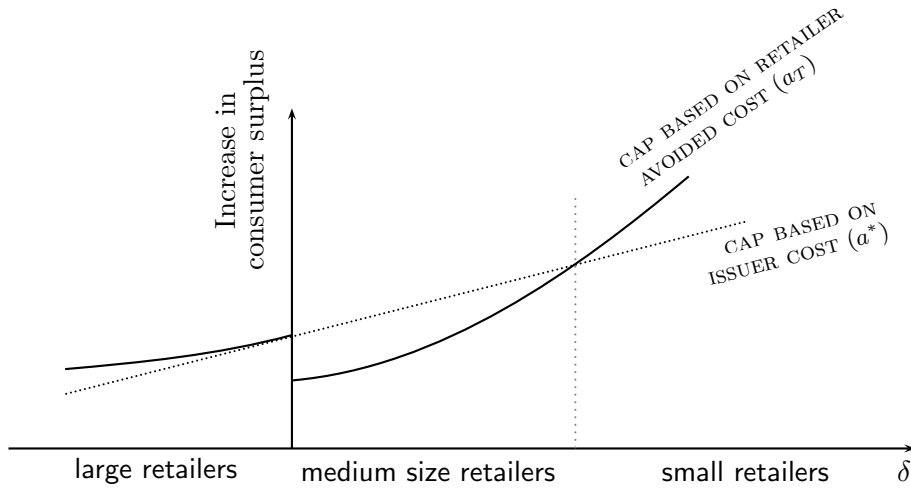


Figure 4: Gains from regulation.

Figure 4 plots these gains from regulation.¹³ It shows how the gains from regulation relate to the three regimes defined in proposition 3. The incremental cost parameter δ can be expected to be smaller for large retailers than for small retailers reflecting that it is likely to be more costly for small retailers to accept store credit.

Consider the situation facing a regulator in practice that is focused on maximizing short-term consumer surplus but does not know the true values of the parameters δ , θ etc. As argued above, the regulator should be able to rule out regime 1 applying (except perhaps for the very largest merchants) since it relies on store credit being more efficient than general purpose credit cards. This leaves the regulator the problem of determining which of regimes 2 or 3 applies. A conservative approach to regulation would be to cap the interchange fee at a_T . As can be seen from figures 2 and 3, this always raises consumer surplus compared to the unregulated outcome, and would never be too low. In contrast, the interchange fee a^* (which is based on the issuers' cost) represents a lower bound for desirable interchange fees if either of the more realistic regimes 2 or 3 hold, since setting the interchange fee below a^* would always lower consumer surplus. This could be used as a check to make sure the calculated a_T is not too low.¹⁴ Thus, regulations which base interchange fees on issuer costs could be rationalized in our framework, although realistically, only as a lower bound on what might be desirable for consumers.

It seems feasible to construct the regulatory cap a_T in practice. 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). Note, however, that a_T will differ for different classes of retailers (reflecting their different values of c_S) in contrast to a^* which only depends on the issuer.

The regulatory cap a_T proposed above is conceptually the same as Rochet and Tirole's (2008) "tourist test threshold", i.e. the maximum level of the interchange fee such that the

¹³Figure 4 is drawn to take into account the various properties of the functions $\Delta CS^*(\delta)$ and $\Delta CS^T(\delta)$; namely, that $\Delta CS^T(\delta)$ is increasing in δ except at $\delta = 0$ (where it has a downward jump), $\Delta CS^*(\delta)$ is linear and increasing in δ , $\Delta CS^*(\delta) < \Delta CS^T(\delta)$ for $\delta < 0$, $\Delta CS^*(\delta) = \Delta CS^T(\delta)$ at $\delta = 0$, and $d^2\Delta CS^T(\delta)/d\delta^2 > 0$ everywhere except at $\delta = 0$.

¹⁴For example, since $a^* > 0$, this implies the regulated credit card interchange fee should be strictly positive.

merchant fee $m = c_A + a$ is less than the cost c_S of the relevant alternative technology for the merchant. This threshold has recently been considered by the European Commission as a possible way (from a regulatory perspective) to set Visa’s interchange fees in Europe. The tourist test threshold generalizes the idea of the Baxter interchange fee (Baxter, 1983), which was developed in the context of a perfectly competitive banking sector, and corresponds to Farrell’s (2006) “Merchant Indifference Criterion”. In contrast to these earlier contributions, the proposed interchange fee here is worked out specifically for credit cards, where merchant indifference relates to the alternative cost to the merchant of offering credit directly.

4 Endogenizing 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 in the first place. Prior membership decisions are potentially important since if interchange fee are set at very low levels (for instance, at zero), then consumers can be expected to face higher card fees since they would pay the full cost of issuing cards (as well as any issuing margin) at which point some 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 3, they would always prefer to use cash instead. 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 maximize consumer surplus (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 would otherwise not hold a credit card.

The framework 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 $(-\infty, \bar{v}]$

for some $\bar{v} > 0$. 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 analysis of retailers' decisions (which treat x as given), and the various surplus expressions are similar to before but with some simplifications. In particular, store credit is not used for ordinary purchases since cash is always preferred to store credit and it is only used for credit purchases by cardholders when $f > 0$. The expression for total user surplus (4) therefore simplifies since the first term is zero, the second term only arises for $f > 0$, and the third term simplifies since when $f < 0$, all cardholders will use credit cards rather than store credit. This implies

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

which is now independent of a if $a > a^*$ (so that $L_c = 1$):

$$a > a^* \Rightarrow \phi(a) \equiv \theta(E(c_B) + c_S) - (1 + \theta)(c + \pi).$$

Assumption (9) in section 3 implies $\phi(a) > 0$. Without making this assumption, retailers would reject cards whenever $a > a^*$. Taking this into account, the card network would always set the interchange fee at most at a^* and there would be no rationale for any regulation to lower interchange fees.

Given (9), 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, 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

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

so consumers will hold cards if

$$v > L_c f - \theta \int_{(1-L_c)f}^{\bar{c}_B} (c_B - f) dH(c_B)$$

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

$$CS(a) = \int_{L_c f - \theta \int_{(1-L_c)f}^{\bar{c}_B} (c_B - f) dH(c_B)}^{\bar{v}} (v + \phi(a)) d\Psi(v).$$

As before, consumer surplus has a downward jump at $a = a^*$ corresponding to the use of credit cards for ordinary purchases as opposed to cash. This change in consumer surplus equals $-x(a^*)(c + \pi)$.

If the interchange fee which maximizes consumer surplus is less than or equal to a^* (i.e. regime 1), then it remains equal to $c_S - c_A$ as in the benchmark model. This reflects that with $a \leq a^*$, cardholder fees are non-negative, and consumers will still choose between store credit and credit cards depending on their draw of c_B as in the benchmark model. We prove this result formally in the Appendix (see the proof of Proposition 4).

$CS(a)$ behaves slightly differently for $a > a^*$. Given our assumption that store credit is always costly to use from consumers point of view, credit cards will be used wherever possible. Thus, consumer surplus is maximized by setting an interchange fee which gets the optimal level of credit card membership, taking into account the costs of excessive card usage for ordinary purchases, and the efficiency benefits of credit cards for credit purchases. We show in the Appendix that the interchange fee which maximizes consumer surplus in this case is

$$\hat{a}_T = \frac{\theta}{1 + \theta} c_S - c_A. \quad (11)$$

which is lower than the corresponding interchange fee a_T in the benchmark model.

The purpose of the interchange fee \hat{a}_T 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 off by c_A due to the excessive usage of cards. Following this interpretation, \hat{a}_T can also be written:

$$\hat{a}_T = \frac{\theta}{1 + \theta} (c_S - c_A) + \left(1 - \frac{\theta}{1 + \theta} \right) (-c_A). \quad (12)$$

Thus, the interchange fee \hat{a}_T can be interpreted as a weighted average of a_T (i.e. the net avoided costs of credit as in the benchmark model) and net avoided cost of cash (i.e. the standard tourist-test or Baxter interchange fee) which is $-c_A$ given that we have assumed

the cost of accepting cash is the same as accepting credit cards for ordinary purchases (and that the social cost of cash is assumed to be zero).¹⁵

Recall from section 3 that we defined $\delta = c_S - c - \pi$ so that $\delta > 0 \Leftrightarrow a_T > a^*$. The equivalent in the current setting is

$$\hat{\delta} = \theta c_S - (1 + \theta)(c + \pi) = (1 + \theta)(\hat{a}_T - a^*)$$

so that $\hat{\delta} > 0 \Leftrightarrow \hat{a}_T > a^*$. Notice if $\delta < 0$ (regime 1) then we have $\hat{\delta} < 0$ so that $CS(a)$ is decreasing in a in the neighborhood of a^* . Figure 1 still applies. Alternatively, if $\hat{\delta} > 0$ then we have $\delta > 0$ so that $CS(a)$ is increasing in a in the neighborhood of a^* . We therefore have the equivalent of regime 2 and regime 3 in figures 2 and 3, but with a_T replaced by \hat{a}_T . Finally, we have a new regime (regime 4) which arises when $\delta > 0$ and $\hat{\delta} < 0$, in which case $CS(a)$ is increasing in a for $a \leq a^*$ and $CS(a)$ is decreasing in a for $a > a^*$, so that consumer surplus is maximized at a^* . Figure 5 illustrates this new regime. Proposition 4 summarizes our results.

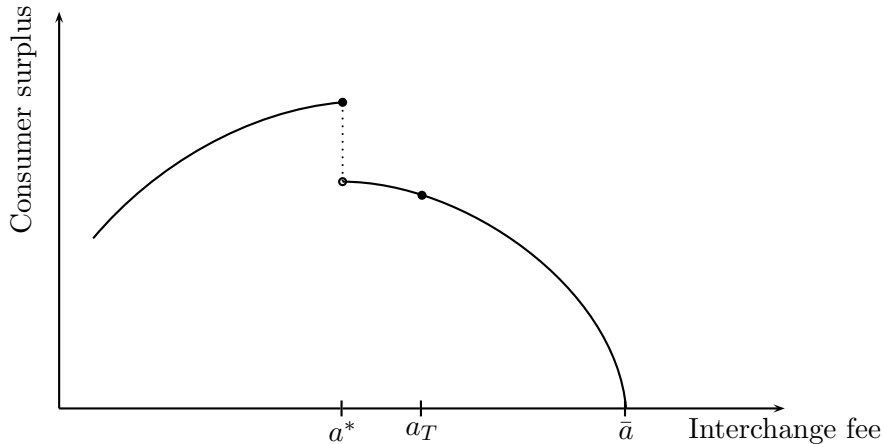


Figure 5: Fourth regime. Consumer surplus is maximized at $a = a^*$.

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 three cases must be considered depending on the values of δ and $\hat{\delta}$:*

¹⁵Indeed, if b_S measures the additional transactional benefit to merchants of accepting credit cards as opposed to cash, then it is straightforward to check that the formula (12) becomes $\hat{a}_T = \frac{\theta}{1+\theta}(c_S - c_A) + \left(1 - \frac{\theta}{1+\theta}\right)(b_S - c_A)$.

a) If $\delta \leq 0$ (regime 1) then the regulatory cap should be $a_T = c_S - c_A$.

b) If $\hat{\delta} > 0$ and

$$\int_{-\theta E(c_B) - \hat{\delta}}^{-\theta E(c_B)} (v + \theta E(c_B) + \hat{\delta}) d\Psi(v) \geq \int_{-\theta E(c_B)}^{\bar{v}} (c + \pi) d\Psi(v) \quad (13)$$

(regime 2), the regulatory cap should be $\hat{a}_T = \frac{\theta}{1+\theta} c_S - c_A$.

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

Proof of Proposition 4: See the Appendix.

The proposition implies that the same interchange fee cap as before ($a_T = c_S - c_A$) achieves the global maximum of consumer surplus if $\delta < 0$, which corresponds to regime 1. Alternatively, if $\hat{\delta} > 0$ and provided (13) holds, which corresponds to regime 2, $\hat{a}_T = \frac{\theta}{1+\theta} c_S - c_A$ maximizes consumer surplus. In all other cases, the cost based regulatory cap $a^* = c_I + \pi$ maximizes consumer surplus.

In case that $\hat{\delta} > 0$, increasing the interchange fee from a^* to \hat{a}_T will lead to more consumers to hold credit cards, thereby reducing the costs associated with using store-credit. At the same time, all cardholders will now use their cards inefficiently for ordinary transactions. This is the trade-off captured by the inequality in (13). Given it may be hard to measure θ with any confidence or to evaluate whether (13) 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) for large θ , it approximates the maximization of (short-term) consumer surplus; and (iii) yet is never too low from the point of view of maximizing (short-term) consumer surplus. Alternatively, to the extent regime 2 is thought to be the most empirically relevant case, our results here motivate the use of (12), which is a weighted average of two interchange fees — the regular Baxter interchange fee (or the tourist-test generalization of it) in which the alternative to credit cards is cash and the credit card interchange fee $a_T = c_S - c_A$ in which the alternative to credit cards is store credit, with the weights being the proportion of each type of transaction (i.e. ordinary purchases versus purchases where credit is required).

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 ordinary purchases and the importance of getting cardholders to internalize retailers' avoided costs arising from their credit card usage. Starting from a setting in which an unregulated card network always sets the interchange fee too high, in this paper we show 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.

The regulatory cap we propose is conceptually the same as the well known Baxter interchange fee, and its more recent generalization, which Rochet and Tirole (2008) have called the "tourist test threshold". Indeed, the tourist test threshold has recently been proposed by the European Commission as an appropriate way to set interchange fees in Europe. Our results, which emphasize the retailers' net avoided cost from not having to provide credit themselves, can help regulators properly interpret the tourist test threshold and apply it in the case of credit card interchange fees. Allowing for cardholder membership to be endogenous, the regulatory cap implies credit card interchange fees would be set as a weighted average of the retailers' net avoided cost from not having to accept cash and the retailers' net avoided cost from not having to provide credit, with the weights being the proportion of each type of transaction (ordinary purchases versus purchases where credit is required). We also discussed the idea that the regulatory cap would vary with the size (or other characteristics) of the merchant which would reflect the extent of cost savings from accepting credit cards compared to providing store credit.

The possibility that the regulator's preferred interchange fee predicted by our theory may still be moderately high reflects the potentially high merchant benefits of accepting cards in these circumstances, which cardholders may not otherwise internalize. Put differently, if the interchange fee is set too low (say at zero) so that consumers were sometimes not willing to hold or use cards for such transactions, then competing merchants will instead tend to rely more on store credit (or other forms of credit) so as to attract business. Quite plausibly, the additional costs to society of making greater use of these more expensive forms of credit will outweigh any benefit from encouraging debit rather than credit card transactions for ordinary purchases. Some excessive use of credit cards may be unavoidable given merchants cannot easily observe if credit is needed or not by

their customers. This seems no different from the fact merchants that offer interest-free installment plans to their customers, will sometimes (perhaps often) end up offering these plans to consumers who actually do not need them.

One important direction for future research is to extend our model to allow retailers to offer different prices (through the use of discounts, interest-free periods and rewards) when consumers make use of store credit. In not allowing this possibility, we had in mind such price differentials being exogenous for each retailer, determined perhaps by a third party or due to the inability (or unwillingness) of retailers to set differential prices based on a consumer's choice of payment technology. If retailers were able and willing to discriminate based on the use of store credit, they may be able to induce consumers to use credit cards and store credit efficiently. However, any individual retailer would still not be able to (or have the incentive to) use its differential pricing to get consumers to make the right decision about whether to hold a credit card in the first place. As such, our results from section 4, in which card membership is endogenized, are likely to carry over to the case retailers can price discriminate in this way.

Another possible direction for future research is to extend our model to allow for competing payment networks. By adapting the arguments of Guthrie and Wright (2007) one should be able to show similar results to those shown here still hold when there is competition between multiple card payment networks. The idea is that competing networks will seek to choose an interchange fee somewhere between the one maximizing total user surplus and that chosen by a monopoly network, depending on the extent to which consumers choose to hold multiple cards, and so these may still be too high.

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 required 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 helps explain why 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 setting of interchange fees that promote the widespread use of general purpose credit cards.

Appendix

Proof of Proposition 1:

From the text we have

$$\pi_i = ((1 + \theta)(p_i - \gamma) - (H(0) + \theta)c_S - x\Gamma(a)L_i) \left(\frac{1}{2} + \frac{(1 + \theta)(p_j - p_i)}{2t} + xS(a) \frac{L_i - L_j}{2t} \right)$$

Differentiating we get

$$\frac{2t}{1 + \theta} \frac{\partial \pi_i}{\partial p_i} = t + (1 + \theta)(p_j - p_i) + xS(a)(L_i - L_j) - (1 + \theta)(p_i - \gamma) + (H(0) + \theta)c_S + xL_i\Gamma(a).$$

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

$$(1 + \theta)(2p_i - p_j) = t + \gamma(1 + \theta) + (H(0) + \theta)c_S + x(S(a)(L_i - L_j) + L_i\Gamma(a)).$$

Solving for p_i , we obtain formula (5):

$$(1 + \theta)p_i = t + \gamma(1 + \theta) + (H(0) + \theta)c_S + x\Gamma(a)L_i + \frac{x}{3}\phi(a)(L_i - L_j). \quad (14)$$

Substituting (14) into (3) and using (4) implies formula (6).

Proof of Proposition 3:

This follows from the comparison of $\phi(a_T)$ and $\phi(a^*)$. Recall the expression for $\phi(a)$ in (4). Evaluating $\phi(a)$ at a_T using that $\delta = c_S - c - \pi = a_T - a^*$ we have

$$\phi(a_T) = L_c \int_{-\delta}^0 (c_B + \delta) dH(c_B) + \theta \int_{-\delta}^{\bar{c}_B} (c_B + \delta) dH(c_B) - \int_0^{\bar{c}_B} (c + \pi) L_c dH(c_B),$$

while evaluating $\phi(a)$ at a^* we have

$$\phi(a^*) = \theta \int_0^{\bar{c}_B} (c_B + \delta) dH(c_B).$$

The difference is

$$\phi(a_T) - \phi(a^*) = L_c \int_{-\delta}^0 (c_B + \delta) dH(c_B) + \theta \int_{-\delta}^0 (c_B + \delta) dH(c_B) - L_c \int_0^{\bar{c}_B} (c + \pi) dH(c_B). \quad (15)$$

When $\delta < 0$, $L_c = 0$ and substituting this into (15) we have

$$\phi(a_T) - \phi(a^*) = -\theta \int_0^{-\delta} (c_B + \delta) dH(c_B) > 0$$

(since $c_B + \delta \leq 0$ when c_B belongs to $[0, -\delta]$). Thus ϕ is maximized at $a = a_T$.

When $\delta > 0$, $L_c = 1$ and substituting this into (15) we have

$$\phi(a_T) - \phi(a^*) = (1 + \theta) \int_{-\delta}^0 (c_B + \delta) dH(c_B) - (c + \pi)(1 - H(0)).$$

This establishes Proposition 3.

Proof of Proposition 4:

First note that differentiating $CS(a)$ for $a \leq a^*$ implies

$$\frac{dCS(a)}{da} = \left(\begin{array}{l} \theta(1 - H(f))^2 \psi \left(-\theta \int_f^{\bar{c}_B} (c_B - f) dH(c_B) \right) \\ + \left(1 - \Psi \left(-\theta \int_f^{\bar{c}_B} (c_B - f) dH(c_B) \right) \right) h(-a + c_S - c_A) \end{array} \right) \theta(-a + c_S - c_A) \quad (16)$$

and differentiating $CS(a)$ for $a > a^*$ implies

$$\frac{dCS(a)}{da} = (1 + \theta)(\theta c_S - (1 + \theta)(a + c_A)) \psi((1 + \theta)f - \theta E(c_B)). \quad (17)$$

Case (i). Suppose $\delta \leq 0$ so that $\hat{\delta} < 0$. For $a \leq a^*$, (16) implies $dCS(a)/da > 0$ if $a < c_S - c_A$ and $dCS(a)/da < 0$ if $a > c_S - c_A$. For $a > a^*$, (17) implies $dCS(a)/da < 0$ given $\hat{\delta} < 0$. Since CS has a downward jump at a^* , this implies the global maximum of CS is at $a = c_S - c_A$.

Case (ii) Suppose $\hat{\delta} \geq 0$ so that $\delta > 0$. For $a > a^*$, (17) implies $dCS(a)/da > 0$ for $a < \frac{\theta}{1+\theta}c_S - c_A$ and $dCS(a)/da < 0$ for $a > \frac{\theta}{1+\theta}c_S - c_A$. For $a \leq a^*$, (16) implies $dCS(a)/da > 0$. Since CS has a downward jump at a^* , this implies the global maximum of CS is either at a^* or at $a = \frac{\theta}{1+\theta}c_S - c_A$. Comparing $CS(\hat{a}_T)$ when $\hat{a}_T > a^*$ with $CS(a^*)$ gives the inequality (13).

Case (iii) Suppose $\delta > 0$ and $\hat{\delta} < 0$. For $a \leq a^*$, (16) implies $dCS(a)/da > 0$, while for $a > a^*$, (17) implies $dCS(a)/da < 0$. Since CS has a downward jump at a^* , this implies the global maximum of CS is at $a = a^*$.

This establishes Proposition 4.

References

- BAXTER, W.P. (1983) "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics*, 26: 541-88.
- BOLT, W. AND S. CHAKRAVORTI (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.
- CHANG, H., D. EVANS, AND D.G.SWARTZ (2005) “The Effect of Regulatory Intervention in Two-Sided Markets: An Assessment of Interchange-Fee Capping in Australia,” *Review of Network Economics*, 4: 328 - 358.
- CHAKRAVORTI, S. AND W. EMMONS (2003) “Who pays for credit cards?,” *Journal of Consumer Affairs*, 37: 208-230.
- CHAKRAVORTI, S. AND T. TO (2007) “A Theory of Credit Cards,” *International Journal of Industrial Organization*, 25: 583-95.
- EVANS, D AND R. SCHMALENSEE (2005) “Paying with plastic: The digital revolution in buying and borrowing” 2nd ed, MIT Press, Cambridge Mass.
- FARRELL, J. (2006) “Efficiency and Competition between Payment Instruments,” *Review of Network Economics*, 5: 26-44.
- FRANKEL, A. (1998) “Monopoly and competition in the supply and exchange of money,” *Antitrust Law Journal*, 66: 331-361.
- 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.