Optimal card payment systems

Julian Wright[*]

Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand

Abstract

This paper presents a model of a card payment system to address the pricing and rules that govern such systems. It evaluates the social optimality of privately set interchange fees and the adoption of a rule by payment systems to prevent merchants surcharging for card transactions using two extremes of merchant pricing—monopolistic pricing and perfect competition. Both types of merchant pricing constrain the ability of card schemes to use interchange fees and the no-surcharge rule in anticompetitive ways, although for quite different reasons. The positive role of the no-surcharge rule in preventing excessive merchant surcharging is also highlighted.
© 2002 Elsevier B.V. All rights reserved.

JEL classification: G21; L31; L42

Keywords: Payment systems; Banks; Credit; Debit cards; Interchange fees

1. Introduction

Recently, the pricing and rules governing card payment systems such as those offered by MasterCard and Visa have come under attack from policymakers in a number of jurisdictions. In the United Kingdom, a report commissioned by the government (see Cruickshank, 2000) suggested a number of problems with the banking sector there, including the high level of fees set between banks in payment systems (interchange fees). In response, the government has given its competition authority new powers to regulate payment systems, including interchange fees (HM Treasury, 2001). The European Commission has also been investigating interchange fees and the rules set by the members of card associations (see European Commission, 2000, 2001). In Germany, a heated debate has arisen concerning the banking system's intention to adopt a common interchange fee for all debit card payments. Outside Europe, antitrust cases are pending

---

[*] Tel.: +64-9373-7999; fax: +64-9373-7427.

E-mail address: jk.wright@auckland.ac.nz (J. Wright).

against MasterCard and Visa in the United States, and in Australia the central bank has moved to regulate card associations (Reserve Bank of Australia, 2002).

Despite the strong public interest in payment systems, there is only a small body of academic research that policymakers can draw on to analyze the pricing and rules of such systems. This is all the more surprising given the dramatic increase in usage of card payments in the last decade. Based on data from the Bank for International Settlements and the European Central Bank, Krueger (2001) calculates debit and credit cards in 12 industrial countries (Belgium, Canada, Denmark, Finland, France, Germany, Italy, Netherlands, Sweden, Switzerland, United Kingdom, United States) which grew from 9.3 billion transactions in 1987 to 33.7 billion transactions in 1998.

This paper builds on the existing literature by providing a model that provides new insights into the role of the rules and pricing that govern such payment schemes. An open scheme (such as the debit and credit card systems offered by MasterCard and Visa) involves the issuance of cards by issuing banks (issuers) to cardholders. These cardholders use their cards to make purchases with merchants which have been signed up by acquiring banks (acquirers). Such schemes differ from closed schemes (such as that offered by American Express) in that they allow, and in fact encourage, competing banks and other institutions to issue cards and acquire merchants. Since for a typical transaction, cardholders pay a different institution from that which receives payment from the merchant, the card association plays the role of clearing payments between issuers and acquirers. In addition, the card association sets an interchange fee, which is a payment made by acquirers to issuers on each transaction. Apart from the setting of the interchange fee, two other rules set by card associations have attracted the attention of policymakers: (i) The no-surcharge rule (often called the 'no-discrimination' rule in Europe), which says that merchants cannot set a surcharge on goods purchased using cards (as opposed to other forms of payment) and (ii) the honour-all-cards rule, which says that merchants must accept all legitimate cards within a card system, regardless of the particular institution that issues them.

In the context of an open payment scheme, we study the welfare implications of the lack of merchant surcharging. We also examine how the level of the interchange fee between issuers and acquirers affects consumers' and merchants' behavior, discussing the optimal setting of such an interchange fee from a private and social perspective.

Previous studies have focused on the positive role interchange fees play in reallocating funds between merchants and cardholders, so as to best align private costs and benefits with those of the network as a whole. Baxter (1983) provides the first formal analysis. Baxter notes that to maximize surplus in the card network, the sum of cardholder and merchant benefits for the marginal transaction should equal the sum of the respective marginal costs. This can be achieved if cardholders are charged a fee equal to the sum of the issuing and acquiring marginal costs, less the merchants' transactional benefits. In this case, cardholders will face the joint transactional benefits and joint costs of using cards. However, with different banks competing for cardholders and merchants, there is no reason to expect that without an appropriately set interchange fee, issuing banks will set their prices in this way. By setting an interchange fee equal to the difference between merchant benefits and the acquirers' marginal cost, perfect

competition between issuers will cause issuers to set cardholder fees at the right level. We refer to this interchange fee as the Baxter fee.

Schmalensee (2002) extends Baxter's analysis by dropping Baxter's assumption of perfect competition between issuers and between acquirers. He derives several implications for the resulting joint profit maximizing interchange fees, including a decomposition of the optimal interchange fee into a term that depends on the difference in demand elasticities across cardholders and merchants, and a term that depends on the difference in costs across issuers and acquirers.

Rochet and Tirole (2002) are the first to study the welfare implications of interchange setting and merchant surcharging in a model of a card payment system in which consumers' and merchants' decisions are derived from first principles. In their model, merchants compete according to Hotelling competition. They obtain two key results. Firstly, assuming surcharges are not allowed, the payment card association's optimal interchange fee is either the same as the socially optimal one, or is higher, thus leading to an overprovision of card payment services. Secondly, when surcharges are allowed, the interchange fee becomes neutral, there is an underprovision of cards, while the effects on social welfare are ambiguous. [1]

In this paper, we take Rochet and Tirole's framework and consider two different extremes of merchant pricing–monopolistic pricing and perfect competition. [2] By considering these extremes of merchant pricing, unambiguous results on the welfare effects of the no-surcharge rule and the privately chosen interchange fee are obtained. These different approaches to merchant pricing, as well as the possibility of a membership fee faced by cardholders, are also used to illustrate some important features of payment systems which have previously been overlooked.

In the case where merchants have local monopolies but are free to surcharge, we show they will do so excessively, so as to extract surplus from inframarginal cardholders. The result will be too few cardholders and too little card usage. Under these circumstances, both the banks and a regulator concerned with maximizing total surplus prefer the adoption of a no-surcharge rule. Moreover, both will set the interchange fee at the socially optimal level. Since the socially optimal interchange fee turns out to be equal to the Baxter fee, our model provides a justification for Baxter's fee from both a private and social perspective. When merchants are monopolists, but cannot surcharge, they will only accept cards when they receive transactional benefits that exceed the merchant fee they are charged. This means the merchant fee is constrained by the

---

[1] Wright (2001) extends Rochet and Tirole's model to take into account merchant heterogeneity, but focuses on interchange fee determination and not the no-surcharge rule. In addition to these papers, Evans and Schmalensee (1999), Chang and Evans (2000) and Chakravorti and Shah (2001) provide useful surveys of the historical and institutional details of payment systems, Frankel (1998) puts forward arguments for why interchange fees should be set at zero, Small and Wright (2001) model the consequences of requiring interchange fees be set bilaterally in existing credit card schemes, while Gans and King (2002) provide a general characterization of conditions under which interchange fees are irrelevant.

[2] Written contemporaneously with this paper, Schwartz and Vincent (2002) obtain mixed results on the welfare effects of the no-surcharge rule in a model in which merchants are monopolists. However, they do not allow consumers to decide whether to use cash or cards, and so in this sense their analysis is incomplete, especially since they also assume cardholders and merchants receive no transactional benefits from the use of cards in deriving their welfare results.

transactional benefits that merchants obtain, at which point the interchange fee will play the role of passing onto consumers the costs of the acquiring side of the market and the transactional benefits that merchants obtain.

Our model also provides an answer to the question, why negotiations along the lines of Coase (1960), in which consumers and merchants negotiate prices that internalize any externalities that exist between them, will not take place when surcharging is allowed? Before making payments with a payment card, consumers first have to decide whether to subscribe to the card network or not. As the benefits of holding a card depend on the value obtained ex post, and as each individual merchant can safely ignore the effect of its decisions on consumers' decision to hold a card (since it is just one of millions of merchants that consumers have the chance of dealing with), individual merchants have no incentive to negotiate agreements ex post which lead to the efficient membership by cardholders ex ante. [3]

This aspect of merchant behavior is modeled by assuming that consumers face a fee to join the card network, and that monopolistic merchants set their prices after consumers have decided whether to join the card network or not. Thus, the ability for consumers and merchants to engage in Coasian style bargains will be limited. The no-surcharge rule and an appropriately set interchange fee turn out to play an important efficiency enhancing role. Without the no-surcharge rule in place, merchants end up destroying the card network altogether through excessive surcharging regardless of the interchange fee set. Conversely, when the no-surcharge rule is imposed, monopolistic merchants will leave sufficient surplus for consumers to join the card network, and in fact the level of the interchange fee preferred by banks is the socially optimal one.

We contrast these results with the case in which merchants compete according to Bertrand competition. An important condition for the no-surcharge rule and interchange fees to matter is that merchants face limited competition. With Bertrand competition, bank profits and social welfare are independent of the no-surcharge rule or interchange fees, since merchants will 'separate' into those that accept cards and those that do not. Under the no-surcharge rule, some merchants will accept card payments and charge more (assuming card payments are more expensive for merchants to handle than cash), while others will only accept cash and charge less. Any firm that accepts both card and cash payments is vulnerable to a competitor that undercuts its price and just accepts the low-cost cash customers. We also show the same behavior applies to the Hotelling model that Rochet and Tirole analyze when the degree of product differentiation becomes small. [4]

The rest of the paper proceeds as follows. Section 2 sets up the basic model. Section 3 defines the first-best solution which will be adopted by a hypothetical central planner that can determine which consumers will subscribe to the card network and which merchants will accept cards. In Section 4, we consider the case where merchants

---

[3] Equivalently, each merchant knows that a customer will use the card across a wide range of merchants, so that an individual merchant would never be able to capture the benefits of its own subsidy, designed to get cardholders to join the network. Merchants free ride on the actions of other merchants to provide surplus for cardholders.

[4] Rochet and Tirole assume that product differentiation is large enough so that no merchant ever corners the market.

have monopoly power over their prices and examine the welfare implications of the no-surcharge rule, as well as the optimal level of the interchange fee from a private and social perspective. We do this both when consumers face a card membership fee and when they do not. Section 5 repeats the analysis when merchants compete according to Bertrand competition. Finally, Section 6 concludes with a discussion of some extensions.

2. Model setup

Our basic setup (and notation) is the same as that of Rochet and Tirole's. There are two basic differences in our model. Firstly, we consider two different types of merchant pricing, these being monopoly pricing and Bertrand pricing. Secondly, for some of the analysis we distinguish membership and usage decisions, so that consumers make a separate subscription decision (facing a membership fee) and then a usage decision (earning rebates).

We start by summarizing the main assumptions of Rochet and Tirole, noting some minor differences in (A3) and (A4):

(A1) All consumers want to buy a fixed number (N) of 'goods'. The gross benefit of each purchase to a consumer, ignoring any benefit to using a particular form of payment, is v per good. The gross cost of the item sold by each merchant is d, where $v > d > 0$.

(A2) Using a card for a transaction generates a benefit of b_B to cardholders (buyers) and b_S to merchants (sellers). These parameters measure the convenience value from using a payment card to conduct a transaction rather than the alternative, say cash. Consumers have types b_B which are continuously distributed on the interval $[\underline{b}_B, \bar{b}_B]$ according to the density function $h(b_B)$ and distribution function $H(b_B)$. All merchants have the same value of b_S. Consumers know their own b_B, and both parties know the distribution of b_B and the value of b_S. The monotone hazard function $h(b_B)/(1 - H(b_B))$ is assumed to be increasing in b_B.

(A3) A transaction that is done using cards costs the issuer c_I and the acquirer c_A. These capture technological costs as opposed to costs that acquirers might face due to the interchange fee (denoted a), or costs that issuers may face from providing rebates to cardholders. The costs of the alternative to cards (cash) are normalized to zero. One view of the transactional benefits of accepting a payment card (b_S) is that it measures the costs that merchants save by not having to handle cash for a transaction. We assume $\bar{b}_B + b_S > c_I + c_A$ since otherwise payment cards will generate no social surplus. We also make the additional assumption (not made by Rochet and Tirole), that $\underline{b}_B + b_S < c_I + c_A$, so that it is not socially optimal for all consumers to join the card network.

(A4) The equilibrium fee set by symmetric issuers is denoted f, which is assumed to be an increasing function of issuers' net costs $c_I - a$ per transaction. Issuers will not price at or below marginal net cost, so that $f > c_I - a$ in equilibrium. In contrast to Rochet and Tirole, we interpret the fee f as a fee per transaction

(a negative fee will capture cardholder rebates per transaction). A fee per card-holder is introduced in Section 4.4. If total demand for card use can be written as $1 - H(f)$, each issuer's equilibrium profit is assumed to be increasing in the interchange fee (that is, decreasing in net costs). As Rochet and Tirole discuss, these assumptions are met if there is a single monopoly issuer, if issuers compete according to a symmetric Cournot oligopoly with elasticity of demand exceeding one, or if issuers compete according to a symmetric differentiated Bertrand oligopoly. [5] Acquirers are assumed to be perfectly competitive, setting the equilibrium merchant fee $m = c_A + a$. [6]

Assumptions (A1)–(A4) are maintained throughout the paper. Given these assumptions, a cardholder of type b_B will receive indirect utility of $v - p_{card} + b_B - f$ from a purchase made with a card and $v - p_{cash}$ from a purchase made with cash, where p_{card} is the price charged by the merchant if the customer uses a card and p_{cash} is the price charged by the merchant if the customer uses cash. The marginal cardholder is defined as the consumer with benefits b_B per transaction, such that the consumer just wants to join the card network. Provided merchants accept both cards and cash, the marginal cardholder will have a level of b_B equal to

$$b_B^m = f + p_{card} - p_{cash}. \tag{1}$$

Merchants, on the other hand, will receive $p_{card} - d + b_S - m$ if a card is used, and $p_{cash} - d$ if cash is used. The timing of the game is then as follows.

(1) Payment system rules are set. In particular, a rule is set whereby merchants are either allowed to set a surcharge for card payment, or not. Also, the centralized interchange fee a is set.
(2) Issuing and acquiring banks set their prices for issuing and acquiring respectively (f and m).
(3) Consumers and merchants decide whether to join the payment network.
(4) Merchants set prices for goods (p_{card} and p_{cash}).
(5) Consumers decide which merchant to purchase from and what payment method to use.

The ordering of these decisions is quite natural. Payment system rules are set first. Given these, banks set their prices for issuing and acquiring. Consumers and merchants observe these and decide whether to join the network. Each individual merchant, realizing that its own pricing will have a negligible effect on the decision of any potential customer to join the card network, ignores the effect of its pricing on consumers' membership decision in working out its optimal prices. This is modeled by having merchants

---

[5] In the latter case, the consumers' transportation costs of using a particular issuer for a transaction is imbedded in b_B, as explained in Appendix 1 of Rochet and Tirole.

[6] Given this last assumption, the case in which a single proprietary scheme (such as American Express) sets the cardholder and merchant fees directly is formally equivalent to the special case of this model in which the issuer is a monopolist. By setting a and f, a single issuer directly sets m and f. Thus, our model also applies to a single proprietary card scheme.

set their prices after cardholder membership decisions are made. When consumers do not incur a cost of joining the network, all consumers are assumed to join the payment network, and so this distinction does not matter. On the other hand, consumers will generally know which merchants accept the payment card and which do not.[7] Thus, their purchase decision is modeled as occurring in the last stage taking all these factors as given.

Where there are multiple equilibria, the equilibria in which neither consumers nor merchants join the network solely because of self-fulfilling beliefs is ruled out by invoking the elimination of weakly dominated strategies. Self-fulfilling beliefs arise when consumers do not hold cards because they expect merchants will not accept cards. If merchants expect consumers to hold these beliefs, a weak best response is not to join the network themselves, thus reinforcing the cardholders' beliefs. Provided consumers believe merchants will join whenever merchants are indifferent about joining or not, such an equilibrium can be ruled out where other equilibria exist in which at least some merchants join.

The final step in closing the model is to detail how many merchants there are, how these merchants set their prices, and how then consumers choose which merchant to purchase from. Before doing so, the central planner's first-best solution is characterized.

3. First-best solution

The first-best solution involves calculating the optimal size of the card network (how many consumers should hold cards), and when cards should be used? In the first-best solution, consumers should join the card network whenever the expected social benefit arising from using cards exceeds the technological costs of doing so; that is,

$$(b_B + b_S)N \geqslant (c_I + c_A)N$$

or

$$b_B \geqslant c_I + c_A - b_S. \tag{2}$$

Once joined, all cardholders should use their cards and merchants should accept cards. Total surplus is then

$$TS = N \int_{c_I + c_A - b_S}^{\bar{b}_B} (b_B + b_S - c_I - c_A)h(b_B)\,\mathrm{d}b_B$$

$$+ N \int_{\underline{b}_B}^{\bar{b}_B} (v - d)h(b_B)\,\mathrm{d}b_B. \tag{3}$$

In practice, this solution is only likely to be a second-best solution, since cash and other payment alternatives are unlikely to be optimally provided and used. However, the design of alternative payment systems is outside the scope of this paper, and so the costs and benefits of cash are taken as given (normalized to zero). One distortion

---

[7] This assumption matters when competition between merchants is allowed. The impact of this assumption is discussed in footnote 12.

that will prevent the first-best solution being achieved in a decentralized solution is the extent to which issuers and acquirers set cardholder and merchant fees above costs. As will become clear, the no-surcharge rule and the interchange fee will be powerless to eliminate such markups. In this sense, the optimal rules analyzed are also second-best solutions.

4. Monopoly outcomes

While issuers often set membership or annual cardholder fees, there is no reason for them to do so in the model above given the assumption that consumers make a fixed number of transactions (N). Sections 4.1–4.3 below analyze the model without cardholder fixed fees, and show that both the card association and the regulator will want to adopt the no-surcharge rule.

In a richer setting in which issuers face fixed costs of handling individual customers, and in which demand is elastic, it may be optimal for issuers to set fixed fees. Unfortunately, such a framework is not amenable to obtaining definite results. Instead, Sections 4.4–4.6 consider what happens when consumers face an exogenous fixed fee to obtaining a card. This fee can represent the costs and effort that the cardholder incurs in obtaining a card. When surcharging is possible, the introduction of such a cost leads to the breakdown of the card network, providing an additional reason why the no-surcharge rule is desirable. Section 4.7 considers alternative solutions to the no-surcharge rule as a way of preventing the breakdown of the card network, and explains why such alternatives give rise to additional costs. Throughout, it is assumed that each merchant produces one of the N different goods that the consumers want to buy.

4.1. Merchant surcharging

When surcharging is allowed, merchants with monopoly power will exploit their power by setting a price to extract surplus from inframarginal cardholders. The following proposition characterizes the unique equilibrium under surcharging.

Proposition 1. *Under surcharging, a unique equilibrium exists in which the marginal cardholder, defined by b_B^m, satisfies the equation*

$$b_B^m = f(b_S - c_A) + \frac{1 - H(b_B^m)}{h(b_B^m)}. \tag{4}$$

Proof. Under surcharging, merchants will sell to those paying by cash to extract their full surplus, setting $p_{cash} = v$. Merchants will then set a price to card users p_{card} to maximize their profits:

$$\pi = H(b_B^m)(v - d) + (1 - H(b_B^m))(p_{card} - d + b_S - m),$$

where the marginal card user, defined by b_B^m, equates the additional benefits of using cards b_B with the additional costs $f + p_{card} - v$. This problem is equivalent to choosing the marginal card user b_B^m to maximize

$$\pi = (v - d) + (1 - H(b_B^m))(b_B^m + b_S - m - f).$$

The solution is

$$b_B^m = f + m - b_S + \frac{1 - H(b_B^m)}{h(b_B^m)}. \tag{5}$$

Since $f + m - b_S = f + a + c_A - b_S$, the solution to (5) is independent of the interchange fee. To see this, note that both Eq. (5) and issuers' profits can be written solely in terms of $f + a$ (rather than f and a separately). Without loss of generality, Eq. (5) can be evaluated at the Baxter fee $a = b_S - c_A$, which gives the result in (4). Note since the left-hand side of (4) is increasing in b_B^m and the right-hand side is decreasing in b_B^m (from A2), any solution to (4) must be unique. The proof that a solution to (4) exists is provided in the appendix.

So far it has been assumed that merchants will indeed want to accept cards. Another choice for merchants is simply to reject cards and set $p_{cash} = v$, giving them profits of $v - d$. Merchants earn a profit from accepting cards of $(v - d) + (1 - H(b_B^m))(b_B^m + b_S - m - f)$, which exceeds $v - d$ from (5). It follows that merchants will indeed be willing to accept cards. □

Proposition 1 implies the marginal cardholder can be defined without reference to the interchange fee. This neutrality of interchange fees reflects the fact issuers can charge more through the cardholder fee f or through the interchange fee a (and so the merchant fee), but the effect will be the same given that when merchants price separately to cardholders, all parties (merchants, cardholders, issuers and acquirers) only care about the total amount charged by issuers ($f + a$). Whether merchants actually surcharge (or discount) for card transactions depends on the particular interchange fee set. However, regardless of the price set by merchants, the total price cardholders face is $p_{card} + f$. From above, this equals $v + b_B^m$, which does not depend on the interchange fee.

As Section 4.3 will show, the total price faced by cardholders is too high, in the sense a lower merchant surcharge will induce more card usage and raise welfare. Monopolistic merchants will engage in excessive surcharging. For now, note that the total profit obtained by the members of the card association when they allow surcharging is [8]

$$\Pi = N \int_{f(b_S - c_A) + (1 - H(b_B^m))/h(b_B^m)}^{\bar{b}_B} (f(b_S - c_A) + b_S - c_I - c_A)h(b_B)\,\mathrm{d}b_B, \tag{6}$$

---

[8] To see this, recall issuers' per transaction margins under surcharging do not depend on a particular interchange fee, and so can be evaluated at the Baxter interchange fee $a = b_S - c_A$. The resulting margins are $f(b_S - c_A) + (b_S - c_A) - c_I$.

while the total surplus under surcharging is

$$TS = N \int_{f(b_S - c_A) + (1 - H(b_B^m))/h(b_B^m)}^{\bar{b}_B} (b_B + b_S - c_I - c_A) h(b_B) \, db_B$$

$$+ N \int_{\underline{b}_B}^{\bar{b}_B} (v - d) h(b_B) \, db_B. \tag{7}$$

4.2. Equilibrium under the no-surcharge rule

Under the no-surcharge rule, each merchant will set a common price for cash and cards.[9] This common price is denoted p, and the merchants' profit maximizing price p^*. In setting p a merchant faces two alternatives. A merchant can set $p = v$, so consumers who want to use cash will purchase. In this case, a merchant's profit is

$$\pi = v - d + (1 - H(b_B^m))(b_S - m).$$

Consumers will use cards whenever their transactional benefits b_B exceed the additional cardholder fee f (that is, $b_B^m = f$). Merchants will accept cards whenever their transactional benefits b_S exceed the merchant fee m. Thus, if a merchant sets $p = v$, then regardless of the interchange fee, the merchant will earn a profit of at least $v - d$.

Alternatively, a merchant can set $p > v$, in which case the merchant will only sell to consumers who use cards, so that

$$\pi = (1 - H(f + p - v))(p - d + b_S - m).$$

Here, the marginal consumer equates the additional benefits of making a purchase with a card b_B with the additional cost $f + p - v$, so that $b_B^m = f + p - v$.

To determine which price merchants will set, we put an upper bound on the maximal profits a merchant can obtain if it excludes cash customers. Provided the surplus of the good itself is sufficiently large, merchants will not want to exclude cash customers. In this case, an upper bound on merchants' profits when they exclude cash customers turns out to be less than a lower bound on merchants' profits when they price at v. It follows that merchants will set $p^* = v$. To obtain this result, an additional parameter restriction is made, which will be maintained throughout the rest of Section 4.

(A5) The surplus of the good itself (that is, $v - d$) is assumed to be sufficiently large. In particular, we assume

$$v - d > \frac{1 - H(f(b_S - c_A))}{h(f(b_S - c_A))}.$$

This is not likely to be a restrictive assumption. For instance, in the case in which b_B follows the uniform distribution over $[\underline{b}_B, \bar{b}_B]$, it requires that for any consumer the

---

[9] At the privately and socially optimal interchange fee derived below, merchants will wish to set a surcharge equal to $(1 - H(b_B^m))/h(b_B^m)$ for card purchases. Thus, the no-surcharge rule is, in fact, binding.

surplus created from the good itself is larger than the surplus created from the use of the card payment to purchase the good (that is, $v - d > \bar{b}_B + b_S - c_I - c_A$). Using this assumption, Proposition 2 shows that even if cash consumers are excluded, merchants' profit decreases as they increase their price above v (the proof is contained in the appendix).

Proposition 2. *Under the no-surcharge rule, monopolistic merchants will set a uniform price $p^* = v$.*

Given merchants set $p^* = v$, consumers will use cards if and only if $b_B \geqslant f$ and merchants will accept cards if and only if $b_S \geqslant m$. The particular level of f and m depend on the interchange fee set by the card association. Given (A4), issuers' profit will be maximized by setting the highest possible interchange fee subject to the condition that merchants accept cards. Subject to the merchants' participation constraint, this minimizes issuers' net cost per transaction and maximizes the number of transactions using cards. This result is stated as Proposition 3.

Proposition 3. *Under the no-surcharge rule, the card association will set the interchange fee $a^* = b_S - c_A$, consumers will get and use cards if and only if $b_B \geqslant f(b_S - c_A)$, and monopolistic merchants will accept cards.*

Interestingly, the card association's optimal interchange fee derived here corresponds exactly to the Baxter fee. To justify his interchange fee, Baxter did not look at maximizing the banking sector's profit (since he assumed perfect competition between issuers and between acquirers). In the next section, we show that our model can justify the Baxter fee from a social perspective as well.

Before doing so, we note that given the marginal card user is defined as

$$b_B^m = f(b_S - c_A), \tag{8}$$

the total profit obtained by the members of the card association when they impose the no-surcharge rule and set their optimizing interchange fee is

$$\Pi = N \int_{f(b_S - c_A)}^{\bar{b}_B} (f(b_S - c_A) + a^* - c_I) h(b_B) \, db_B$$

$$= N \int_{f(b_S - c_A)}^{\bar{b}_B} (f(b_S - c_A) + b_S - c_I - c_A) h(b_B) \, db_B, \tag{9}$$

while the total surplus under the no-surcharge rule is

$$TS = N \int_{f(b_S - c_A)}^{\bar{b}_B} (b_B + b_S - c_I - c_A) h(b_B) \, db_B$$

$$+ N \int_{\underline{b}_B}^{\bar{b}_B} (v - d) h(b_B) \, db_B. \tag{10}$$

4.3. Welfare implications of the no-surcharge rule

The implications of the no-surcharge rule for the profits of the members of the card association can be determined by comparing (6) and (9). As the following proposition demonstrates, an association of issuers and acquirers prefers to adopt the no-surcharge rule (setting the interchange fee at $b_S - c_A$). The proposition also shows the same result applies to a regulator concerned with maximizing total surplus.

Proposition 4. *The imposition of the no-surcharge rule on monopolistic merchants is preferred by both the card association and the regulator. Both will set the interchange fee equal to the Baxter fee.*

Proof. The no-surcharge rule is preferred by the card association since the expression in (9) is at least as high as the expression in (6). The profit per transaction is the same in the two cases, but under the no-surcharge rule there are more card transactions. The latter result follows from the fact that with surcharging the marginal cardholder has transactional benefits of

$$b_B^m = f(b_S - c_A) + \frac{1 - H(b_B^m)}{h(b_B^m)},$$

which are at least as high as the transactional benefits of the marginal cardholder under the no-surcharge rule (which equal $f(b_S - c_A)$).

The no-surcharge rule is also preferred by the regulator since the expression in (10) is at least as high as the expression in (7). This follows from the fact, when evaluated for the marginal cardholder under the no-surcharge rule, the total surplus per transaction is positive. That is,

$$f(b_S - c_A) + b_S - c_I - c_A > c_I - (b_S - c_A) + b_S - c_I - c_A = 0,$$

where the inequality follows from the assumption in (A4) that $f > c_I - a$. Total surplus can be increased by encouraging more, not less, consumers to use cards. This is achieved by setting the highest interchange fee consistent with merchants' accepting cards; that is, $a^* = b_S - c_A$. □

If merchants are not constrained from surcharging, they will exploit their monopoly power by surcharging excessively. This leads to fewer cardholders, which hurts issuers. It also lowers total surplus. The latter result reflects the fact when merchants have monopoly power they will have a high resistance to accepting cards under the no-surcharge rule. Merchants will only accept cards if the transactional benefits exceed the merchant fee acquirers charge. This will limit the interchange fee that issuers will want to set to the socially optimal level, and make the imposition of the no-surcharge rule unambiguously desirable. [10]

---

[10] These results contrast with the ambiguous results found by Rochet and Tirole. In their model, competition between merchants reduces merchants' resistance to accepting cards, as merchants accept cards in part to steal business from each other. Competition between merchants also helps control any excessive surcharging in the case the no-surcharge rule is lifted. This implies that the interchange fee chosen by an issuer controlled card association (under the no-surcharge rule) may be at, or above, the socially optimal level, and that the welfare effects of the no-surcharge rule are ambiguous.

Given consumers face the same merchant price for using cash or cards, they will only use cards when their transactional benefits exceed the cardholder fee issuers charge. The cardholder fee they face will incorporate the issuers' cost less the interchange fee issuers' receive. When the interchange fee is set at the Baxter level, it will reflect the merchant transactional benefits less acquiring costs. Thus, cardholders will face the joint costs of providing a card transaction (c_I and c_A) but will also receive the joint transactional benefits created by the use of a card (b_B and b_S). Absent any issuer markup, this will lead consumers to choose the first-best level of card holding and usage. To the extent issuers' margins also have to be covered, there will inevitably be some reduction in card usage compared to the first-best solution.

4.4. Surcharging with membership fees

So far, the welfare effects of the no-surcharge rule have only been analyzed assuming issuers set transaction fees (either positive or negative). This was optimal given that consumers are assumed to make a fixed number of purchases. However, in practice, issuers often set annual fees for card membership, especially in the case of credit cards. Such fees could be optimal when issuers incur costs of signing up cardholders in the face of elastic consumer demand for goods. Such fees could also be chosen by an individual issuer so as to avoid attracting the wrong type of cardholder at the margin (those who will not make enough purchases to cover the costs of signing them up and those who are poor credit risks since they would otherwise be put off by a joining fee). Alternatively, membership fees could represent the costs and effort that the cardholder incurs in obtaining a card and carrying the card. Since allowing for elastic demand turns out to greatly complicate the analysis, we simplify by introducing an exogenous cost (or fee), denoted F, that cardholders face to join the card network. A small transaction cost of membership will be sufficient for our arguments to apply. In Section 4.7, we discuss some issues which arise if this fee is endogenized.

The introduction of such a fee may increase or decrease the level of cardholder fees f that results from competition between symmetric issuers for any given level of the interchange fee. It is assumed only that the assumptions (A4) and (A5) continue to apply to this new function f.

With the introduction of a membership fee F, surcharging leads to no one joining the card system, even though card payments can be efficient for a large number of transactions.

Proposition 5. *Under surcharging, when consumers face a membership cost F, a unique equilibrium exists in which monopolistic merchants set a cash price of $p_{cash} = v$. No consumers hold cards.*

Proof. Under surcharging, merchants will set a price for using cards that at least extracts all the payment benefits that the marginal cardholder receives. That is, p_{card} will be greater than or equal to $v + b_B^m - f$. Knowing this, the cardholder of type b_B^m will never join the network in the first place, since joining entails a fixed fee F but provides no usage benefits. This means the valuation of the marginal cardholder would in fact

have to be greater than b_B^m, which contradicts the definition of b_B^m as the valuation of the marginal cardholder. No matter how we define the marginal cardholder, only cardholders with valuations an amount F/N higher will actually want to join. By logical deduction, no consumer will choose to hold cards in the first place. Consequently, there will be a single price of goods sold, $p_{cash} = v$, which is the maximum firms can extract from cash-paying customers. □

Although it can be in the collective interests of merchants to pay cardholders to join the network, since each takes the supply of cardholders as given, each merchant will behave opportunistically. This demonstrates the role that the no-surcharge rule plays in preventing merchants expropriating the additional surplus that arises from the use of cards. In the simple setting considered here, if merchants are not constrained from such pricing, customers will not be prepared to pay anything to join the payment network.[11] Total social surplus is then easily calculated as the value created when all sales are cash. This is simply

$$TS = N(v - d). \tag{11}$$

The result proven in Proposition 5 does not depend on the level of interchange fees. Merchants, in surcharging, will always set prices that extract all the value from the marginal consumer, leaving no surplus to cover their costs of joining the card network in the first place. With no cardholders, the level of the interchange fee is irrelevant.

4.5. The no-surcharge rule with membership fees

In this case, a merchant that accepts cards will set a common price p for cash and cards. As was the case in Section 4.2, in setting the common price a merchant faces two alternatives. A merchant can set $p = v$, in which case consumers who prefer to purchase with cash will still want to purchase. The merchant's profits will be

$$\pi = v - d + (1 - H(b_B^m))(b_S - m).$$

Consumers will hold cards whenever their draw of b_B exceeds the cardholder usage fee f plus their membership fee F spread over the transactions they make (that is, $b_B^m = f + F/N$). Cardholders will always use cards when they can, and merchants will accept cards whenever their transactional benefits b_S exceed the merchant fee m. Thus, if a merchant sets $p = v$, then regardless of the interchange fee, the merchant will earn a profit of at least $v - d$.

Alternatively, a merchant can set $p > v$ in which case the merchant will only sell to consumers who use cards, so that

$$\pi = (1 - H(f + p - v))(p - d + b_S - m).$$

---

[11] A common theme in the literature on holdups is the role of competition or dual sourcing in protecting "specific" investments against ex post holdup. Here, the investment is the membership fee that consumers must sink before merchants and consumers trade. In this sense, the Rochet and Tirole model looks at a case in which competition among merchants helps protect consumers from this kind of holdup, whereas the present model considers what happens in the absence of this protection.

Here, the marginal consumer to use cards equates the additional benefits of making a purchase with a card b_B with the additional cost $f + p - v$. However, when merchants sell only to cardholders, they will always price to at least extract the surplus from the marginal cardholder, so that p will be greater than or equal to $v + b_B^m - f$. Just as in Proposition 5, no consumer will want to hold cards in the first place. This implies:

Proposition 6. *Under the no-surcharge rule, when consumers face a membership cost F, in any equilibrium where cards are used for some transactions, the following conditions will be satisfied*:

$$b_B^m = f + \frac{F}{N}, \quad b_S \geqslant m \text{ and } p^* = v.$$

The second condition of Proposition 6 implies that for there to be an equilibrium whereby cards are used it must be that

$$a \leqslant b_S - c_A. \tag{12}$$

Even under this condition it does not necessarily follow there will be such an equilibrium. It may be, for a particular interchange fee that satisfies (12), that merchants still want to price to exclude cash customers (that is, set $p > v$). This case is ruled out in the following proposition.

Proposition 7. *Under the no-surcharge rule, when consumers face a membership cost F, monopolistic merchants will set the uniform price $p^* = v$.*

Proof. Applying the proof of Proposition 2 to the new cardholder fee f that arises in the presence of the membership cost we have

$$\max_{p \geqslant v} (1 - H(f + p - v))(p - d + b_S - m) \leqslant v - d.$$

Since

$$\max_{p \geqslant v + F/N} (1 - H(f + p - v))(p - d + b_S - m)$$

$$\leqslant \max_{p \geqslant v} (1 - H(f + p - v))(p - d + b_S - m),$$

it is still the case that merchants can do no better than setting $p = v$. □

Proposition 7 demonstrates that if merchants do not want to price high to exclude cash customers in the absence of the fixed fee F, they will also not want to exclude cash customers when card customers have less surplus to extract due to the presence of the fixed fee F. As in Section 4.2, issuers' profit will be maximized by setting the highest possible interchange fee subject to the condition that merchants accept cards. This generates the maximal number of card transactions and the minimum net costs for issuers, and so maximizes the profits of the members of the card association. This result is stated as Proposition 8.

Proposition 8. *Under the no-surcharge rule, when consumers face a membership cost F, the card association will set the interchange fee $a^* = b_S - c_A$, consumers will get and use cards if and only if $b_B \geqslant f(b_S - c_A) + F/N$, and monopolistic merchants will accept cards.*

The introduction of a fixed membership cost for cardholders might suggest that a higher interchange fee is required to help offset the extra costs imposed on cardholders. However, since interchange fees are already set at the highest level consistent with merchants accepting cards, there is no room to further increase the interchange fee.

Given the marginal card user is defined as $b_B^m = f(b_S - c_A) + F/N$, the maximal profit obtained by the members of the card association when they impose the no-surcharge rule and set their optimizing interchange fee is

$$\Pi = N \int_{f(b_S - c_A) + F/N}^{\bar{b}_B} (f(b_S - c_A) + b_S - c_I - c_A) h(b_B) \, db_B \tag{13}$$

while the total surplus under the no-surcharge rule is

$$TS = N \int_{f(b_S - c_A) + F/N}^{\bar{b}_B} (b_B + b_S - c_I - c_A) h(b_B) \, db_B$$

$$+ N \int_{\underline{b}_B}^{\bar{b}_B} (v - d) h(b_B) \, db_B. \tag{14}$$

4.6. Welfare implications of the no-surcharge rule

Section 4.4 showed that with surcharging there will be no equilibrium involving card transactions. Instead, card payment systems will do better by imposing the no-surcharge rule and setting an interchange fee equal to $b_S - c_A$. As Proposition 9 shows, the same result applies to the regulator.

Proposition 9. *When consumers face a membership cost F, the imposition of the no-surcharge rule on monopolistic merchants is preferred by both the card association and the regulator. Both will set the interchange fee equal to the Baxter fee.*

Proof. Under surcharging there are no card transactions, banks earn no profits, and total surplus is defined by (11). Proposition 8 implies that, under the no-surcharge rule, the interchange fee which maximizes the profits of the members of the card association is $a^* = b_S - c_A$. At this interchange fee, banks earn positive profits since

$$f(b_S - c_A) + b_S - c_A - c_I > c_I - (b_S - c_A) + b_S - c_I - c_A = 0, \tag{15}$$

where the inequality follows from the assumption in (A4) that $f > c_I - a$. The card association strictly prefers the use of the no-surcharge rule.

To show that total surplus is higher in (14) than in (11), note that

$$N \int_{f(b_S - c_A) + F/N}^{\bar{b}_B} (b_B + b_S - c_I - c_A) h(b_B) \, db_B > 0,$$

which follows since even when evaluated for the marginal cardholder ($b_B = f(b_S - c_A) + F/N$), the term $b_B + b_S - c_I - c_A$ is positive as a consequence of (15). Thus, a regulator concerned with maximizing total surplus will prefer the adoption of the no-surcharge rule. The regulator will also want to set the interchange fee at a^* since, although there are too few cardholders compared to the first-best solution, at any higher interchange fee merchants will not accept cards. □

The imposition of the no-surcharge rule is strictly preferred by both the card payment system and the regulator since it increases bank profits and total surplus. The no-surcharge rule prevents merchants expropriating the additional surplus that arises from the use of the payment card. If merchants are not constrained from such pricing, customers will not be prepared to pay anything to join the card network. Increasing the interchange fee will not help this unfortunate situation, since merchants will simply increase the price that customers using cards face, in order to expropriate any additional surplus cardholders get through rebates arising with a higher interchange fee. In aggregate, merchants value a constraint that prevents them setting a surcharge, even though each merchant, taking the card network as given, will like to expropriate all the cardholders' surplus from using cards.

Given consumers face the same price for using cash or cards under the no-surcharge rule, they will only hold cards when their transactional benefits exceed the cardholder fee issuers charge as well as the membership fee they face. Compared to the outcome without membership fees, there will be fewer cardholders. However, provided the no-surcharge rule is maintained, there are no new distortions introduced by the need to cover membership costs. This is not the case when surcharging is allowed.

4.7. Endogenous membership fees

Section 4.4 showed that when cardholders face fixed costs of cardholder membership, merchant surcharging will result in no one choosing to hold a card. This highlights an additional reason why the no-surcharge rule may be desired by card associations, and why it can play an efficiency enhancing role. The rule can help eliminate merchants' ex post expropriation of consumers' investment in cards. Ex post expropriation arises from free-riding behavior. In the absence of the no-surcharge rule, each individual merchant hopes other merchants will not surcharge (so that consumers will hold cards), but then surcharges itself.

The results of Section 4.4 also suggests that if for some reason the card association cannot avoid the surcharging outcome, it will have a strong incentive to eliminate card membership fees or subsidize membership costs for consumers. This, however, may not be possible. Subsidizing membership costs may also entail efficiency losses.

The problem of excessive surcharging that results in consumers not wanting to hold cards is a problem that cannot be solved by individual issuers reducing membership fees. When there are a large number of issuers, each will realize its impact on the surcharging behavior of merchants is minimal. To the extent individual issuers set a fixed membership fee as part of their optimal tariff design, the elimination of fixed fees will tend to reduce an issuer's profits, taking as given the decision of other

issuers. Similarly, individual issuers will be reluctant to pay people to sign up, so as to offset their joining costs, especially when this causes adverse selection problems. This suggests some fixed fees are likely to persist, even if merchant surcharging is allowed.

If instead a card association tried to ban membership fees directly this could be deemed illegal by competition authorities given such fees are retail prices that would otherwise be set by competitors. Alternatively, the card association could try to induce individual issuers to eliminate membership fees (or subsidize cardholders' membership costs) by introducing a payment from acquirers to issuers based on the number of cardholders signed up by issuers. However, arrangements not based on usage are vulnerable to abuse, with consumers obtaining multiple cards simply to earn rebates for signing up, or issuers signing up households with multiple cardholders to earn greater payments from acquirers.

For these reasons, even if the setting of membership fees is endogenized, allowing for such fees will still provide an additional justification for why the no-surcharge rule is a desirable rule when merchants have monopoly power.

5. Bertrand competition

In this section, we take the opposite extreme to the previous section by assuming all merchants offer an identical product and compete over price according to Bertrand competition. Thus, we assume there are potentially many (at least three) merchants, all producing an identical good. We also proceed by abstracting from any membership fee F faced by cardholders since this will not affect our analysis. When merchants compete aggressively, they will not be able to sustain excessive surcharges, and so there can be no unravelling of the card network even if card membership fees are introduced.

5.1. Merchant behavior under surcharging

With Bertrand competition, prices will be driven down to (net) costs.

Proposition 10. *Under surcharging, the equilibrium prices set by competitive merchants are $p_{cash} = d$ and $p_{card} = d + m - b_S$, with at least one merchant accepting cards at these prices. The marginal cardholder is defined by $b_B^m = f(b_S - c_A)$.*

Proof. At these prices, merchants are indifferent about accepting cards or not, but in equilibrium at least one merchant will accept cards. To see this is an equilibrium, first note that while the merchants do not earn any margins on transactions, no merchant can do better by changing its decision on accepting cards. If any merchant offers a lower price than those above, it will make a loss, while if any merchant increases its price above the levels above, it will receive no customers. Merchants cannot do better by changing their prices or acceptance decisions.

This will not be true if all merchants accept cards, in which case at least one merchant can do better by specializing in handling cash-only customers. Similarly, if

all merchants accept cash only, a merchant can do better by specializing in handling card-only customers. In equilibrium, at least one merchant accepts cards and the marginal cardholder is defined by substituting the equilibrium prices into (1). This gives

$$b_B^m = f + m - b_S$$
$$= f(b_S - c_A), \tag{16}$$

where the second line follows because $f + m - b_S = f + a + c_A - b_S$, which is independent of a and so can be evaluated at $a = b_S - c_A$. $\square$

Merchants pass the full costs and benefits of card usage back to cardholders, so that there can be no reallocation of surplus between cardholders and merchants. Given that merchants do not surcharge excessively, there is also little need to reallocate surplus in this case. The marginal cardholder is exactly the same as that arising under the no-surcharge rule when merchants have monopoly power. As with the result there, the fact that cardholders face all the costs and all the benefits of the card network helps achieve the efficient levels of card usage. Despite this, given consumers do not internalize the banks' markups when making their usage decisions, there will generally be too little card usage compared to the first-best solution. A final implication of Proposition 10 worth noting is that it implies that if the merchant fee is lower than the merchant's transactional benefit of accepting cards, the merchant will discount for card purchases.

5.2. Merchant behavior under the no-surcharge rule

Provided merchants do not get cost savings from accepting cards that exceed their merchant fee ($b_S \leq m$), each merchant will not *want* to discount for transactions made with cards. With the no-surcharge rule imposed they will not be *allowed* to surcharge for such transactions. Merchants will therefore set a single price p^*. Moreover, as Proposition 11 shows, regardless of the level of a, the equilibrium prices are identical to those under surcharging.

Proposition 11. *Under the no-surcharge rule, if $b_S \leq m$, then in equilibrium competitive merchants will either accept only cash sales at a price of $p_{cash} = d$, or accept cards as well and set a common price $p^* = d + m - b_S$. If $b_S > m$, then in equilibrium competitive merchants will discount for card purchases, setting $p_{cash} = d$ and $p_{card} = d + m - b_S$, with at least one firm accepting cards at these prices. In either case, the marginal cardholder is defined by* (16).

Proof. Consider first the case $b_S \leq m$. To see the above prices represent an equilibrium, note that merchants earn zero margins on all transactions. Those merchants that accept cards will only attract customers who pay by card (since $p^* \geq p_{cash}$). If any individual merchant increases its price, it will attract no customers, while any merchant that decreases its price will make a loss. If $b_S > m$, then the equilibrium without surcharging

is identical to the case with surcharging, since the no-surcharge rule on cards is not binding. In either case, since the prices of cash and card transactions are identical to those in Proposition 10, so too is the definition of the marginal cardholder. □

Under perfect competition, the no-surcharge rule will cause merchants to 'separate' into those that accept cards and those that do not. Those merchants that accept cards will set a price to cover the cost of doing so. Given this, they will only attract cardholders. The remaining merchants will set a lower price, but only accept cash customers. Only if the merchant fee is so low that merchants want to discount for cards there will be no separating equilibrium. In this case, firms accept both cards and cash, but discount for card transactions.

Given the equivalence of the outcomes with surcharging and those without, the choice of optimal interchange fee corresponds exactly to the analysis of Section 5.1. With Bertrand competition amongst merchants, the level of the interchange fee is irrelevant from a private and social perspective.

It is natural to expect these results to extend to the case where merchants have only a small amount of market power. Any attempt to force merchants to price the same for both card and cash customers will also lead merchants to specialize in servicing cash or card customers. That this is true is shown in the Appendix. In a sense, when horizontal differentiation is slight, it is the vertical differentiation aspect of credit card acceptance that dominates.

5.3. Welfare implications of the no-surcharge rule

Since consumers face the same prices regardless of whether merchants' surcharge or not, subscription, usage, and welfare are identical in both cases. This implies:

Proposition 12. *Under Bertrand competition between merchants, both the card association and the regulator are indifferent as to whether surcharges should be allowed or not. Bank profits and total surplus do not change when the no-surcharge rule is lifted, regardless of the level of the interchange fee.*

With perfect competition, there can be no subsidy from cash-paying customers to card-paying customers, since competitive merchants will only accept cards if they can recover the net costs of doing so from their card customers. As a result, the cost of recovering any interchange fees set by the card association along with the issuers' fee is ultimately borne by consumers who use cards. This means the level of the interchange fee will be irrelevant regardless of whether there is surcharging or not, and there is no consequential difference between the outcome with surcharging and that without.

It is insightful to compare this result to those derived earlier. In the earlier results, monopolistic merchants placed a constraint on the interchange fee that could be charged by the card scheme under the no-surcharge rule. This ensured that the interchange fee was limited to the role of aligning the private benefits and private costs of each party with joint benefits and joint costs.

In a world of perfect retail competition, the interchange fee will not be allowed to play the role of aligning joint benefits and joint costs, but nor will it be needed for this purpose. Under perfect competition, merchants will simply pass through any additional costs or benefits they face from accepting cards back to cardholders. Provided consumers observe which merchants accept cards and which do not, consumers will always face the full costs and benefits of holding and using cards. [12] In such a world, the interchange fee becomes irrelevant, and it is strong competition between merchants, rather than merchants' market power, which prevents the card scheme from using the interchange fee and the no-surcharge rule in any anticompetitive way.

Note as with the earlier models, there will be too little card usage from the central planner's perspective. Cardholders do not internalize the markups they generate for issuing banks when making their usage decisions. When these markups are zero, as Baxter assumed, then the first-best outcome is obtained regardless of the level of interchange fees or whether surcharging is allowed. In contrast to our results in the monopoly model, where we provided new reasons why an appropriate interchange fee and the no-surcharge rule are needed, our model of Bertrand competition puts limits on the conditions under which interchange fees and the no-surcharge rule will matter.

6. Conclusions

The existing literature on payment systems has tended to either emphasize the role of interchange fees in balancing demand by cardholders and merchants so as to maximize network benefits – see Baxter (1983), Chang and Evans (2000), Rochet and Tirole (2002), and Schmalensee (2002) – or, as with Frankel (1998), argued for regulation of interchange fees to zero on the grounds that only then will cardholders and merchants face the true costs of the services they use, rather than be cross-subsidized by cash customers. Relative to this literature our paper makes two contributions.

Firstly, it puts limits on the extent to which interchange fees can matter. When merchants earn negligible surplus (due to intense retail competition), interchange fees cannot play their normal reallocative role. Instead, facing the no-surcharge rule, merchants will specialize in catering either to cash- or card-paying customers. Cross-subsidies of the type Frankel considers will not be possible, although in the setting of perfectly competitive merchants, nor will they be needed.

Secondly, when merchants do have significant market power, the interchange fee can be set so as to appropriately allocate costs and benefits between cardholders and merchants, but only under the no-surcharge rule. With merchant surcharges allowed, merchants will surcharge excessively, resulting in too little holding and usage of cards. In the case consumers first face a membership cost of joining the card network, this undersubscription of the card network can be extreme. Although merchants value having

---

[12] If consumers are not aware of which merchants accept cards and which do not, merchants will only accept cards in the face of the no-surcharge rule if their transactional benefits from doing so are at least as high as the merchant fee they pay. Provided the interchange fee is set less than or equal to the Baxter fee, the merchants' participation constraint will be met, and issuers' profits and welfare will still be invariant to whether surcharging is allowed or not. In this case, all merchants accept cash and cards.

cardholders, each individual merchant takes as given the number of cardholders, realizing that its own contribution to a system-wide subsidy will have a negligible effect on the joining decision of any potential customer. Since merchants ignore the effect their own price discrimination has on a consumer's decision to join the card network, they surcharge too much, the result being too few cardholders from the card network's as well as the social planner's perspective. The no-surcharge rule prevents this free-riding behavior, as well as excessive surcharging more generally, while an appropriately set interchange fee then balances the need to promote cards to consumers while keeping merchants onboard the network.

Our results demonstrate that both types of merchant pricing constrain the ability of an issuer-controlled card association to use interchange fees or the no-surcharge rule in inefficient ways, although for quite different reasons. Merchants with a lot of market power will not be willing to accept cards unless merchant fees are at or below the cost savings card acceptance provides them. This ensures the Baxter interchange fee is privately optimal, since it is the highest interchange fee for which merchants' participation constraint is still met. On the other hand, merchants that compete aggressively will not be able to sustain any cross-subsidy between cash and card customers under the no-surcharge rule. In either case, the no-surcharge rule and privately set interchange fees cannot reduce welfare, and in the case of monopolistic merchants, it will be welfare enhancing.

A simplifying assumption throughout the paper was that merchants are identical. In practice, different types of industry structures co-exist. A straightforward extension of the model is to allow for some industries to be fully competitive, and others to be monopolistic. Since the level of the interchange fee and surcharging are irrelevant in competitive industries, the results with monopolistic merchants (and no membership fees) should then carry over to this more general setting.

Another type of heterogeneity that can be incorporated into the model is to allow variation in merchants' transactional benefits of accepting cards. In the case of monopolistic merchants this extension is straightforward, provided merchants cannot surcharge. Such merchants will only accept cards if their transactional benefits exceed the merchant fee. This provides rigorous foundations for the partial demands approach assumed by Schmalensee (2002). For the case in which merchants compete according to Bertrand competition, the extension to heterogenous merchant benefits will not change our results, provided there are at least two merchants that share the highest transactional benefits of accepting cards. These merchants will be the only ones accepting cards, with all other merchants rejecting cards and selling just to cash consumers. [13]

Throughout we have assumed away any frictions that would prevent merchant surcharging in the absence of the no-surcharge rule. Frankel (1998) has argued merchants may not surcharge even if they are free to do so, provided merchant fees are not too great. He calls this phenomenon price coherence. It may be worth exploring the case for price coherence. If price coherence holds primarily for moderately competitive merchants, it suggests the implications of the no-surcharge rule will then depend on

---

[13] In Wright (2001), I provide a model in which merchant heterogeneity is incorporated for the Hotelling model of merchant competition.

the effect of the rule on the remaining merchants. Since these may be only the most competitive merchants and those with considerable market power, the results from our analysis of the no-surcharge rule may become especially relevant.

Acknowledgements

I would like to thank two referees for insightful comments and Visa International for funding earlier related research (Wright, 2000). The views expressed here are strictly my own, as is the responsibility for any errors.

Appendix A.

Proof of existence for Proposition 1. A solution to (4) exists with $\underline{b}_B < b_B < \bar{b}_B$ since $\bar{b}_B > f(b_S - c_A)$ and $f(b_S - c_A) + 1/h(\underline{b}_B) > \underline{b}_B$. The first condition holds since if $f(b_S - c_A) \geqslant \bar{b}_B$, issuers will earn no profits, and so issuers can do better by lowering fees. That issuers can profitably do this is guaranteed by the assumption in (A3) that $\bar{b}_B + b_S > c_I + c_A$. The second condition holds since

$$f(b_S - c_A) + 1/h(\underline{b}_B) > f(b_S - c_A)$$

$$> c_I - (b_S - c_A)$$

$$> \underline{b}_B,$$

where the second to last inequality follows from the assumption in (A4) that issuers price above cost, and the last inequality follows from the assumption in (A3) that $\underline{b}_B + b_S < c_I + c_A$. $\quad\square$

Proof of Proposition 2. If a merchant sets $p > v$, it will only face demand from cardholders, so that

$$\pi = (1 - H(b_B^m))(p - d + b_S - m),$$

where $b_B^m = f + p - v$. This implies

$$\pi = (1 - H(b_B^m))(b_B^m + v - d + b_S - m - f)$$

$$= (1 - H(b_B^m))(b_B^m + v - d - f(b_S - c_A))$$

$$= (1 - H(f(b_S - c_A) + p - v))(p - d), \tag{A.1}$$

where the second line follows because $b_S - m - f = b_S - c_A - a - f$ so that the merchant's profit is independent of the interchange fee and can be evaluated at $a = b_S - c_A$. It follows that

$$\lim_{p \to v} \pi \leqslant v - d. \tag{A.2}$$

Differentiating (A.1) with respect to price, we get

$$\frac{d\pi}{dp} = 1 - H(f(b_S - c_A) + p - v) - (p - d)h(f(b_S - c_A) + p - v),$$

so that given (A5)

$$\lim_{p \to v} \frac{d\pi}{dp} = 1 - H(f(b_S - c_A)) - (v - d)h(f(b_S - c_A)) < 0. \tag{A.3}$$

Evaluating at any $p > v$,

$$\frac{d\pi}{dp} = 1 - H(f(b_S - c_A) + p - v) - (p - d)h(f(b_S - c_A) + p - v)$$

$$< 1 - H(f(b_S - c_A) + p - v) - (v - d)h(f(b_S - c_A) + p - v)$$

$$= (1 - H(f(b_S - c_A) + p - v))\left(1 - \frac{(v - d)h(f(b_S - c_A) + p - v)}{1 - H(f(b_S - c_A) + p - v)}\right)$$

$$\leqslant (1 - H(f(b_S - c_A) + p - v))\left(1 - \frac{(v - d)h(f(b_S - c_A))}{1 - H(f(b_S - c_A))}\right)$$

$$\leqslant 0, \tag{A.4}$$

where the second to last line follows from (A2) and the last line follows from (A5).

Taken together, the results in (A.2)–(A.4) demonstrate that a merchant's profit decreases as it increases prices above v. Merchants will optimally set $p^* = v$, and there will be both cash and card customers.[14] $\quad\square$

Proof of merchant specialization in the Hotelling model. To show that our results of Section 5 extend to the case where merchants have a small amount of market power, we extend our model to the case in which merchants compete according to the Hotelling model of competition. Following Rochet and Tirole's derivations (in Appendix 2 of their paper), assuming the no-surcharge rule is in effect, one merchant accepting cards and one rejecting is an equilibrium if and only if

$$\left[t - \frac{(1 - H(f))(\beta(f) + b_S - m)}{3}\right]^2 \geqslant t^2 \tag{A.5}$$

and

$$\left[t + \frac{(1 - H(f))(\beta(f) + b_S - m)}{3}\right]^2 - \beta(f)(m - b_S)H(f)(1 - H(f)) \geqslant t^2, \tag{A.6}$$

where t is the standard transportation cost in the Hotelling model, and[15]

$$\beta(f) = \frac{\int_f^{\bar{b}_B}(b_B - f)h(b_B)\,db_B}{1 - H(f)}.$$

---

[14] Clearly the condition in (A5) is only a sufficient condition to obtain this result. For instance, to the extent there exist consumers who prefer to pay by cash when they face the cardholder fee $f(b_S - c_A)$, then $H > 0$ and π will be strictly lower than $v - d$, even as $p \to v$.

[15] Rochet and Tirole obtain the same result with $\beta(f)$ defined in terms of gross benefits b_B rather than net benefits $b_B - f$. This is because they view f as a cardholder joining fee rather than as a fee per transaction. This does not affect the results here.

Rochet and Tirole argue that if $\beta(f) + b_S - m > 0$ then condition (A.5) will not hold and so this is not an equilibrium. On the other hand, if $\beta(f) + b_S - m \leqslant 0$ then $t + (1 - H(f))(\beta(f) + b_S - m)/3 \leqslant t$ and $m > b_S$ so that $\beta(f)(m - b_S)H(f)(1 - H(f)) > 0$ and condition (A.6) will not hold. Thus, they argue, regardless of a, there will not be any 'hybrid' equilibrium in which one firm accepts cards and one does not.

The apparent inconsistency in this result and our result on Bertrand competition can be resolved by noting what happens to the above conditions for small t. For instance, consider the simple case the interchange fee is set so that $m = b_S$. Then both (A.5) and (A.6) will hold for small enough t. An equilibrium will exist where one merchant will reject cards and another will accept cards. The same result will hold for lower interchange fees, and for somewhat higher interchange fees. A low value of transportation costs t also raises the possibility that all consumers that prefer to use cards will go to the merchant that accepts cards, regardless of their location, and all consumers that prefer to use cash will go to the merchant that only accepts cash, regardless of their location. This violates the assumption in Rochet and Tirole that product differentiation is large enough so that no merchant ever corners the market, suggesting conditions (A.5) and (A.6) may not even apply. $\square$

References

Baxter, W.F., 1983. Bank interchange of transactional paper: Legal perspectives. Journal of Law and Economics 26, 541–588.

Chakravorti, S., Shah, A., 2001. A study of the interrelated bilateral transactions in credit card networks. Emerging Payments Occasional Paper Series, EPS-2001-2.

Chang, H.H., Evans, D.S., 2000. The competitive effects of the collective setting of interchange fees by payment card systems. The Antitrust Bulletin 45(3), 641–677.

Coase, R.H., 1960. The problem of social cost. Journal of Law and Economics 3, 1–44.

Cruickshank, D., 2000. Competition in UK banking. Report to the Chancellor of the Exchequer, www.bankreview.org.uk.

European Commission, 2000. Commission plans to clear certain Visa provisions and challenge others. Press release, DN:IP/00/1164, Brussels, 16 October.

European Commission, 2001. Notice pursuant to Article 19(3) of Council Regulation No 17-Case Comp/29.373-Visa International. Official Journal of the European Communities 2001/C226.

Evans, D., Schmalensee, R., 1999. Paying with Plastic: The Digital Revolution in Buying and Borrowing. MIT Press, Cambridge, MA.

Frankel, A.S., 1998. Monopoly and competition in the supply and exchange of money. Antitrust Law Journal 66, 313–361.

Gans, J., King, S., 2002. The neutrality of interchange fees in payment systems. Mimeo., University of Melbourne, 18 July.

HM Treasury, 2001. Competition in payment systems: A response to consultation. Treasury press notice (95/01).

Krueger, M., 2001. Interchange fees in the line of fire. Mimeo., Institute for Prospective Technological Studies, Seville, Spain.

Reserve Bank of Australia, 2002. Reform of credit card schemes in Australia: IV final reforms and regulation impact statement. www.rba.gov.au.

Rochet, J., Tirole, J., 2002. Cooperation among competitors: The economics of payment card associations. Rand Journal of Economics, forthcoming.

Schmalensee, R., 2002. Payment systems and interchange fees. Journal of Industrial Economics 50, 103–122.

Schwartz, M., Vincent, D., 2002. Same price, cash or card: Vertical control by payment networks. Working paper 02-01, Georgetown University.

Small, J.P., Wright, J., 2001. The bilateral negotiation of interchange fees in payment schemes. Mimeo., NECG and University of Auckland.

Wright, J., 2000. An economic analysis of a card payment network. Mimeo., NECG and University of Auckland.

Wright, J., 2001. The determinants of optimal interchange fees in payment systems. Working Paper No. 220, University of Auckland.