Same Price, Cash or Card:

Vertical Control by Payment Networks

Marius Schwartz
Daniel R. Vincent

Working Paper 02–01

February 2002

Georgetown University
Department of Economics
580 Intercultural Center
Washington, DC 20057

Tel: (202) 687–5830
Fax: (202) 687–6102
http://econ.georgetown.edu/workingpapers/

Same Price, Cash or Card:

Vertical Control by Payments Networks

by

Marius Schwartz

Daniel R. Vincent[*]

December 2001

Abstract: The no-surcharge rule (NSR) prohibits merchants from charging different prices to consumers that use credit cards instead of cash. We show that, while an NSR raises card company profits, it may reduce both cash and card transactions. If the card company can offer rebates to its cardholders, it will do so. Rebates benefit card users and harm cash users; they raise total surplus if and only if the proportion of cash users relative to card users exceeds some threshold. A similar condition determines whether total surplus rises under the NSR with rebates compared to no NSR; aggregate consumer surplus moves in opposite direction to total surplus. If the card company cannot limit its member banks from competing vigorously, then an NSR, by cross-subsidizing card purchases, can still reduce total surplus.

JEL Classification: D42, G28, L42, L8.

---

[*] Schwartz: Department of Economics, Georgetown University, Washington DC 20057-1036, <schwarm2@georgetown.edu> Vincent: Department of Economics, University of Maryland, College Park, MD 20742 <dvincent@wam.umd.edu>. For helpful comments and suggestions, we would like to thank seminar participants at Georgetown University, Tulane University, University of Virginia, Patrick Greenlee, Alex Raskovich, George Rozanski and, especially, Andrew Dick. We alone are responsible, however, for the views expressed in this paper.

1. Introduction

Transactions through electronic payment networks (EPNs) in the U.S. exceeded $1.18 trillion in 1999 and are growing rapidly.[1] Several practices in this important industry have attracted controversy and antitrust scrutiny.[2] One such practice involves constraints on the ability of merchants to set different prices depending on the means of payment employed such as credit or debit cards, cash or checks. We examine these constraints as instruments of vertical control, assess their welfare effects, and show that their presence may explain the growing phenomenon of rebates and reward programs in payments markets.

Uniform pricing constraints were at various times imposed by law or by EPN rules that prohibited merchants from imposing surcharges (or adverse non-price terms) for payments with an EPN card, even though merchants may face higher costs for such transactions due to fees charged by the EPN.[3] Even in the absence of formal prohibitions, merchants are often reluctant to set different retail prices depending on the means of payment.[4] We refer to all these limits as the No-Surcharge "Rule" (NSR). Our analysis is relevant to assessing the desirability of the rule

---

[1] This figure includes $1.09 trillion in transactions using credit cards and offline debit cards (of which $561.5 billion was through Visa, $277.9 billion through Mastercard, and the rest through proprietary networks such as American Express and Discover), and $84.9 billion using online debit (mainly through regional electronic funds transfer (EFT) networks). *Nilson Report*, May 2000, issue 716. <http://www.nilsonreport.com/issue_716.html.>

[2] For example, there is debate over whether the joint setting of certain network fees by EPN-member banks (as in "bank card" associations such as Visa, Mastercard, and regional networks) is anti-competitive. (See Salop 1990; Carleton and Frankel 1995, 1995a; or Evans and Schmalensee 1995, 1999.) Also, a common EPN requirement is that merchants must accept all the EPN's cards if they wish to accept any, e.g., merchants that accept Visa's credit card must also accept its online debit card. This requirement is the target of a major lawsuit brought by Walmart and other retailers against Mastercard and Visa, alleging anti-competitive tying and seeking over $8 billion in damages. "Visa, Mastercard Battle with Retailers," by Paul Beckett, *Wall Street Journal*, Nov. 24, 1998. "The Perils of Plastic," by Joseph Webber and Ann Therese Palmer, *Business Week*, February 14, 2000.

[3] Surcharges on credit card transactions were prohibited by federal statutes from 1968 to 1985 and remain prohibited by some states (e.g., Florida). For a detailed history of the U.S. legislative and regulatory treatment of surcharges, see Chakravorty and Shah (2001).Visa long had its own no-surcharge rule. It relaxed the rule quite recently. Mastercard currently prohibits its merchants from "surcharging" customers for credit purchases, though it allows cash "discounts". In Europe, both associations prohibit discounts and surcharges. (Rochet and Tirole 1999 and www.mastercard.com/consumer/cust_serv.html) Allowing cash discounts but prohibiting card surcharges still limits a merchant's ability to set different prices for transactions using noncash instruments such as checks.

[4] Cash discounts are rare. According to one retailer survey, fewer than 1% of merchants offer cash discounts. Chain Store Age, *Fourth Annual Survey of Retail Credit Trends*, January 1994, section 2.

when it is imposed by law or private regulations. For example, the practice of prohibiting surcharges is banned in the United Kingdom and has come under scrutiny in Australia (Reserve Bank of Australia (RBA) 2001 and RBA/ACCC Study 2000). When the "rule" instead derives from other characteristics of merchant's trading environment (and its repeal is therefore not an option), our analysis helps evaluate the effects of restricting an EPN's ability to offer rebates to its cardholders. Finally, the analysis is a necessary step towards evaluating card tying policies (see fn. 2 above), since such tying would have no force if merchant surcharges were unrestricted.

We consider a monopolist EPN that contracts with a representative merchant. The merchant faces downward-sloping demand from consumers, who pay with the EPN card or other means ("cash").[5] The EPN may set charges to the merchant and to card users. In this setting, one might expect an NSR to increase total consumer surplus and overall welfare, drawing on intuition from optimal taxation (or Ramsey pricing), where inefficiency is reduced by using a broader tax base so as to lower the tax rate. The analogy is that, for a given above-cost charge from the EPN to the merchant for card transactions, an NSR leads a merchant to set a "moderate" uniform price for card and cash transactions instead of a higher price for card transactions alone; such price uniformity can reduce misallocation between card and cash transactions.

The optimal tax analogy is flawed, however. Here, the EPN is unregulated and its profit is not held constant. Allowing it to tax non-card sales as well—indirectly via the NSR—leads the EPN to raise its charge to the merchant. As a result, the NSR can lead to a higher retail price for both cash and card transactions. In that case, welfare declines for both cash users and card users as well as merchants. This contrasts with standard comparisons of uniform pricing and third-degree price discrimination where, under regularity conditions also satisfied here, a requirement of uniform pricing causes at least some price(s) to fall (Nahata et al. 1990; Malueg 1992).

Our analysis also highlights a contrast between the NSR and a well-known instrument of vertical control—maximum resale price maintenance (RPM). Both practices enable a supplier to reduce the margin charged on its product by an imperfectly-competitive downstream firm.

---

[5] We use "card" to denote any electronic payment instrument, and "cash" to denote the alternative means of payment. Also, we sometimes will refer to the EPN as the card company. We abstract from the credit role of some electronic payments instruments and focus solely on its payment function. Chakravorti and Emmons (2001) present a model where some consumers use cards for both functions while others use them only as a payment instrument, and investigate the presence of cross-subsidies under an NSR from the former to the latter.

Maximum RPM, however, affects only the targeted product, lowering its price and benefitting consumers and overall welfare (Tirole 1988). In contrast, an NSR squeezes the merchant's margin indirectly, by requiring the same retail price to be charged for the other product (here, cash transactions). Thus, the retail price for cash users rises; moreover, since the EPN is induced to raise its charge to the merchant, the merchant's (now uniform) retail price is pushed up.

In addition, an NSR alters the EPN's preferred structure of charges between merchants and cardholders. If merchant surcharges to consumers were unrestricted, only the EPN's aggregate charge would matter, its division between cardholders and merchants would be irrelevant (Baxter 1983). With an NSR constraint, however, the EPN concentrates its charges on merchants. Indeed, given the NSR, if rebates (negative charges) to cardholders are feasible, they will be offered.

Cardholder rebates are often viewed as reflecting the inability of an EPN to prevent its member banks from competing for cardholders and dissipating rents generated by high EPN charges to merchants. Our analysis reveals a different possible role for rebates: they enable an EPN to magnify the effect of an NSR constraint on merchant pricing. Consistent with this interpretation, cardholder rebates have been offered not only by bank-card associations, but also by proprietary networks such as Discover, where a single entity controls the charges to both merchants and cardholders. The EPN grants rebates to boost cardholders' demand and raises its charge to merchants knowing that they will absorb part of the increase, since the NSR requires that any increase must apply equally to cash users. Rebates thus misallocate transactions towards cards, the opposite of what occurs absent an NSR.

Despite the significant and growing importance of electronic means of payment (Evans and Schmalensee 1999), there are relatively few formal economic analyses of payment networks.. The standard reference on the operation of a payment system that involves multiple banks (bank-card networks, such as Visa or Mastercard) is Baxter (1983). As Baxter illustrates, typical payment transactions involve four parties. A cardholder exchanges with a merchant a promise to pay (a credit card receipt, for example) in return for goods. The merchant sells the promise to pay to a bank with which it has contracted, known as the (merchant's) "acquiring bank." The merchant receives the face value of the promise minus a fee called the "merchant discount." The acquiring bank then sells the receipt to the bank that issued the card to the cardholder ("issuing bank"), again at face value minus a fee known as the "interchange fee." The issuing bank collects from the cardholder the amount promised to pay. Conceptually, and sometimes in practice, the issuing bank may also charge the cardholder a fee (or grant a rebate).

In a bank-card network, each member bank sets its own terms to its cardholders or merchants. In *proprietary* networks such as American Express, Discover and Diner's Club, the same integrated entity deals with both merchants and cardholders. The interchange fee in this case is a fiction and there are only two true prices for credit card services, the merchant discount fee and any cardholder fees. Our model considers a profit-maximizing agent, the EPN, setting the charge to merchants and, for most of the paper, also to cardholders. This model is most obviously interpreted as one of a proprietary card network. It is also an approximation of a bank-card network under two conditions: (a) merchant acquisition is competitive (variations in the interchange fee are then fully passed through to the merchant discount, and one can view the merchant discount as being set by the EPN's issuing banks—through their choice of interchange fee—to maximize their profit); and (b) competition is weak between issuing banks in pricing to cardholders (so that cardholder charges can also be viewed as being chosen to maximize overall profits of issuing banks).[6] Later we modify the model to allow competition among EPN banks in setting cardholder charges.

The complexity of this market implies that EPN pricing practices will influence multiple decisions. For example, merchants must decide whether or not to accept a card, and customers must choose whether or not to use a card and, also, their level of purchases. Schmalensee (1999) examines how the joint setting of interchange fees affects the marginal decisions of merchants to accept cards (as well as the decision of consumers to carry cards). Rochet and Tirole (1999) analyze the impact of interchange fees and the NSR on the decision of a consumer to use a card versus an alternative means of payment. In their model, consumer have unit demands for a good but differ in their private values of paying with cards versus cash. The net cost of using cards affects the number of consumers who choose cards rather than cash. However, as Rochet and Tirole observe, the total quantity of purchases is unaffected by the NSR given the assumptions of unit demands and that all consumers are always served in the Hotelling competition among merchants.

---

6 Rochet and Tirole (1999) and Schmalensee (1999), discussed further below, assume imperfect competition at the merchant level and in card issuance, but perfect competition in merchant acquisition (citing empirical evidence that acquisition is significantly more competitive than issuance).

Our focus is on the impact of an NSR on the *quantities* of purchases. We assume that the means of payment for a given consumer is exogenous; a fraction of consumers use cards and the rest use cash. In our model, however, consumers have continuous demand functions for goods. Thus, while EPN pricing will not affect the number of cardholders in our model, it can and will affect the total quantities of card purchases and, through the NSR, of cash purchases as well.[7]

The remainder of the paper is organized as follows. Section 2 presents our model without the NSR, where a monopolist EPN can commit to setting (linear) charges to a monopolist merchant and to cardholders. Section 3 shows that the NSR increases the EPN's profit and, holding the sum of merchant and cardholder fees fixed, that the EPN now strictly prefers to load all the fees on the merchant. Section 4 considers the case where rebates to cardholders are not feasible, showing that the NSR can then harm both cash *and card* users (as well as the merchant). Section 5 allows for cardholder rebates and specializes to the case of linear demand. Given the NSR, the introduction of rebates benefits card users (relative to no rebates), and harms cash users; overall welfare increases if and only if the proportion of cash users relative to card users is not too small. A similar condition determines whether total surplus rises under the NSR with rebates, as compared to no NSR. Aggregate consumer surplus moves in opposite direction to total surplus. Thus, when total surplus falls, consumers as a whole (and the EPN) gain at the merchant's expense; when total surplus rises, the EPN gains at the expense of the merchant and of consumers as a whole.

Section 6 modifies the industry structure by assuming that the EPN cannot control the prices charged by its card-issuing member banks to cardholders. An NSR again will induce rebates to cardholders, benefitting them but harming cash users. If competition among the

---

7 Wright (2000) uses the Rochet and Tirole framework to analyse the role of the NSR in preventing ex post opportunistic pricing against cardholders by a monopolist merchant. Other differences between our approach and Rochet and Tirole's are less consequential. For example, they assume (as do we) that EPN members set the interchange fee (and ultimately the merchant discount) jointly, but may compete in their fees to cardholders. This is modeled by treating the equilibrium cardholder fee as a reduced form function that decreases in the interchange fee. (Raising the interchange fee raises a cardholder bank's margin, thereby inducing the bank to cut its cardholder fee so as to expand card usage.) The two cases we consider—an EPN monopolist that sets cardholder fees directly, or Bertrand competition among EPN members—are special cases of their specification. Regarding merchant behavior, they assume symmetric duopolists in Hotelling competition while we assume a monopolist, but in both cases the number of merchants is fixed and each faces downward-sloping demand. Like Rochet and Tirole, we address whether the merchant will agree to accept a card.

member banks is strong (Bertrand), then—for linear demand—an NSR increases overall consumer surplus regardless of the relative sizes of the two consumer groups. However, as long as the merchant's benefit from card use is low, the NSR reduces total surplus, because the cross-subsidy to card use biases the mix of payment modes.

2. The Model and Pricing With Surcharges Possible

Consumers: We consider two types of consumers. Type e consumers ("cardholders") buy units of a good using only cards from an electronic payments network; their mass is 1. Type c consumers buy units of a good using only an outside means of payment, call it cash; their mass is α. There is no substitutability across means of payment for these consumers. Consumers are otherwise identical and have quasilinear preferences: $U(p_j, q_j)=V(q_j)-p_j q_j$, $V'(\bullet)>0$, $V''(\bullet)<0$, $j=c,e$.

Throughout, q_j is the *per capita* number of transactions of a consumer of type $j = c, e,$ and p_j is the net price per unit of transaction paid by such a consumer. The net price paid by cash users equals the price charged by the merchant but the two prices may differ for card users. Let p_e^M denote the price charged by the merchant to a card-using consumer: $p_e = p_e^M+t,$ where t is the per unit charge (or rebate if $t < 0$) imposed by the EPN company on card users. For each type of consumer, the (downward sloping) inverse demand function is given by $V'(q_j) = p_j$.

Merchants: We assume that merchants are local monopolists who treat the above inverse demand curve as the demand for their product from each type of consumer. The marginal cost of providing a good to a cash consumer is assumed constant and is normalized to zero. The merchant may also gain a benefit, $b \geq 0,$ from being paid by card instead of cash, reflecting potential savings on cash-handling costs. The merchant is charged a per-unit fee i by the EPN.

The merchant's profit is $\alpha p_c q_c$ from cash users and $p_e^M q_e - (i-b)q_e$ from card users, where quantities are given by $p_c = V'(q_c)$ and $p_e^M = V'(q_e) - t$. For given values of i and t, the merchant's problem can therefore be formulated as choosing a level of x to solve

$$\max_x (V'(x)-(i+t-b))x.$$

Observe that $i=t=b=0$ yields the merchant's problem vis-a-vis the cash market. Written this way, the term $i+t$ can be interpreted as a tax imposed on the card market by the EPN. For given (i,t), we denote the solution to this problem by $\Pi^M(i,t)$.

Electronic Payments Network: As noted, we suppress the distinction between the interchange fee and merchant discount, and simply view the EPN as setting the charge to merchants, *i*, monopolistically. With the exception of Section 6, we also assume that the EPN acts monopolistically in the setting of any cardholder fees. The timing of price setting is in a Stackelberg manner: that is, the EPN sets *t* and *i* and commits to this profile of prices and, given *t* and *i,* the merchant sets her monopoly price. The EPN's marginal cost of servicing a card transaction is assumed to be zero.

We assume that two-part tariffs are not available either to the EPN or to the merchant. What is important for our analysis is that the sequential monopoly environment between the card company and the merchant lead to some inefficient pricing at both the merchant and EPN levels.[8] For simplicity, we assume that only linear pricing is feasible for each agent.

The first-order conditions from the merchant's problem yields a derived inverse demand curve for card transactions defined, implicitly, by

$$i+t=V'(x)+xV''(x)+b \tag{1}$$

Therefore, the EPN maximizes $(i+t)\,x$ or

$$\Pi^e(b)= \max_x\,(V'(x)+xV''(x)+b)\,x \tag{2}$$

Since x is a function of $i+t$ but not i or t separately, the card company varies x by varying the sum of charges, $i+t$. This leads immediately to the following well-known result. (See Baxter, 1983.)

[8] There are a variety of reasons why fully efficient two-part tariffs (or other nonlinear pricing) may not be achievable for the EPN to eliminate such double marginalization. A typical EPN has relationships with a vast number of merchants, and contracting costs could make merchant-specific, two-part tariffs prohibitively expensive. Furthermore, merchants aggregated together in a single market place, such as a mall, may be able to avoid most of the impact of a fixed fee by channeling all card purchases to a single merchant. Additionally, in the context of asymmetric information, for example with heterogeneous merchants, the optimal two-part tariff generally yields some surplus to the high demand merchant and pricing at levels above marginal cost.

Proposition 1: *Suppose merchant surcharges are allowed. If (i,t) maximizes the profits of the EPN, then so too does any pair (i′,t′) where t′+i′=t+i.*

That is, it is irrelevant whether the EPN charges its fee to the consumers, to the merchant, or to a combination of the two. Since the sum, $i+t$, can be viewed as a transactions tax, Proposition 1 echoes the familiar result that the effects of a tax are invariant to whether the obligation to pay the tax is placed on buyers or on sellers. However, the next section shows that, when a no-surcharge rule is in effect, EPN profits will vary, for a given $i+t$, depending on the relative values of i and t.

The remainder of the paper imposes some further restrictions on the environment:

A1) *i) The merchant's cash market revenue function, $p\,f(p)$ where $f(\bullet)$ is the inverse of $V'(\bullet)$, is concave in price; ii) $x\,V'''(x)+V''(x) \leq 0$ which implies the merchant's revenue function is also concave in quantity;*

A2) *The EPN's revenue function $x(x\,V''(x)+V'(x))$ is concave in quantity;*

A3) *For $x_0 = \text{argmax}_x\; x\,V'(x)$, $b < -x_0^{\,2}\,V'''(x_0) - 2\,x_0\,V''(x_0)$.*

The first two assumptions are concavity assumptions which are sufficient to ensure a unique solution to the various optimization problems that arise in the market where surcharging is feasible.[9] Assumption **A3)** implies that the merchant's benefit from card use (b) is not so great that -- in spite of the double-marginalization on cards -- card transactions would exceed cash transactions. (See Lemma 2.) Note that a special case that satisfies all three assumptions is the linear demand case:

LD) *$V'(x) = \max\{0,1-x\}$.*

In this case, $b < 1$ is sufficient to ensure **A3)** is satisfied.

The subsequent analysis uses the following lemma frequently. Define $x(k)$ as $x(k)=\text{argmax}_x\,(V'(x)-k)\,x$, the profit-maximizing quantity for a monopolist that faces inverse

[9] Assumption **A1)ii)** is, of course, slightly stronger than concavity. It implies that if a monopolist with a revenue function $x\,V'(x)$ incurs an increase in his (constant) marginal cost, this increase is not fully passed on in price to consumers.

demand $V'(x)$ and marginal cost k. Assumption **A1)** ensures that the profit-maximizing quantity is unique, though this is not required for the Lemma.

Lemma 1: *$k'>k$ implies $x(k')<x(k)$.*

The proof, which is in the Appendix along with all other proofs, uses a standard revealed preference argument (see, for example, Tirole 1988, pp. 66-67). Note that for a given i and t, the quantity of card transactions is $x(i+t-b)$ while the quantity of cash transactions is $x(0)$. For later reference, we define $x(0) \equiv x_0$. Lemma 1 implies that when the sum of the fees to the merchant and to the card user exceeds the merchant's benefit from card use, $i+t > b$, then per capita card transactions are lower than per capita cash transactions.

When is it optimal for the EPN, incorporating the merchant's behavior, to choose a total charge that exceeds the merchant's benefit from card use, $i+t>b$, and, thereby bring about a higher net price for card transactions than for cash? Lemma 2 exploits equations (1) and (2) to show that this will occur if the merchant benefit from card use is not too high (**A3**).

Lemma 2: *Assume **A1)-A3)**. In the environment with merchant surcharges allowed, per capita card use is less than per capita cash use.*

3. The Allocation of EPN Charges Under a No Surcharge Rule

Now suppose that the EPN requires any merchant that accepts its card to charge no more to card users than to cash users, $p_e^M \leq p_c$. Of course, a merchant may refuse and forgo card transactions. The merchant's alternative is to serve the cash market alone, yielding a per-capita level of transactions x_0 and total profit $\alpha\, x_0\, V'(x_0)$. The merchant must be assured at least this amount under the NSR. We refer to this as the "individual rationality" or IR constraint. This section examines merchant's pricing and EPN incentives under an NSR assuming the merchant's IR constraint does not bind (as will occur for some values of α).

When will a pricing constraint such as the NSR bind on the merchant? Recall from Proposition 1 that, for any given $i+t$, EPN profits are constant; however, the prices that a merchant charges to different consumers will vary depending on how the EPN divides its aggregate "tax" between the merchant and cardholders. Lemma 2 shows that the EPN's optimal aggregate fee with no NSR will satisfy $i+t > b$, so the merchant faces a higher net marginal cost of serving card users than cash users; if $t = 0$, a card user's inverse demand is equal to that of a cash user, hence the merchant's higher marginal cost dictates setting a higher price to card users ($p_e^M > p_c$). Lemma 3 below shows that the same effect occurs if, instead, cardholders are levied a positive charge by the EPN but this charge is sufficiently small.

An implication of this observation is that, with a low cardholder fee (which we show is desired by the EPN), we can formulate the no-surcharge rule mathematically as the inequality constraint, $p_e^M \leq p_c$. Although this may seem like the obvious formulation since credit card companies, for example, have historically imposed such rules on their merchant clients, the inequality constraint may obscure other reasons for merchant pricing constraints. Some merchants argue that even without a formal no-surcharge rule, social conventions make it very difficult for them to charge different prices for users of different means of payments. Such an environment would be better captured by the constraint, $p_e^M = p_c$ (that is, a uniform pricing rule rather than a no-surcharge on card use rule). Under **A3)**, however, the effects of the two constraints are the same.

Lemma 3 also shows that, holding constant the sum of the EPN's fee to cardholders (t) and to the merchant (i), when t is low, imposing the NSR lowers cash purchases but raises the quantity of noncash purchases, thereby benefitting the EPN. For any total charge $i + t \equiv k$, define $t_k^* \equiv V'(x(k-b))-V'(x_0) > 0$. Note that $t_k^* < k$.

Lemma 3: *Assume **A1)-A3)** and $k > b$. When merchant surcharges are allowed, $t < t_k^*$ implies $p_e^M > p_c$. Holding k constant and for $t < t_k^*$, if an NSR is imposed and accepted, then cardholder purchases rise and cash purchases fall.*

The intuition for this result is straightforward. Suppose the EPN sets t low enough and i high enough that the merchant charges $p_e^M > p_c$. The NSR then binds on the merchant and induces her to choose a uniform price between these two levels; starting from a uniform price equal to p_e^M, a small move towards p_c imposes a zero first-order loss in the card market while yielding a first-order gain in the cash market (we move closer to the optimal cash price), and similarly starting from p_c^M and moving towards p_e^M.[10]

Lemma 3 suffices to establish that the EPN's profit rises when a binding NSR is accepted. With no NSR holding *i+t* fixed, the EPN generates the same profit by with low *t* and high *i.* Lemma 3 shows that, with an NSR, the EPN's profit increases at the same level of aggregate charges, *i+t.* By revealed preference, any departure from these charges post NSR would further increase its profit. In which direction would the EPN alter its charges? Low card user fees (*t*) ensured that the NSR was binding on the merchant. Next we show that the EPN benefits by choosing a small reduction in *t* accompanied by an equivalent increase in *i*, since this further increases card transactions. Lemma 4 assumes that the merchant continues to accept the card with the no-surcharge rule restriction as *t* falls.

Lemma 4: *Assume* ***A1)-A3)*** *and suppose that the NSR is in place. Fix the sum of EPN charges k > b, and suppose t < t*. If the merchant continues to accept the NSR, as t falls (and therefore i rises), per capita card purchases and EPN profits both rise.*

The intuition behind Lemma 4 is as follows. A cut in *t* and an offsetting increase in *i* leave the EPN's margin unchanged, and hence leave its profit unchanged only if the per-capita quantity of card transactions remains unchanged. This in turn will only happen if the merchant were to raise her price to card users by the full increase in *i*, since card users' inverse demand shifts up by an amount equal to the fall in *t* (equivalently, to the increase in *i*). But since the NSR forces the merchant to charge the same price to cash users as to card users, and since the marginal cost of serving cash users has not risen, the merchant prefers to raise its uniform price by less than the full increase in *i* and accept a lower margin on card sales (given that the merchant's acceptance constraint was not binding). Interestingly, a stronger sufficient condition is required to show that cash transactions decline as *t* falls.[11] However, we show later that the EPN's profit-maximizing prices (*i*, *t*) under the NSR indeed cause cash transactions to decline.

[10] This result echoes the finding that a prohibition of third-degree price discrimination will lead a monopolist to charge an *intermediate* uniform price. Sufficient conditions are that the marginal cost be non-decreasing and the monopolist face independent demands in the various markets, each with quasi-concave profit functions (Nahata et al. 1990, Malueg 1992).

[11] A proof along the same lines as that for Lemma 4 shows that $V'(\bullet)$ convex is a sufficient condition.

Lemma 4 shows that with the NSR, it becomes relevant how $i+t$ is distributed: the EPN prefers lower t (provided the merchant still accepts). Rebates to cardholders (negative t) are often taken as evidence of strong competition among the EPN member banks for cardholders. Lemma 4 offers an alternative interpretation: rebates may be a pricing tactic by a monopolist, designed to increase the impact of the NSR.

Given the incentives for an EPN to raise i and reduce t, what are the constraints that determine the EPN's equilibrium prices? One limit may be institutional. For historical, practical or other reasons, cardholder rebates may not be an option for the EPN.[12] Section 4 investigates the effects of the NSR when rebates are not possible. In this case, the binding constraint may be the non-negativity of t, the merchant's willingness to accept the restrictions implied by the NSR, or both.[13] In fact, as Proposition 2 illustrates, the non-negativity constraint always binds. Section 5 allows for rebates. In this case, we show that the EPN may be constrained either by the IR constraint or by the constraint that the merchant continue to be willing to serve the *cash* market (a sort of incentive compatibility constraint).

4. EPN Pricing With No Rebates ($t \geq 0$)

Lemma 3 showed that an NSR will decrease cash transactions but increase non-cash transactions, holding constant the sum of the EPN's charges to cardholders and to the merchant. However, the NSR also alters the derived demand facing the EPN from the merchant, inducing the EPN to adjust its charges. Furthermore, Lemma 4 shows that the EPN will benefit more from an NSR the lower it sets the cardholder fee, t. Let i_0 be the EPN's optimal charge given $t = 0$ and (for the moment) ignoring the merchant's IR constraint. Whether or not the IR binds depends on

---

12 The phenomenon of cardholder rebates is relatively recent. While credit cards date to the late 1960s/early 1970s, cash rebates were first offered, by Discover in 1986. Rebate cards only became common in the early 1990s with the introduction of the GM Mastercard and other cards that offer reward points associated with co-branding partner companies (such as frequent-flier miles). See generally, Evans and Schmalensee (1999). Today, roughly half of all credit volume is associated with rebates of various sorts. Faulkner and Gray (2000).

13 In (i,t) space, under the NSR, the merchant's level sets have slope strictly less than -1. Therefore, for any given k, the line $t=k-i$ eventually crosses the line given by $\Pi^M(i,t)=\alpha\, x_0\, V'(x_0)$. Thus, if the EPN holds $i+t$ fixed and lowers t, it eventually runs against the merchant IR constraint.

the relative size of the cash market, α. If, at $(0,i_0)$, the IR does not bind, then, by Lemma 4, these prices are optimal for the EPN. If the IR is violated at these prices, in Proposition 2 we provide sufficient conditions under which setting $t = 0$ is still optimal for the EPN.

The analysis in Proposition 2 of the case where the merchant's IR binds requires an additional assumption that enables us to focus solely on the first order conditions of the EPN's optimization problem under an NSR.[14] Define the merchant's profit function and the merchant's profit maximizing choice of cardholder purchases under an NSR as

$$\Pi^{NSR}(i,t) \equiv \max_x \alpha f(V'(x)-t)(V'(x)-t)+x(V'(x)-i-t+b),$$

$$q_e(i,t) \equiv \operatorname{argmax}_x \alpha f(V'(x)-t)(V'(x)-t)+x(V'(x)-i-t+b).$$

The EPN's profit maximization problem, assuming an NSR is imposed, can be expressed as

$$P^{EPN}: \quad \max_{i,t}(i+t)\,q_e(i,t)$$

$$\text{s.t.} \quad \Pi^{NSR}(i,t) \geq \alpha x_0 V'(x_0) \qquad (IR)$$

$$t \geq 0 \qquad (\text{no rebates}).$$

We have not found clearly interpretable general conditions that would ensure that this optimization problem is a concave program for all α, however, it is clear that, for low α, the IR constraint does not bind. In this case, Lemma 4 and Assumptions **A1-A3)** imply that the first order necessary conditions for an optimum are also sufficient. Similarly, for values of α slightly above the level of α that the IR binds, first order necessary conditions are sufficient. Assumption **A4)** is a technical assumption that ensures the first order necessary conditions are also sufficient for an optimum no matter how large is the cash market.

A4) *For all* α, *if* (i^*,t^*) *satisfy the Kuhn-Tucker first order conditions for* P^{EPN}, *then* (i^*,t^*) *solve* P^{EPN}.

The special case **LD)** satisifes **A4)** for all values of α.

---

14 Proposition 2iv) does not rely on **A4).**

Proposition 2: *Assume **A1)-A4)** and suppose that cardholder rebates are not feasible ($t \geq 0$).*

i) The EPN's optimal fee implies $t = 0$ (no cardholder fees). Thus, per capita card and cash purchases are equal.

ii) For all α, merchant profits fall and EPN profits rise with an NSR.

iii) There exists α such that the merchant IR constraint binds if and only if $\alpha > \alpha^*$.*

iv) For $\alpha \leq \alpha^*$ (IR does not bind): under the NSR, cash transactions are lower; card transactions are unchanged if $b = 0$ and are lower if $b > 0$. Thus, total surplus and consumer surplus of each consumer group fall with the NSR.

v) For $\alpha > \alpha^*$ (IR binds): under the NSR, per capita transactions increase with α and, in the limit, approach x_0 (defined in Lemma 1) . Thus, total surplus is lower with the NSR than without for α close to α, and is higher for large α. For α large enough, per capita cardholder purchases and consumer surplus rise. For all α, cash user purchases and consumer surplus fall.*

vi) For $\alpha > \alpha^*$ (IR binds) and in the linear demand case, aggregate purchases ($q_e + \alpha\, q_c$) and total consumer surplus fall with an NSR.

Proposition 2i) shows that the EPN's desire for lower cardholder fees illustrated in Lemma 4 is stronger than the merchants' preference for higher such fees. Even if the merchant's IR constraint binds before the EPN achieves its optimal fee schedule, the EPN will choose to move down the merchant's IR locus to set cardholder fees to zero. Notice that this result also implies that if the non-negativity constraint is relaxed (cardholder rebates are allowed) then the EPN will typically choose to lower t even further. This argument also implies that EPN profits always rise since Lemmas 3 and 4 show that, holding $i+t$ fixed, imposing an NSR and then lowering t until either the IR constraint or the no-rebate constraint binds raises EPN profits. The merchant's profits obviously fall when the IR constraint binds; when the IR constraint does not bind, her profits also fall since the NSR constrains the merchant's pricing and leads to a higher aggregate EPN charge.

The merchant IR constraint binds if and only if the cash market is not too small. (In the linear demand case, (**LD)**, and $b = 0$ this occurs if $\alpha > \alpha^* = 1/3$). When the IR constraint is not binding at the optimal EPN fee schedule, $(i_0, 0)$, the welfare consequences of an NSR are especially stark. The NSR will leave card transactions unchanged if there are no benefits to the merchant from card use ($b=0$), and *reduces* card transactions if $b>0$. Since per capita card and cash transactions are equal under the NSR (and since $t = 0$), cash transactions therefore fall in both cases. Thus, the NSR harms *both* types of consumers (card users are unaffected if b=0).

One might have expected that optimal pricing by the EPN under an NSR would induce a net price for card transactions that is a convex combination of cash and card prices absent the NSR. Instead the ultimate price with the NSR is *higher* than even the original price for card transactions! The conclusions of Proposition 2iv) emerge because of the way the NSR affects the derived demand curve of the EPN. With $b=0$, the NSR induces a proportional rise (by a factor of $(1+\alpha)$) in the EPN's inverse demand curve (i as a function of q_e). Thus, the non-NSR optimal quantity is also the optimal quantity with the NSR but at a higher interchange fee. With $b>0$, the NSR also lowers the elasticity of the derived demand curve and thus prompts the EPN to induce a strict fall in q_e. Since both quantities fall, total surplus must fall.

In the case where the merchant IR binds, Proposition 2v) implies that as the cash market becomes very large, the NSR may induce a rise in social surplus by reducing the double-marginalization distortion compared to no NSR. The logic is that, as α increases, the EPN must cut i in order to continue satisfying the merchant's IR constraint. The EPN nevertheless benefits because the NSR reduces the merchant's margin on card transactions, so that the increased volume of card transactions compensates for the lower i.

The effect on total quantity depends on the precise form of the merchant IR constraint which, in turn, depends on $V(\bullet)$. Proposition 2vi) shows that in the linear demand case, total quantity of transactions falls. This result follows because with equal linear demands by card and cash users (since $t=0$), imposing the NSR would leave total quantity unchanged only if the EPN's charge to the merchant, i, were unchanged; but i increases as the EPN exploits the decreased elasticity of demand that it faces from the merchant. A consequence is that the combined consumer surplus of card and cash users always declines. To see this, observe that total consumer surplus would have fallen even if total quantity had remained unchanged: in such a case, the losses to cash users from the price increase to them would have exceeded the gain to card users from their price decrease, because the two quantities would be equal with the NSR, while initially the cash quantity is higher. In actuality, the NSR causes total quantity to decline, which reduces consumer surplus even further. Thus, while the NSR can increase overall welfare, any such efficiency gains (at least in the linear demand case) are more than appropriated by the EPN.

Proposition 2i) illustrates that the outcome $t=0$ is not stable, since the EPN's profit generally rises as t falls. The next section characterizes what occurs in a special case when the no rebate constraint is relaxed.

5. EPN Pricing With Rebates ($t<0$)

We have already discussed one obvious constraint on the EPN's charges —the merchant's option to refuse to deal with the EPN. If the EPN can offer cardholder rebates, then another, less evident, constraint emerges: the merchant's willingness to continue serving *cash* customers.[15] With very favorable cardholder rebates and given the NSR's prohibition on offering cash customers a lower price than to cardholders, if the cash market is relatively small then a merchant may prefer to set price high enough that cash customers choose not to purchase at all. Such an outcome is clearly not in the EPN's interest, however, since cross-subsidization of cash to card customers then disappears. Observe that this issue does not arise with $t\geq 0$, since in that case (per capita) inverse demand from cash users always exceeds that from card users, so any price that yields sales to the former will also yield sales to the latter.

General results are not available because of the complicated nature of the constraint sets, so we restrict attention to linear demand (**LD)** and where the merchant derives no gross benefit from processing card rather than cash transactions ($b = 0$). To isolate the incremental effects of rebates, Proposition 3 compares the equilibrium under the NSR when rebates are feasible to that without rebates (examined in Proposition 2). Proposition 4 compares the equilibrium under the NSR when rebates are feasible with the equilibrium without the NSR.

15 Technically, this constraint arises because the non-negativity constraint on quantities implies that the merchant's objective function is not quasi-concave.

Proposition 3: *Assume* ***LD)****, b = 0, the NSR holds, and rebates (t<0) are feasible. Then:*

i) For all α , the EPN's optimal choice involves granting card users rebates (t<0), implying a misallocation of transactions towards cards and away from cash.

ii) For low α (<1/4), the requirement that (i,t) induce the merchant to continue to sell to cash customers is a binding constraint on the EPN; for high enough α (above approximately.18) the IR constraint binds (so both constraints bind for α in [.18,.25]).

iii) Compared to the constrained NSR-equilibrium without rebates, card users' consumer surplus is higher while cash users' consumer surplus is lower with rebates. Total consumer surplus is higher with rebates.

iv) Compared to the constrained equilibrium without rebates, total surplus is higher with rebates if and only if α >.18.

The EPN always grants rebates (3i)) because this is a more effective way to stimulate card transactions than by cutting the fee to the merchant. Cutting *t* boosts cardholders inverse demand directly, while the effect of cutting *i* operates through the merchant's price and is therefore dampened because the NSR requires any price cut to apply equally to cash transactions.[16]

If the cash market is large enough (α *>1/5*), the EPN will cut *t* and raise *i* until the merchant's IR binds. For a small cash market (*α <1/5*), the floor on *t* is not the merchant's IR but the need to induce the merchant to continue serving the cash market; consequently, the EPN is not able to extract the full surplus from the merchant solely by choosing *(i,t)* appropriately in conjunction with the NSR.[17]

Part iii) of Proposition 3 is understood as follows. Under the NSR, a cut in *t* induces a rise in *p*, because the increased demand of cardholders prompts the merchant to increase its retail

---

16 Gerstner and Hess (1991) obtain a similar effect in a somewhat different context. They consider a monopolist manufacturer selling through a monopolist retailer (as in our model), that faces two customer groups: low demanders and high demanders, and the latter also incur a higher transaction cost than low demanders of using a rebate coupon. In our model, the NSR plays roughly the same role as their differential transaction costs in motivating rebates.

17 Analytic results are not available for this case (*α <1/4*), and for those values of α all the results in Proposition 3 are proved computationally.

price.[18] Cash users therefore would be harmed by the granting of rebates to cardholders ($t<0$), even if rebates did not imply an increase in the EPN's charge to the merchant. In fact, when the EPN lowers t it also will raise i somewhat, which puts further upward pressure on the retail price p. Cardholders, on the other hand, get the rebate directly while any change in price due to a rise in i is mitigated partly because the cost is partly shared by cash users.

To see why aggregate consumer surplus across cash and card users rises with rebates, we examine the equilibrium with rebates more closely when the equilibrium values of (i, t) are determined by the merchant's IR constraint ($\alpha > 1/4$). With linear demand, if $i+t$ is constant, total quantity stays fixed. Total consumer surplus would remain unchanged *if* quantities remained unchanged, so if we show that $i+t$ falls with a rebate, aggregate consumer surplus will rise.

Suppose the merchant's IR constraint does not bind at i_0, the EPN's optimal interchange fee when $t=0$. Figure 1 illustrates this case. Recall from Lemma 4 that, for a fixed $i+t$, the EPN wishes to lower t in the absence of other constraints. Thus, a movement downward and to the right along the line $i+t = i_0$ (i.e., a cut in t and an equal increase in i) raises EPN profits. The level set of $\Pi^M(i,t)$ equal to the profit the merchant gets from serving the cash market alone yields the IR constraint. This line is downward sloping with slope steeper than -1. Point B in Figure 1 represents the intersection of the line $i+t = i_0$ with this manifold. The optimal solution for the EPN is, then, to move downward and to the right from B along the IR line, until it reaches an indifference curve in (i,t) space that is just tangent to it (point C in Figure 1). This point represents a lower $i+t$ and a lower t compared to $(i_0,0)$. If, instead, the IR constraint binds at $t=0$, (hence $i < i_0$) with rebates feasible, the EPN immediately moves down and to the right along the IR manifold to a point of tangency, so again $(i+t)$ falls relative to the constrained equilibrium with no rebates.

Regarding overall welfare, total output rises with rebates, however, with $t<0$ there is now a misallocation relative to no rebates. Can this misallocation outweigh the quantity effect? It is easiest to compare results with the case where the merchant's IR was binding when $t=0$. In this case, the solution with rebates remains on the IR curve, so the merchant's welfare remains constant. We have shown that aggregate consumer surplus rises and EPN's profits rise. Thus,

[18] Gerstner and Hess (1991) cite empirical evidence that retailers indeed raise their prices in response to manufacturers' granting of rebates to consumers.

total surplus has risen. The remaining cases are not proven analytically, but computations show that total surplus rises then as well.

Can rebates reverse the negative effects of the NSR on total consumer surplus, found in Proposition 2vi) when rebates were not feasible? Proposition 4 addresses the welfare impact of an NSR by comparing the NSR equilibrium when rebates are feasible with the outcome without an NSR, for all values of the relative size of the cash market α. It shows that when the merchant's outside option is weak (the cash market small), an NSR lowers total welfare as well as consumers in aggregate.[19]

Proposition 4: *Assume* ***LD),*** *b = 0, and cardholder rebates are feasible. Compared to the equilibrium with no NSR, the outcome with the NSR results in:*

i) lower total surplus if and only if $\alpha<1/3$*;*

ii) higher consumer surplus of card users and lower consumer surplus of cash users for all α*, and higher total consumer surplus if and only if* $\alpha \in [.22,.33]$.

Proposition 2iv) showed that with $t=0$ and $b=0$, card purchases were the same with and without an NSR. When cardholder rebates are feasible and $b=0$, 4ii) shows that card purchases rise with the NSR. Nevertheless, the NSR reduces overall welfare if the cash market is relatively small ($\alpha<1/3$, result 4i)). Interestingly, this harm occurs even though the total fee ($i+t$) is the same as without the NSR. The source of welfare harm is the misallocation of transactions from cash users towards card users. In this case where $b=0$, the efficient allocation entails equal quantities for each type of consumer. Without the NSR, the EPN's monopolistic charge to the merchant induces too few card transactions, hence a misallocation away from cards. With the NSR and cardholder rebates, the pattern is reversed and misallocation again emerges, albeit in the other direction. This contrasts with the case of $t=0$ addressed in Proposition 2, where the NSR can only reduce welfare if there is a decrease in total quantity (since, for any given total quantity, with $t=0$, allocation between cash and cards is then efficient).

---

19 Private communication from Michelle Bullock (Reserve Bank of Australia) alerted us to an error in a previous version of this Proposition discovered by Julian Wright.

Finally, with rebates, there is a conflict between total surplus and total consumer surplus in that consumer surplus falls under the NSR even when the cash market is large enough that overall welfare rises ($\alpha > 1/3$). In that case, the EPN's pricing drives the merchant's profit down to the IR level, hence the merchant also loses from the NSR. Thus, when the NSR does increase welfare, the EPN's benefits come partly at the expense of the merchant and consumers as a whole.

6. Competitive Card Issuers

To this point, our analysis applies most directly to the case of proprietary networks (such as American Express or Discover) where the EPN is a single card issuer. Alternatively, it describes outcomes when despite multiple card issuers, the issuing industry behaves as if were maximizing issuing banks' joint profits. How do the results change if the EPN is an association of *competitive* issuing banks? In this scenario, member banks issue the cards, and they, rather than the network, set most of the terms to cardholders, including prices (annual fee, interest rate, rebates). Member banks set cardholder fees non-cooperatively and compete with one another for card users. This section explores the effects of an NSR when the EPN is unable to control t.[20]

A sequential/simultaneous game emerges in this case. First, through their partnership with the EPN, banks set the merchant discount fee i and commit to it. Merchants continue to set prices taking i as given but recognizing that t is determined through competition for cardholders by member banks. Member banks of the EPN charging the lowest cardholder fee split the total card transactions equally, while other member banks have no sales. That is, taking i as given, banks compete as Bertrand price setters to cardholders. Suppose that banks can only set cardholder fees in discrete units, ϵ. In this case, the standard Bertrand result follows. The equilibrium t satisfies $t = -i + \epsilon$. Card issuers compete away (virtually) all their rents by granting

[20] The EPN's control over t may also be limited by an inability to credibly commit to cardholder fees even if it is a monopolist in the issuing market. As in the competitive game, a sequential/simultaneous move game results where i is first set, and then the merchant and the EPN choose prices and cardholder fees simultaneously. An earlier version of our paper (Schwartz and Vincent, 2000) analyses this case as well. Rebates tend to be yet lower but, as in the competitive case discussed here, when an EPN cannot commit to t, total surplus declines when an NSR is imposed.

their cardholders a per unit rebate virtually equal to the interchange fee.

When surcharging by the merchant is feasible, total card transactions is derived from equation (1) as

$$\frac{1}{2}-\frac{i+t-b}{2}.$$

A similar argument holds with the NSR except that now the quantity of card transactions becomes[21]

$$\frac{1}{2}-\frac{i+t-b+2\alpha t}{2(1+\alpha)}.$$

It may appear ambiguous from these expressions whether the NSR actually raises card transactions. However, recall the Bertrand assumption implies $i+t = \epsilon \approx 0$. Applying Lemma 3, we have the NSR will bind on the merchant if and only if $t < t_0^* \approx -b/2$. Direct algebra implies that for t below $-b/2$ the second quantity is higher than the first and therefore the NSR increases demand for card transactions. Furthermore, as long as the EPN enjoys *some* profits from transactions ($\epsilon > 0$), it will wish to generate the largest quantity of such transactions possible. Since transactions are decreasing in t, this leads to an incentive for the EPN to fix a high i, inducing its competing issuers to offer large negative values of t (i.e., large card user rebates). While this response induces greater card transactions, a large negative t induces misallocation as between cash and card transactions, biased towards cards. The results are gathered in Proposition 5.[22]

---

21 See equation 3* in the Appendix.

22 The theorem is shown for $b = 0$, however, given the continuity of the environment, quantity and welfare results will continue to hold for b small and positive. They may not hold for b large since the bias towards card purchases may then be socially desirable.

Proposition 5: *Assume **LD)** and b = 0. With perfectly competitive issuers, in the equilibrium under the NSR:*

i) if the mass of cash users is less than the mass of card users, $\alpha<1$, the merchant strictly prefers to accept the NSR and the EPN sets i until merchants are just indifferent between selling to cash customers or not; if $\alpha>1$, the merchant's IR constraint binds;

ii) cash sales are lower than with no NSR, card sales are higher but the total transactions remain the same;

iii) for all values of α, overall consumer surplus is higher than with no NSR but merchant profit and total surplus are lower;

iv) in the limit as the mass of cash users becomes large, the cash-transaction quantity approaches the single monopoly level and the card-transaction quantity approaches the competitive level.

Part 5i) illustrates that, with competitive issuers, the constraint that the EPN ensures that the merchant continues to serve the cash market binds for a larger size of the cash market ($\alpha \leq 1$ rather than $\alpha \leq 1/5$). This is because the stronger tendency to offer rebates under competition among card issuers makes the option of pricing cash users entirely out of the market relatively more attractive to merchants. Part 5ii) is an implication of the constant price constraint of the NSR along with linear demand. Given that total market quantity remains constant and $b = 0$, social surplus must fall since now the NSR along with cardholder rebates forces the per consumer sales to be different across cash and card users. For a similar reason, overall consumer surplus rises: the (quantity weighted) rise in price to cash users is outweighed by the corresponding fall in price to the card users. Result 5iv) shows that as the cash market becomes large relative to cards, the NSR in conjunction with competitive rebates by card issuers succeed in eliminating the distortion in the pricing of card transactions due to the monopolist merchant. The merchant charges a uniformly high (monopoly) price to both card and cash users, but card users receive a rebate and therefore obtain a net price close to the competitive price. However, the net price to cash users is the (uniform) price charged by the merchant. Since the cash market is relatively large by hypothesis, the merchant's price will be driven by the cash market and thus will approach the simple monopoly level.

7. Conclusion

The complex cycle which makes up a typical payments network offers a rich field for economic analysis. Prices play important roles at every link of the cycle. We have shown how attempts to reduce distortions caused by double marginalization by imposing a No Surcharge "Rule" on the merchant at one link spawn further distortions at other links— the NSR induces an electronic payment network both to raise total charges and to shift charges away from cardholders to merchants. Throughout, the EPN profit increases while the merchant and cash users are harmed. Card users typically gain (but can lose when rebates are not feasible). Overall welfare can increase or decrease with the NSR, because of the tradeoff between the expanded total quantity of transactions and the misallocation from cash to cards. An increase is more likely if the number of cash users relative to card users is large enough. In such case, however, total consumer surplus tends to fall with the NSR. Allowing cardholder rebates, under the NSR, harms cash users but benefits card users by more. Thus, cardholder rebates (which entail somewhat higher charges to merchants) are beneficial *if* the relevant policy goal is total consumer welfare, and if repeal of the NSR is not an option (e.g., because it is dictated by merchants' inherent reluctance to set different prices not by explicit prohibitions).

Our emphasis has been on the impact of limits on pricing behavior at the merchant level when there exists some power over price. To highlight this effect, we assumed monopoly pricing at both the merchant and EPN levels, but we conjecture that similar effects will arise whenever there remains a margin between price and cost at both levels. We have also abstracted away from consumers' choice of the means of payment in order to focus on the impact of the constraint on the level of purchases. Extensions of this research would include analyses of the effects of the "rule" under a broader class of merchant market structures and with endogenous consumer choice of the means of payment. Another direction will be to examine how the NSR influences the nature of competition among rival electronic payments networks both in pricing and in other practices such as the tying of multiple cards to merchants. The growing role of new payments technologies such as online and offline debit cards makes it increasingly important to obtain a fuller understanding of the impact of constraints on merchants' abilities to charge different prices for different means of payments.

Appendix

Proof of Lemma 1: Let $x=x(k)$, $x'=x(k')$. By definition, $(V'(x)-k)x \geq (V'(x')-k)x'$ and $(V'(x)-k')x \leq (V'(x')-k')x'$ so $(k'-k)\, x \geq (k'-k)x'$.This implies $x \geq x'$. Now, suppose $x=x'$. The first order condition for $x(k)$ satisfies $xV''(x)+V'(x)-k=0$. For $k'>k$, then, $xV''(x)+V'(x)-k'<0$. ‖

Proof of Lemma 2: The derivative of the EPN's profit function is

$$d\Pi^e(x)/dx=(xV''(x)+V'(x))+x^2\, V'''(x) +2x\, V''(x) +b.$$

At x_0, the term in parentheses is zero by the optimality of x_0 for the merchant in the cash market. Assumption **A3)** implies the remaining term is negative. The concavity of the EPN profit function implies the right side declines as x rises. Therefore, EPN profits are declining in x for $x \geq x_0$. ‖

Proof of Lemma 3: By Lemma 1, $x(k-b) < x_0$ and therefore, $V'(x(k-b))>V'(x_0)$. Since $t_k^* \equiv V'(x(k-b))-V'(x_0)$, for all t,i such that $i+t=k$, $x(k-b)$ remains constant and $t < t_k^*$ implies $V'(x(k-b))-t=p_e^M>V'(x_0)=p_c^M$. Consider an optimal choice of (q_e, q_c) that solves the merchant's profit maximization problem with an NSR. Suppose that $q_e < x(k-b)$. The pair $(x(k-b),q_c)$ is also feasible for the merchant since $V'(q_c)+t \geq V'(q_e)$ implies $V'(q_c)+t \geq V'(x(k-b))$ by the concavity of $V(\bullet)$. But the choice of (q_e, q_c) over $(x(k-b),q_c)$ then implies that

$$q_e\, (V'(q_e)-k+b) \geq x(k-b)(V'(x(k-b))-k+b)$$

which violates the definition of $x(k-b)$. A similar proof shows $q_c \leq x_0$. Now suppose $q_e=x(k-b)$. The merchant's first order condition with respect to q_e under the NSR constraint is $q_eV''(q_e)+V'(q_e)-i-t+b-\lambda V''(q_e)$ where $\lambda > 0$ is the multiplier on the constraint imposed by the NSR. Evaluating this expression at $x(k-b)$ yields $-\lambda V''(x(k-b))>0$ since the first terms are the merchant's first order condition with no NSR and equal zero at $x(k-b)$. Therefore, merchant profits are strictly increasing in q_e at $q_e=x(k-b)$. ‖

Proof of Lemma 4: From **A1)**, $f(p)$ is the demand curve of cash users with $pf(p)$ concave. The demand curve of card users is $f(p+t)$. Let $i+t = k$ and let p denote the optimal (uniform) price charged by the merchant under an NSR when the card user fee is t (so $i = k-t$). Similarly, let p' denote the optimal uniform price charged by the merchant when the card user fee is $t' < t$. Finally, for convenience, set $t-t' \equiv \Delta > 0$. By definition of p, charging a price p under the fee profile, $(k-t,t)$ yields higher merchant profits than charging a price $p'-\Delta$. Note that this second price implies a net price to card users of $p'+t'$. Thus,

$$\alpha pf(p)+(p+t-(k-b))f(p+t) \geq \alpha(p'-\Delta)f(p'-\Delta)+(p'+t'-(k-b))f(p'+t').$$

Similarly, under the fee profile, $(k-t',t')$, p' raises more profits than charging a price $p+\Delta$.

$$\alpha p'f(p')+(p'+t'-(k-b))f(p'+t') \geq \alpha(p+\Delta)f(p+\Delta)+(p+t-(k-b))f(p+t).$$

Adding the two inequalities and eliminating the common terms which denote revenues in the card market and dividing by α, yields

$$p\,f(p)-(p+\Delta)\,f(p+\Delta) \geq (p'-\Delta)f(p'-\Delta)-p'f(p').$$

Recall that p and p' are higher than the price which maximizes $p\,f(p)$. Suppose that $p'+t'>p+t$. This implies $p'-\Delta>p$. But this violates the assumption of concavity of $p\,f(p)$ since the slope of the revenue function must become steeper as we move further to the right of the maximum point. ‖

Proof of Proposition 2: We first prove 2iv) by analysing the outcome when the IR constraint does not bind. Lemma 4 implies that in this case, the optimal solution is $t = 0$ which implies that per capita cash and card purchases are the same. This gives the first order condition of the merchant, $i = b+(1+\alpha)(V'(x)+xV''(x))$. Define $q_\alpha = \text{argmax}_x\, x(b+(1+\alpha)(V'(x)+xV''(x)))$ to be the quantity of card-user transactions which maximizes EPN profits with the NSR. Assumption **A2)** implies this is unique. Note that q_0 maximizes profits with no NSR. If $b=0$, then the definition indicates that $q_0 = \text{argmax}_x\,(1+\alpha)\,x((V'(x)+xV''(x)))$ and so q_0 also solves the EPN's problem with the NSR. Therefore, assume $b > 0$. By definition,

$$q_0\,(b+V'(q_0)+q_0\,V''(q_0)) \geq q_\alpha(b+V'(q_\alpha)+q_\alpha\,V''(q_\alpha))$$

and

$$q_0\,(b+(1+\alpha)(V'(q_0)+q_0\,V''(q_0))) \leq q_\alpha(b+(1+\alpha)(V'(q_\alpha)+q_\alpha\,V''(q_\alpha))).$$

Subtract the two inequalities and divide by $-\alpha$ to get

$$q_0(V'(q_0)+q_0 V''(q_0)) \leq q_\alpha(V'(q_\alpha)+q_\alpha V''(q_\alpha)).$$

Suppose that $q_\alpha > q_0$,. Then $b\, q_0 < b\, q_\alpha$. This implies

$$q_0(b+V'(q_0)+q_0 V''(q_0)) < q_\alpha(b+V'(q_\alpha)+q_\alpha V''(q_\alpha))$$

which violates the definition of q_0. The EPN first order conditions with the NSR, evaluated at q_0, indicates that EPN profits are strictly declining in quantity at that point:

$$\frac{\partial \pi^e(x)}{\partial x}_{|x=q_0} = b(1-(1+\alpha)) = -\alpha b < 0$$

so $q_\alpha < q_0$ given $b>0$.

Now suppose the IR binds and consider (t,i) space. (For the purposes of this argument, (t,i) space is more convenient than (i,t) space shown in Figure 1.) At $t = 0$, and i such that the merchant IR curve binds, the slope of the EPN level set is steeper than the slope of the merchant IR curve. This implies that this point is a local maximium. Under an NSR, the lagrangian representing the merchant's profit maximization problem is

$$L(q_c\ q_e,\lambda;i,t)=\alpha\, q_c\, V'(q_c)+ q_e\, (V'(q_e)-i-t+b)+\lambda(V'(q_c)+ t - V'(q_e)).$$

The first order conditions from this problem satisfy

1A) $(\lambda+ \alpha\, q_c)V''(q_c) +\alpha\, V'(q_c) = 0,$

2A) $(-\lambda+ q_e)V''(q_e) + V'(q_e) -i-t +b = 0,$

3A) $V'(q_c)+ t - V'(q_e) = 0.$

Totally differentiate this system and evaluate it at $t = 0$, so $q_c = q_e \equiv Q$, $V''(q_c)=V''(q_e) \equiv V''$, and $V'''(q_c)=V'''(q_e) \equiv V'''$. This yields the system

$$\begin{bmatrix} V'' & V'''(\lambda+\alpha Q)+2\alpha Q V'' & 0 \\ -V'' & 0 & V'''(\lambda+\alpha Q)+2\alpha Q V'' \\ 0 & V'' & -V'' \end{bmatrix} \begin{bmatrix} d\lambda \\ dq_c \\ dq_e \end{bmatrix} = \begin{bmatrix} 0 \\ di+dt \\ -dt \end{bmatrix}$$

Using **1A)** to substitute in for λ, we can solve this system to obtain (at $t = 0$)

$$\frac{\partial q_e}{\partial t} = \frac{1+2\alpha-\alpha V'''\frac{V'}{V''}}{(1+\alpha)(QV'''+2V'')} \leq \frac{\partial q_e}{\partial i} = \frac{1}{(1+\alpha)(QV'''+2V'')} < 0.$$

The first inequality follows from the concavity of the merchant revenue function in price and the second from the concavity of the revenue function in quantity (**A1)**. Equations **1A** and **2A** imply that $i = b - \lambda (1+ \alpha) V''/\alpha$. This yields

$$i\left(\frac{\partial q_e}{\partial i}-\frac{\partial q_e}{\partial t}\right) = b\left(\frac{\partial q_e}{\partial i}-\frac{\partial q_e}{\partial t}\right)+\lambda\frac{2V''-V'''\frac{V'}{V''}}{QV'''+2V''} \geq \lambda\frac{2V''-V'''\frac{V'}{V''}}{QV'''+2V''}$$

and

$$i\frac{\partial q_e}{\partial i}+Q = b\frac{\partial q_e}{\partial i}-\frac{\lambda V''}{\alpha(QV'''+2V'')}+Q \leq \frac{V'+QV''}{(QV'''+2V'')}+Q.$$

The inequality follows because q_e is decreasing in i and from the substitution for λ . Combining these results yields

$$\frac{i\left(\frac{\partial q_e}{\partial i}-\frac{\partial q_e}{\partial t}\right)}{i\frac{\partial q_e}{\partial i}+Q} \geq \frac{\lambda(2V''-V'''\frac{V'}{V''})}{V'+3QV''+Q^2V'''}.$$

Now consider the level sets of the merchant and the EPN in *(t,i)* space. The slopes at $t = 0$ are given by

$$\left.\frac{di^e}{dt}\right|_{t=0} = -\left(1 - \frac{i\left(\frac{\partial q_e}{\partial i} - \frac{\partial q_e}{\partial t}\right)}{i\frac{\partial q_e}{\partial i} + Q}\right), \qquad \left.\frac{di^M}{dt}\right|_{t=0} = -\left(1 - \frac{\lambda}{Q}\right).$$

Subtracting the second from the first yields, after substituting the inequality from above,

$$\left.\frac{di^e}{dt}\right|_{t=0} - \left.\frac{di^M}{dt}\right|_{t=0} \geq \left(\frac{-\lambda}{V''Q}\right)\left(\frac{QV''+V'}{V'+3QV''+Q^2V'''}\right)\left(V''+V'''Q\right) \geq 0.$$

The first term is positive because λ is non-negative and V is strictly concave. The numerator in the second term is positive because $Q \leq x_0$ and $xV''(x) + V'(x)$ is decreasing. The denominator is the marginal revenue curve of the EPN and is non-positive because if the IR constraint is binding, i is less than what it would choose (at $t = 0$) if the IR did not bind. The final term is negative by **A1).** Since the slope of the EPN level set exceeds the slope of the merchant's IR curve at $t = 0$, this point represents a local maximum. By Assumption **A4)**, this is also a global maximum. Thus 2i) follows.

To prove 2iii) and 2v) note that at $t = 0$, equations **1A)** and **2A)** imply that $(1+\alpha)(Q^2 V'' + QV') = (i+t-b) Q$ so (substituting in for $(i+t-b) Q$) the IR constraint is equivalent to

$$QV'(Q) - (i+t-b)Q = -Q^2 V''(Q) \geq \alpha x_0 V'(x_0)/(1+\alpha).$$

The right side is increasing in α. Concavity of the merchant revenue function in quantity, **A1),** implies the left side is increasing in Q. For low α, the constraint does not bind when the EPN selects its globally optimal Q at $(i_0,0)$. As α rises, the constraint binds and the EPN must offer a successively higher Q (lower i) in order to induce the merchant to participate. The limit of $\alpha/(1+\alpha)$ as α becomes large is 1 so Q must approach x_0. (Recall that the merchant's first order condition implies $-x_0^{\ 2} V''(x_0) = x_0 V'(x_0)$). Before reaching the limit, though, $Q < x_0$ so cash users purchases and surplus fall with the NSR. Lemma 2 implies that eventually cardholder purchases and surplus must rise with the NSR.

2vi) With an NSR and linear demand, merchant profits are $(1+\alpha)Q^2$ so the merchant's IR constraint can be written $Q^2 \geq \alpha/(4(1+\alpha))$. The participation condition implies that the constraint does not bind for $\alpha \leq 1/3$. With no NSR, the total quantity of transactions is $(1+b)/4+\alpha/2$ and 2iv) implies that $Q=1/4$. For $\alpha > 1/3$, the constraint binds and determines Q. The NSR raises total quantity if and only if,

$$\frac{1+\alpha}{2}\sqrt{\frac{\alpha}{1+\alpha}} = \frac{\sqrt{1+\alpha}\sqrt{\alpha}}{2} \geq \frac{1+b}{4}+\frac{\alpha}{2}$$

$$\alpha+\alpha^2 \geq \frac{1+b^2}{4}+(1+b)\alpha+\alpha^2$$

This is impossible so total quantity falls. Computations show that total surplus exceeds total surplus with no NSR at $\alpha = 1.53$. To see that the net effect on consumer surplus is negative, note that if total quantity had remained constant with card use rising and cash use falling, consumer surplus would have to fall since the loss in consumer surplus for cash users -- whose initial per capita quantity exceeded that of card users -- would necessarily be larger than the gain enjoyed by card users. Since, in fact, total quantity actually falls, the consumer surplus effects are worse. ‖

Proof of Proposition 3: The merchant's optimization problem, yields as a solution,

$$p=\frac{1+\alpha+i-t}{2(1+\alpha)}, \quad q_c=\frac{1+\alpha-i+t}{2(1+\alpha)}, \quad q_e=\frac{1+\alpha-i+t(1+2\alpha)}{2(1+\alpha)} \tag{3*}$$

Thus for changes in i and t, di, dt, we have $dp=(di-dt)/(2(1+\alpha))$. Note that this implies that as t falls, p rises implying that lower t's harm cash users through higher merchant prices. Also,

$$dq_c = -(di-dt)/(2(1+\alpha)), \quad dq_e = -(di+(1+2\alpha)dt)/(2(1+\alpha)).$$

Therefore, change in *total* quantity is $\alpha dq_c+dq_e = -((1+\alpha)di-(1+\alpha)dt)/(2(1+\alpha))=-(di+dt)/2$ and total quantity remains the same for any changes in i and t such that $i+t$ remains constant and rises for any changes such that $-dt>di>0$.

3i) From the envelope theorem, substituting in for q_e and λ,

$$d\Pi^M(i,t)=-q_e di+(\lambda-q_e)dt = -q_e di+-(p-i)dt.$$

In the linear case, with no NSR, $q_e=(p-i)$. With the NSR, $q_e>(p-i)$, the volume of card sales exceeds the profit margin on card sales. A small fall in i increases merchant's profits by the number of card sales. A small fall in t increases card sales by a unit and therefore increases merchant profits only by the per unit profit margin. Thus, in (i,t) space, with an NSR, the merchant's indifference curves are steeper than -1. The EPN's profit function can be expressed as $\Pi^e(i,t)= (i+t)q_e(i,t)$. Differentiating implies

$$d\Pi^E(i,t)=(q_e+(i+t)\partial q_e/\partial i\,)di+(q_e+(i+t)\partial q_e/\partial t\,)dt$$

With an NSR, $\partial q_e/\partial t<\partial q_e/\partial i<0$. Using the definitions of q_e we get for any level set

$$d\Pi^E(i,t)=0=((1+\alpha)-2(i+t)-2\,\alpha t\,)di+((1+\alpha)(1-2(i+t))-2\,\alpha t\,)dt$$

Consider any point such that $t=0$. Since $i>1/2$ (the value which maximizes EPN profits with no NSR), the coefficient on dt is negative and so EPN indifference curves are positively sloped and, if the $t\geq 0$ constraint is relaxed, the EPN will move downward and to the right along the merchant's IR curve.

3ii): By computation. (All conclusions for $\alpha\leq 1/4$ are based on computations.)

3iii): Consumer surplus is given by $CS=(\alpha\, q_c^2+q_e^2)/2$. So,

$$2\Delta CS=q_c\,\alpha\,\Delta\, q_c+q_e\,\Delta q_e=q_c\,(\alpha\,\Delta\, q_c+\,\Delta q_e)+(q_e-q_c)\Delta q_e$$

Suppose that the IR binds at the optimal solution with $t=0$. The optimal solution with the $t\geq 0$ relaxed is at a point downward and to the right of this point. Since the IR curve has slope steeper than -1, this implies a rise in i, and a fall in t and $i+t$. The rise in i and fall in t imply a rise in merchant price and therefore a fall in cash user surplus. The fall in $i+t$ implies a rise in card user surplus, a rise in total quantity and therefore a rise in total consumer surplus. Revealed preference implies that EPN profits rise and, since we remain on the merchant's IR curve, merchant profits stay the same. Thus, total surplus rises when the $t\geq 0$ constraint is relaxed from a point at which the IR constraint binds. The case when the IR constraint does not bind at $t=0$ is less immediate. The optimal interchange fee at this point is $i_0=(1+\alpha)/2$. Note that using the EPN level curves,the locus of points $L=\{(i,t):((1+\alpha)(1-2(i+t))-2\,\alpha t\,)=0\}$ denote when EPN profits are locally stationary in t and so its indifference curves become negatively sloped below that line. For α above $.1$, the IR curve intersects the line $i+t=i_0$ (point B in Figure 1) before it intersects L so EPN indifference curves are positively sloped at B. Therefore, at point B the EPN will wish to move downward and to the right along the IR curve. This implies a further rise in i and fall in t and a fall in $i+t$ which by the argument above, implies the cash users surplus falls, and card and total consumer surplus rises.

3iv): Total surplus is given by $TS=\alpha\,(1-.5q_c)q_c+(1-.5q_e)q_e$. So,

$$\Delta TS=(1-q_c)\,\alpha\,\Delta\,q_c+(1-q_e)\Delta q_e=(1-q_c)(\alpha\,\Delta\,q_c+\Delta q_e)+(q_c-q_e)\Delta q_e$$

While we see that total quantity rises also in the case where the IR constraint does not bind at $(i_0,0)$, which has a positive impact on total surplus, the second term reflects a misallocation and is negative. Computations indicate that total surplus also rises in this case but there do not appear to be formal arguments except for pure algebra supporting the conclusion. Note that if it were the case that total surplus falls, then, interestingly, it would be because the loss in merchant profits outweigh the gains to consumers and the EPN. ‖

Proof of Proposition 4: Shown by computation. ‖

Proof of Proposition 5: 5i): Solving the merchant participation constraint simultaneously with the constraint $t=-i$, yields

$$i^{comp(IR)}\leq\frac{1}{2}\frac{\sqrt{1+\alpha}}{\sqrt{\alpha}}$$

The constraint that the merchant continue to be willing to serve the cash market is $t>i-(1+\alpha)^{.5}/2$. Using $t=-i$, this yields a value

$$i^{comp(IC)}\leq\frac{1}{2}\sqrt{1+\alpha}$$

The lowest value for i^{comp} is the binding constraint. The second one is lower than the first if and only if $\alpha<1$.

5ii),iii): Use the quantity equations from Equation (3*) in Proposition 3 for per capita purchases and $t=-i$ to get $\alpha q_c=\alpha(.5-i/(1+\alpha))$ and $q_e=(.5+\alpha\, i/(1+\alpha))$. Summing the two yields total quantity $(1+\alpha)/2$ which is independent of i and is equal to the total quantity of purchases with competitive issuers and no NSR.. Conditional on total quantity remaining constant, social surplus is maximized when the cash and non-cash quantities are the same. Any value of t strictly less than zero along with the NSR, violates this condition, so social surplus must fall. Consumer surplus rises because, holding total quantity fixed, the loss to cash consumers from the higher price is more than compensated by the gain to EPN consumers from the lower price. Using $q_e=(.5+\alpha\, i/(1+\alpha))$ and letting α grow large yields i approaches $1/2$, EPN quantity approaches 1 and cash quantity approaches $1/2$. ‖

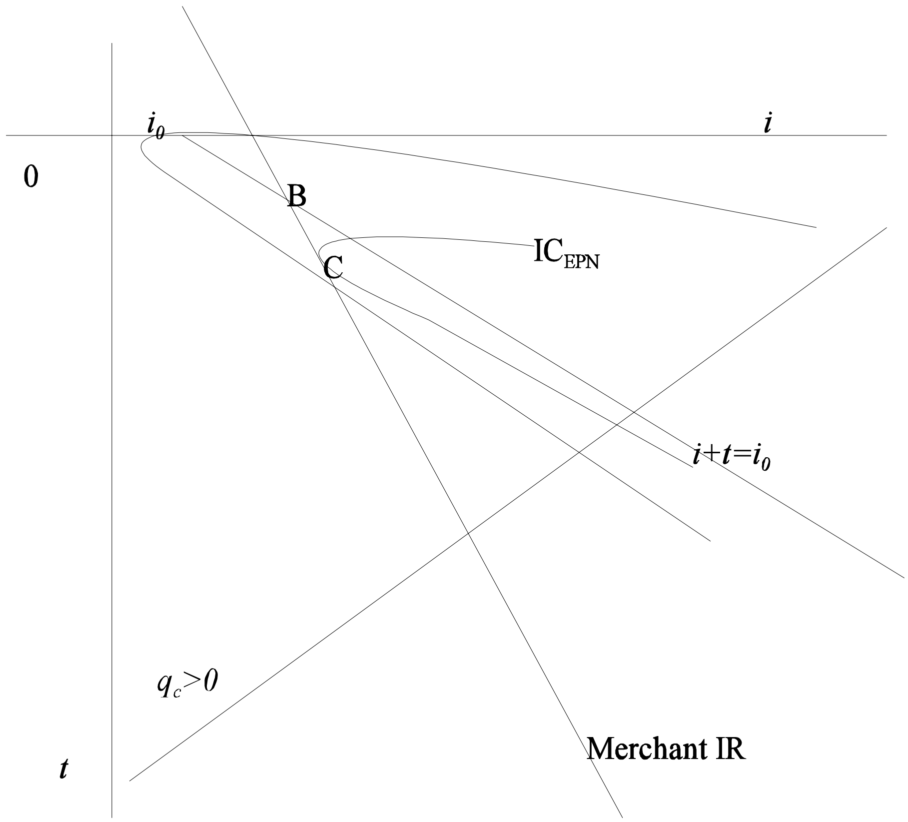

Figure 1

References

Baxter, William (1983). "Bank interchange of transactional paper: Legal and economic perspectives." *Journal of Law and Economics 26* (October): 541-588.

Carleton, Dennis and Alan Frankel (1995a). "The antitrust economics of credit card networks." *Antitrust Law Journal 63 (2)*: 643-668.

Carleton, Dennis and Alan Frankel (1995). "The antitrust economics of credit card networks: Reply to Evans and Schmalensee Comment." *Antitrust Law Journal 63 (3)*: 903-915.

Chain Store Age, *Fourth Annual Survey of Retail Credit Trends,* January 1994, section 2.

Chakravorty, Sujit and Alpah Shah (2001). "A Study of the Interrelated Bilateral Transactions in Credit Card Networks". Federal Reserve Bank of Chicago. EPS-2001-1.

Chakravorty, Sujit and Williams Emmons (2001). "Who Pays for Credit Cards". Federal Reserve Bank of Chicago. EPS-2001-1.

Evans, David and Richard Schmalensee (1995). "Economic aspects of payment card systems and antitrust policy toward joint ventures." *Antitrust Law Journal 63 (3)*: 861-901.

Evans, David and Richard Schmalensee (1999). *Paying with Plastic: The Digital Revolution in Buying and Borrowing.* MIT Press.

Faulkner and Gray (2000), *Card Industry Directory*, 2000 edition, Chicago.

Gerstner, Eitan, and James D. Hess (1991). "A theory of channel price promotions." *American Economic Review* 81 *(September)*: 872- 886.

Malueg, David (1992)."Direction of price changes in third-degree price discrimination: Comment." Freeman School of Business, Tulane University, Working Paper 92-ECAN-95.

Nahata, Babu, Krzysztof Ostaszewsi, and P.K. Sahoo (1990). "Direction of price changes in third-degree price discrimination." *American Economic Review* 80 *(December)*: 1254-1258.

The Nilson Report, May, 2000.

Reserve Bank of Australia/Australian Competition and Consumer Commission (2000). *Debit and Credit Card Schemes in Australia: A Study of Interchange Fees and Access.* Reserve Bank of Australia (October).

Reserve Bank of Australia(2001). "Designation of Credit Card Schemes in Australia". Media

Release, 12 April.

Rochet, Jean-Charles and Jean Tirole (1999). " Cooperation among competitors: The economics of credit card associations." IDEI Mimeo. Version March 1, 1999.

Salop, Steven (1990), "Deregulating self-regulated shared ATM networks," *Economics of Innovation and New Technology* volume 1, numbers 1-2 (December): 85-96.

Schmalensee, Richard (1999). " Payment systems and interchange fees." MIT mimeo. Version of 06/04/99.

Schwartz, Marius and Daniel R. Vincent (2000). " Spreading the Pain Can Aggravate the Injury: Uniform Pricing Constraints in Payments Networks." University of Maryland, mimeo. Version of 10/00.

Tirole, Jean (1988). *The Theory of Industrial Organization.* MIT Press.

Wright, Julian (2000). "An Economic Analysis of a Card Payment Network." Attachment 2, *Credit Card Schemes in Australia.* (Response to RBA/ACCC Study, 2000). VISA International Service Association.. January, 2001.