Differentiated Interchange Fees*

Hans Zenger†

16 December 2011

Abstract

Payment networks typically differentiate their interchange fees (IFs) by setting a variety of sector-specific IFs for the same payment card. While the previous literature on IFs has focussed on the optimal level of IFs, this paper addresses the optimal structure of IFs, i.e. whether or not IF differentiation is desirable. It is shown that it is generally efficient for a regulator to leave the decision on the structure of IFs to payment networks, even in cases where regulation of the (average) IF level itself is welfare-enhancing.

JEL classification: G21, L31, L42

Keywords: interchange fees, two-sided markets, price discrimination

*The views expressed in this article represent those of the author only, and do not necessarily reflect the opinions of other CRA staff or of CRA's clients.

†Charles River Associates, Avenue Louise 81, 1050 Brussels, Belgium. E-mail: hzenger@crai.com. Tel. +32-2-6271400.

1 Introduction

An ever-increasing amount of economic transactions is paid through debit and credit cards.[1] Open payment networks such as Visa and MasterCard typically operate their schemes with an interchange fee (IF), a per transaction payment from acquiring banks (the banks of merchants) to issuing banks (the banks of cardholders). Since these fees have traditionally been agreed collectively by competing banks, they have come under regulatory scrutiny by competition authorities and financial regulators in many jurisdictions including the European Union, the U.S., France, Australia, Switzerland, Spain, and the United Kingdom.

Following Baxter (1983), a large literature has spawned that assesses privately and socially optimal IFs, including Rochet and Tirole (2002, 2011), Schmalensee (2002), Wright (2003, 2004), Guthrie and Wright (2007) and Zenger (2011). These papers typically focus on the efficient level of IFs.[2] For both payment schemes and regulators, another question of key interest has been the optimal structure of IFs. In particular, payment schemes typically do not apply a single IF rate per type of payment card, but individualize fees for different commercial sectors. For instance, MasterCard applies differentiated rates for restaurants, utilities, real estate services, insurance companies, service stations, telecommunications operators, government services, and many other industries.[3] These differentiated fees reflect the heterogeneous benefits that merchants in different industries derive from accepting payment cards.

This paper addresses the question of whether or not an existing IF regulation should encompass restrictions on the freedom to differentiate IFs. I.e., should regulatory caps on IFs apply to each

[1] According to Gerdes (2008), payment cards were used in the U.S. in more than 47 billion transactions in 2006.

[2] An exception is Shy and Wang (2011), who analyze whether proportional or fixed per-transaction fees are more efficient.

[3] See MasterCard Worldwide, U.S. and Interregional Interchange Rates, http://www.mastercard.com/us/merchant/pdf/MasterCard_Interchange_Rates_and_Criteria.pdf.

individual IF, or should they only apply to the weighted average of transactions with a particular type of payment card? To answer this question, I extend the standard IF model with heterogeneous merchants (Wright, 2004, and Rochet and Tirole, 2011) by allowing the payment network to apply IF schedules that make individual fees contingent on the sector in which a merchant operates.

The welfare effects of price differentiation in the presence of a regulatory cap on average prices has previously been analyzed by Armstrong and Vickers (1991) for the case of one-sided markets. They show that in general, price discrimination around an exogenously given average has ambiguous welfare properties. If the regulated average price is sufficiently close to the efficient level (marginal cost), however, price discrimination always *decreases* welfare.[4] As is shown in this paper, this intuition cannot be transposed to two-sided markets. Under a condition that is both theoretically and empirically mild, it is always welfare-enhancing for a regulator not to impose restrictions on the IF structure.

2 Interchange fee differentiation

The base model, which is extended to allow for price differentiation, closely follows Rochet and Tirole (2011). There is a continuum of measure one of industries and of consumers. Each consumer is exogenously matched with each industry. Consumers have inelastic demand and buy one good per industry by using a payment card or cash. All merchants in a given industry derive the same transactional benefit b_S from a card transaction relative to a cash transaction. The value b_S that characterizes an industry is drawn independently from a distribution function $G(b_S)$ with positive and continuously differentiable density $g(b_S)$ over the interval $\left[\underline{b}_S, \bar{b}_S\right]$. In order

[4]Armstrong and Vickers also consider a more sophistocated average price regulation, where the weights that are applied to different transactions are not given by observed demand levels at prevailing prices, but by the counterfactual demand levels that would arise under uniform pricing. However, Armstrong and Vickers rightly point out that the implementation of such schemes would require more information than real world regulators normally possess.

for industry-specific pricing to be feasible, assume that the differences in transactional benefits across industries are observable. A given consumer derives the same benefit b_B from a card payment relative to a cash payment for all transactions (and hence always prefers one or the other payment instrument). The value b_B that characterizes a consumer is drawn independently from a distribution function $H(b_B)$ with positive and continuously differentiable density $h(b_B)$ over $[\underline{b}_B, \bar{b}_B]$.

Each card transaction costs the issuing bank c_B and the acquiring bank c_S. The IF a is a per transaction transfer from acquiring to issuing banks. Payment networks can price differentiate between transactions in different industries. Hence, sector-specific IFs are represented by some schedule $a(b_S)$, which maps industry types into IFs. Since each consumer with a preference for card use makes one purchase in each industry that accepts cards, the issuers' average interchange revenue per card transaction α is the same for all cardholders.[5] Consider the situation where IFs are subject to regulatory oversight. Hence, some cap $\alpha \leq \bar{a}$ is imposed on the (weighted) average IF level by regulation.[6] Since for all practical purposes IF caps on the average IF level bind in real world regulation (there would be no point to regulate, otherwise), it is assumed that $\alpha = \bar{a}$ in what follows.[7]

In addition to setting the IF level, the regulator may also determine the IF structure by imposing a restriction on sector-specific IFs. If differentiation is prohibited, regulation implies $a(b_S) = \bar{a}$ for all $b_S \in [\underline{b}_S, \bar{b}_S]$. Otherwise, the payment network chooses $a(b_S)$ to maximize its member banks' profits subject to the regulatory cap on average fees.

[5] α will be formally characterized below in equation (1).

[6] This regulation is not required to be optimal in any sense, but it might be. Rochet and Tirole (2011) and Wright (2004) derive socially optimal IFs and show that they may exceed privately optimal fees under certain conditions.

[7] For instance, the Australian regulation of the average IF level of credit cards has been binding at any point since its inception. Likewise, the European Commission's settlements with Visa and MasterCard have led to average IF levels that were always binding for both networks.

Given net per transaction costs $c_I = c_B - \alpha$ on the issuing side, competing symmetric issuers set a symmetric equilibrium per transaction fee of $p_B(c_I)$ to consumers, with $p'_B > 0$. Acquirers are assumed to be perfectly competitive. Since merchants of different types b_S pay different IFs $a(b_S)$, net per transaction costs for acquirers are given by $c_A(b_S) = c_S + a(b_S)$. The equilibrium per transaction fee p_S for merchants is therefore given by $p_S(c_A) = c_A(b_S)$. While the acquiring markup is zero due to perfect competition, the issuing markup is

$$m(c_I) = p_B(c_I) - c_I.$$

For given values of p_B and p_S, the value of b_S at which merchants are indifferent between accepting or rejecting card payments is denoted by $b_S^m(p_B, p_S)$.[8] The measure of merchants that accept cards is then given by the merchant demand function $D_S(b_S^m) = 1 - G(b_S^m)$ and the average transactional benefit of merchants that accept cards is

$$\begin{aligned} \beta_S(b_S^m) &= E(b_S \mid b_S \geq b_S^m) \\ &= \frac{\int_{b_S^m}^{\bar{b}_S} b_S dG(b_S)}{1 - G(b_S^m)} \text{ for } b_S^m < \bar{b}_S \end{aligned}$$

with $\beta'_S > 0$.[9] Likewise, for a given value of p_B, the value of b_B at which consumers are indifferent between using a card or cash is denoted by $b_B^m(p_B) = p_B$ (consumers use cards if and only if $b_B \geq p_B$). The measure of consumers that use cards (if accepted by the merchant) is then given by the consumer demand function $D_B(b_B^m) = 1 - H(b_B^m)$ and the average benefit of consumers

[8] As with a uniform IF, it is optimal for the scheme to set $a(b_S)$ so as to serve a continuous interval $[b_S^m, \bar{b}_S]$ (the upper section of industries), i.e. not to suppress card acceptance in industries with relatively higher willingness to accept cards. Given the regulatory cap, it is "cheaper" to induce merchants with large b_S to accept cards.

[9] As shown by Rochet and Tirole (2002, 2011), if card acceptance attracts purchasers to stores, merchants may be willing to accept cards even if the cost of accepting cards exceeds their transactional benefits, so $c_S + a(b_S) > b_S$.

that use cards is

$$\begin{aligned}\beta_B\left(b_B^m\right) &= E\left(b_B \mid b_B \geq b_B^m\right) \\ &= \frac{\int_{b_B^m}^{\bar{b}_B} b_B dH\left(b_B\right)}{1-H\left(b_B^m\right)} \text{ for } b_B^m < \bar{b}_B\end{aligned}$$

with $\beta_B' > 0$. Given D_B and D_S, the total volume of card transactions is given by

$$\begin{aligned}V\left(b_B^m, b_S^m\right) &= D_B\left(b_B^m\right) D_S\left(b_S^m\right) \\ &= \left[1-H\left(b_B^m\right)\right]\left[1-G\left(b_S^m\right)\right].\end{aligned}$$

Moreover, the average IF per card transaction can now be explicitly characterized as

$$\begin{aligned}\alpha &= E\left(a\left(b_S\right) \mid b_S \geq b_S^m\right) \\ &= \frac{\int_{b_S^m}^{\bar{b}_S} a\left(b_S\right) dG\left(b_S\right)}{1-G\left(b_S^m\right)} \text{ for } b_S^m < \bar{b}_S.\end{aligned} \tag{1}$$

The timing of the game is as follows. First, the regulator determines whether or not to impose a restriction on price differentiation. Second, the payment network chooses $a\left(b_S\right)$ subject to the regulation. Third, competing issuers and acquirers set p_B and p_S. Fourth, given their individual realizations of b_B and b_S, consumers and merchants decide whether to use (respectively accept) cards. The equilibrium of this game is described in the following proposition.[10]

Proposition 1 *If the payment network is permitted to differentiate IFs around $\bar{a}$, then it will do so by increasing (decreasing) the IF for merchants with high (low) transactional benefits of card use. The resulting IF differentiation (i) increases the degree of card acceptance D_S among*

[10] As shown by Rochet and Tirole (2011), the payment scheme's maximization problem is well-defined if the distribution of b_B fulfills the monotone hazard rate property.

merchants, (ii) leaves consumers' demand for card usage D_B unaffected, and (iii) maximizes the volume of card transactions V given the regulatory cap on the average level of the IF.

Proof. If the payment network can freely choose to differentiate IFs around the regulatory average $\bar{a}$, it will do so in order to maximize profits

$$\Pi\left(b_B^m, b_S^m\right) = m\left(c_B - \alpha\right) V\left(b_B^m, b_S^m\right). \tag{2}$$

Since a differentiation of sector-specific IFs around $\bar{a}$ does not affect the average interchange fee α, and since m is only contingent on the level of α, but not on individual fees, it follows from (2) that the network chooses profit maximizing differentiation in such a way as to maximize output V. This proves (iii). Since p_B is only contingent on α, the marginal consumer b_B^m is not affected by changes in the structure of the IF around $\alpha = \bar{a}$. This proves (ii). Maximization of V therefore amounts to maximizing merchant acceptance $D_S\left(b_S^m\right)$. This is done by constructing the differentiated scheme $a\left(b_S\right)$ in such a way as to decrease b_S^m as much as possible while keeping $\alpha = \bar{a}$ constant, which proves (i). ■

Proposition 1 provides a positive analysis of IF differentiation. Before turning to the normative assessment, the following definition is introduced.

Condition 1 *Card payments are called "technologically efficient" if*

$$\beta_B\left(b_B^m\right) + b_S^m \geq c_B + c_S. \tag{3}$$

If (3) holds, then (from the perspective of social welfare) the *average* card transaction at merchant outlets that accept cards is efficient.[11] Guthrie and Wright (2007), who model the card acceptance decision of merchants in a strategic environment, show that merchants accept cards *if and only if Condition 1 holds.* Their theoretical result is consistent with the empirical evidence. In particular, both regulators and academic researchers have repeatedly argued that a move towards a greater use of electronic payments would have large benefits to society, among other things because the marginal cost of making payments in a card network are very small, while cash payments involve significant transaction costs.[12] In that case, the following strong result on differentiation can be derived.

Proposition 2 *Suppose card payments are technologically efficient in the sense of Condition 1. Then permitting the payment network to differentiate IFs (i) always increases social welfare and (ii) the resulting differentiation is such that social welfare is maximized given the regulatory cap* $\bar{a}$ *on the average IF.*

Proof. Overall welfare in the model is given by

$$W\left(b_B^m, b_S^m\right) = \left[\beta_B\left(b_B^m\right) + \beta_S\left(b_S^m\right) - c_B - c_S\right] V\left(b_B^m, b_S^m\right).$$

Since differentiation decreases b_S^m but leaves b_B^m unaffected, we consider

$$\begin{aligned} \frac{dW}{db_S^m} &= \frac{d\beta_S}{db_S^m} V\left(b_B^m, b_S^m\right) + \left[\beta_B\left(b_B^m\right) + \beta_S\left(b_S^m\right) - c_B - c_S\right] \frac{dV}{db_S^m} \qquad (4) \\ &= \frac{g\left(b_S^m\right)}{1 - G\left(b_S^m\right)} \left(\beta_S - b_S^m\right) D_B\left(b_B^m\right) D_S\left(b_S^m\right) + \left[\beta_B\left(b_B^m\right) + \beta_S\left(b_S^m\right) - c_B - c_S\right] \left[-g\left(b_S^m\right) D_B\left(b_B^m\right)\right] \\ &= -g\left(b_S^m\right) D_B\left(b_S^m\right) \left[\beta_B\left(b_B^m\right) + b_S^m - c_B - c_S\right] \end{aligned}$$

[11] Note that this does not imply that each card transaction that occurs must be socially efficient. To the contrary, there may well be card transactions occurring for which it would have been socially more efficient to use cash—as long as at least the average card transaction is efficient.

[12] For empirical evidence on the greater efficiency of card payments, see Garcia-Swartz et al. (2006) and the references cited therein.

As a matter of general principle, (4) is indeterminate in sign. However, (3) implies that $dW/db_S^m \leq 0$ in (4). Hence, any decrease in b_S^m induced by larger differentiation increases welfare. This proves (i). (ii) follows from the fact that, by Proposition 1, the payment network will differentiate fees in such a way as to maximize the volume of card transactions. Hence, the payment network will try to reduce the cutoff point b_S^m as much as possible given the cap $\bar{a}$, which by the fact that $dW/db_S^m \leq 0$ maximizes welfare. ■

Note that Propositions 1 and 2 imply some separability between determining the extent of card use on the consumer side (which can be steered by by altering $\bar{a}$) and influencing the extent of card acceptance on the merchant side (which is maximized by allowing differentiation). In conclusion, even in scenarios where regulating the *level* of IFs turns is welfare-enhancing, it is generally more efficient to leave the determination of the *structure* of IFs to payment card associations themselves.

References

[1] Armstrong, M.,Vickers, J., 1991. Welfare effects of price discrimination by a regulated monopolist. RAND Journal of Economics 22, 571-580.

[2] Baxter, W. F., 1983. Bank Interchange of Transactional Paper: Legal and Economic Perspectives. Journal of Law and Economics 26, 541-588.

[3] Garcia-Swartz, D.D., Hahn, R.W., Layne-Farrar, A., 2006. The Move Toward a Cashless Society: A Closer Look at Payment Instrument Economics. Review of Network Economic, 5, 175-198.

[4] Gerdes, G., 2008. Recent Payment Trends in the United States. Federal Reserve Bulletin 94, A75-A106.

[5] Guthrie, G., Wright, J., 2007. Competing Payment Schemes. Journal of Industrial Economics 55, 37-67.

[6] Rochet, J.-C., Tirole, J., 2002. Cooperation among competitors: some economics of payment card associations. RAND Journal of Economics 33, 549-570.

[7] Rochet, J.-C., Tirole, J., 2011. Must-Take Cards: Merchant Discounts and Avoided Costs. Journal of the European Economic Association 9, 462-495.

[8] Schmalensee, R., 2002. Payment Systems and Interchange Fees. Journal of Industrial Economics 50, 103-122.

[9] Shy, O., Wang, Z., 2011. Why Do Card Networks Charge Proportional Fees?. American Economic Review 101, 1575-1590.

[10] Wright, J., 2003. Pricing in debit and credit card schemes. Economics Letters 80, 305-309.

[11] Wright, J., 2004. The Determinants of Optimal Interchange Fees in Payment Systems. Journal of Industrial Economics 52, 1-26.

[12] Zenger, H., 2011. Perfect Surcharging and the Tourist Test Interchange Fee. Journal of Banking & Finance 35, 2544-2546.