Note on the Economic Theory of Interchange

By

Emilio Calvano*
Bocconi University, Department of Economics and IGIER

February 22, 2011

Contact details

Calvano: emilio.calvano@unibocconi.it

***** Professor of Economics, Department of Economics, Bocconi University and IGIER. This paper was prepared on behalf of Market Platform Dynamics which received funding from large members of the Electronic Payments Coalition. The views expressed in this paper are my own and do not necessarily reflect those of any of these institutions. I appreciate comments from Richard Schmalensee on this paper and thank Scott Walster of Market Platform Dynamics for research assistance.

Table of Contents

I. Introduction and Overview

Interchange charges within the payment card industry are transfers set by card networks to cope with a basic feature of two-sided markets. Consumers when deciding to have and use cards and merchants when deciding to accept them do not account for the externality that their choices exert on their current and future trading partners. Without an explicit mechanism for allowing agents to internalize these externalities (for instance by bargaining over the choice of the means of payment) interchange transfers can and should be instituted to enhance market efficiency.[1]

In Section II, I provide a brief survey of the economic literature on market-set and socially efficient interchange fees and on the relation between them, supplementing and updating the survey by Prager et al. (2009) at the Board of Governors of the Federal Reserve System ("Board"). This literature is almost entirely theoretical, and the models it contains cannot support any particular regulatory outcome until they are used as frameworks to construct empirically-based models of real markets. In addition, this literature does not explicitly consider the implications of debit cards' essential links with depository accounts. Nonetheless, we believe the insights from this literature have some important general implications for the proper regulation of debit card interchange.

Abstract, theoretical models can serve to make clear what empirical questions need to be answered in order to devise regulations that serve the public interest. And here the literature is clear and unanimous: information about cost is not enough. Interchange fees serve to determine how costs are shared between acquirers and merchants on the one hand and issuers

[1] See Prager et al. (2009) for a comprehensive introduction to the economic theory of interchange.

and consumers on the other; knowing only the costs provides no information on how they should be shared. Efficient regulation cannot be promulgated without knowledge of the demand functions and behavior of merchants and consumers.

In Section III, I consider the theoretical model of Bedre-Defolie and Calvano (2010) ("BC (2010)") in more detail. This model emphasizes a feature of the market that was emphasized by Dr. Prager in her discussion of market failure in response to Chairman Bernanke at the Board's open meeting to discuss the proposed interchange rules:[2] merchants can only decide which payment methods to accept; consumers get to decide which of the accepted methods to use for any particular transaction. In the BC (2010) model, the market-set interchange fee is generally above the socially efficient fee. The calibration exercise reported below suggests that the gap is small in this case, certainly much smaller than the fee reduction proposed by the Board.

II. A Brief Survey

A first set of insights on the economics of payment systems comes from a rapidly expanding literature on the general features of what have come to be called "two-sided markets." (Notable papers include Schmalensee (2002), Caillaud and Jullien (2003), Rochet and Tirole (2003, 2006), Armstrong (2006), Rysman (2009), and Weyl (2010)). These papers study the incentives of privately owned platforms, such as card systems, that enable interactions between two (or more) different sets of users (e.g. cardholders and merchants). These papers contrast market outcomes with efficient (i.e. welfare-maximizing) outcomes under various market configurations and governance structures.

[2] "Federal Reserve Board of Governors Holds an Open Meeting," *CQ Financial Transcripts*, December 16, 2010, at pp. 8-9 of 28.

On the public policy side a key insight is the distinction between the choice of the price level and the price structure (first introduced by Rochet and Tirole (2003)). If they have any market power, platforms, like ordinary one-sided businesses, will choose a price *level* that is too high; they will set a total price per transaction above the corresponding marginal cost. On the other hand profit-maximizing businesses generally do not exhibit any obvious price *structure* distortion. The key intuition is that structuring prices in a way that increases the value delivered to end-users allows two-sided platforms to charge higher per-transaction markups.

In line with this intuition, all the authors cited above conclude that there is no general bias in the structure of fees charged by two-sided platforms. In the case of payment systems, these papers imply that interchange fees may be inefficiently low as well as inefficiently high. Despite shedding a great deal of light on the workings of two-sided businesses, these analyses deliver no straightforward policy implications. The relationship between the socially and privately optimal interchange fees is found to depend on quantitative considerations: the surplus measures hinge on fundamental attributes of cost and demand functions. Socially desirable interventions are shown to require a great deal of information and could in principle either increase or decrease market-set interchange fees, as Prager et al. (2009) recognize. All these papers imply and sometimes say explicitly that there is no economic basis for cost-based regulation. That is, there is no reason to believe that cost-based regulation would improve social welfare relative to market-set interchange fees.

The more specialized literature on payment cards builds on the above framework and studies those aspects inherent to the payment card business that could potentially lead to different conclusions. Most papers focus on the determinants of merchants' willingness to pay to accept particular card brands. A common theme is that merchants are willing to accept cards

because they internalize the value that they could deliver to consumers through acceptance. Card networks, the argument goes, can exploit this internalization effect to induce individual merchants to accept card payments that merchants, collectively, do not want. The presence of merchants who are captive in this sense tilts the price structure and thus the interchange fee to the benefit of buyers in these models. (See the "must take cards" discussion of Vickers (2005).)

Internalization arises in this literature in several ways. Rochet and Tirole (2003) and Wright (2004, 2010) present models in which internalization is due to a "business stealing" effect. That is, merchants accept cards for strategic reasons, seeking an advantage over their competitors. Farrell (2006) and Rochet and Tirole (2010) assume that by accepting cards, merchants could increase the quality of the service they provide and thus appropriate some (or all) of the buyers' transactional benefits by raising prices. Chakravorti and To (2007) instead focus on the ability of credit cards to induce illiquid customers to purchase more.

A key observation is that these factors decrease merchants' resistance to price increases and thus induce higher interchange fees and merchant discounts. However, that conclusion does not necessarily imply that profit maximizing fees are above the socially efficient level. In fact unless one considers the special case in which banks can fully extract all of the merchants' benefits from accepting cards, there is no unambiguous bias in the interchange fee. Another important observation is that since merchants' incentives are also driven by factors that lie outside the payment arena, there is even less scope to use cost as the basis for regulation. Efficient fees would also depend, for instance, on the degree of retail competition or on the elasticity of demand for goods and services bundled with the transaction. This adds a further cautionary note against determining interchange fees by cost-based regulation.

Different considerations have arisen but similar conclusions have been obtained in theoretical studies of competition among payment systems: see Rochet and Tirole (2003), Guthrie and Wright (2007), and Chakravorti and Roson (2006). These papers show that the effect of competition on privately set interchange fees depends on which side of the market represents the "bottleneck." If consumers generally use a single card while merchants accept multiple cards (a pattern consistent with the empirical evidence presented by Rysman (2007)), then competition amongst networks will tend to increase interchange charges as networks try to woo cardholders (through issuers) from their rivals. The two key observations here are: first, whether competition increases or reduces interchange fees is an empirical matter and, second, in any case the effect of competition is not related either to the existence or to the magnitude of a price structure distortion via interchange. This in contrast with one-sided markets, in which competition typically reduces an upward price level distortion and hence increases welfare.

Two very recent contributions (Wright (2010) and BC (2010)) present models in which the privately set interchange fee is generally above the socially efficient level and thus biased against merchants and in favor of cardholders. Wright (2010) argues that if retailers perfectly internalize the benefit that cardholders receive from the ability to use cards, then interchange fees will be biased against retailers. The analysis of BC (2010) turns on the fact that the choice of the payment instrument at the point of sale is ultimately made by cardholders. Intuitively, lowering usage fees on the cardholder side through higher merchant discounts pays a double dividend: higher interchange not only increases issuance of new cards (higher membership) but also fosters usage of existing cards (higher volume per cardholder).

As with all of the papers referenced here, these two contributions do not prove that real debit card interchange fees are set too high, since these models are based on a variety of

simplifying assumptions. It is an empirical question whether market interchange fees lead to distortions, let alone distortions significant enough to justify price regulation. Moreover, as noted above, these theoretical contributions cannot say anything whatsoever about the magnitude of the difference between the socially efficient and market-established interchange fees. As we now show, a calibration exercise with the BC (2010) framework establishes at least a presumption that the market-established interchange fee is not much higher than the socially efficient interchange fee.

III. The BC (2010) Model

In this section, I first derive the determinants of the efficient interchange fee, denoted i, in a general "two-sided market" model. I then adopt the BC (2010) framework to derive the profit maximizing policy and contrast it with the socially efficient one.[3] I present a calibration exercise that strongly suggests that the Board's proposed regulation would result in interchange fees below the socially efficient level.

Suppose that there is a single network jointly owned by a finite number of issuers and acquirers.[4] Let $m-(c_A+i)$ and $f-(c_I-i)$ be the per transaction markups at per-transaction prices m and f charged by issuers and acquirers respectively, with marginal costs $c_I, c_A \geq 0$ and interchange fee i. Cardholders use a card if only if the benefits of doing so exceed the fee for that transaction. Let $D_B(f)$ be the fraction of transactions that the average cardholder

[3] Although the Board has not consistently embraced any explicit model or framework, this is arguably the closest with respect to the view expressed by Dr. Prager during the Board's open meeting (see footnote 2, above) that consumer choice of means of payment is a source of inefficiently high interchange fees. In addition, this setup is very broad and nests several prior contributions. The paper distinguishes usage and membership choices (a broader action space) and allows for four degrees of heterogeneity (a broader parameter space).

[4] The assumption that the issuers and acquirers own the network is convenient, and nothing I say below would change in the case, as is now true, that the acquirers and issuers are customers of an independently owned network.

makes at price f. Similarly let $D_S(m)$ be the fraction of merchants that accept cards at price m. Finally let N be the number of cardholders (i.e. the number of consumers who subscribe to the network).

The socially optimal interchange fee maximizes the sum of end user surplus. Perhaps the most intuitive way of expressing this measure is as the sum of the net (per transaction) average benefit of cardholders and merchants times the number of transactions in the economy (arguments omitted):

$$(1) \qquad i^* = \arg\max_i \left(v_B + v_S\right) D_B D_S N$$

Suppose N is fixed, interchange charges are fully passed through to cardholders and merchants (i.e. $m'(i) = -f'(i) = 1$) and there are no fixed costs. In this case the socially optimal interchange fee has a relatively simple form:

$$(2) \qquad \frac{f(i^*)}{m(i^*)} = \frac{\eta_B^*}{\eta_S^*} \div \frac{v_B^*}{v_S^*}$$

where η_B^* denotes the elasticity of the consumers' demand computed at the socially efficient usage fee and η_S^* the elasticity of the merchants' demand for membership with respect to the merchant discount again computed at the efficient price. In words, the socially optimal allocation is achieved when relative user prices are equal to the ratio of the relative demand elasticities and the relative average per transaction surpluses of buyers and sellers.

An important observation is that issuing and acquiring costs do not enter directly into the equation. In other words given a total price $f + m$, which fraction of this price has to be paid from one set of users does not depend either on the allocation of issuing and acquiring

marginal costs or on their level. This is true even if the policy-maker could control end user prices and insist that banks price at cost; that is, $f(a) = c_I - a$ and $m(a) = c_A + a$ so that $f + m = c_I + c_A$. This is the two-sided analog to the "price at marginal cost" principle for one-sided markets. Both principles must, of course, be modified if revenues must exceed variable costs in order to cover fixed costs.

Consider now the network's profit-maximizing pricing policy. BC (2010) enrich the above setup by explicitly modeling card membership choices (i.e., N is endogenous) and allowing issuers to charge membership fees. (In our setting one could think of the membership fee as subsumed in the various terms and conditions of a depository account relationship.) When issuers have market power but acquirers do not (a standard assumption in this literature), the network chooses the pricing policy that maximizes the expected buyers' surplus from usage (arguments omitted, as above):

$$(3) \qquad i^M = \arg\max_i \; v_B D_B D_S N = \eta_S v_B - c_A$$

This argument relies on issuers' internalization of cardholder surplus from card usage through membership fees. In reality banks' can not fully discriminate between different types of cardholders, hence i^M in what follows should be interpreted as an upper bound on the fee that the network would choose.[5]

Given the above, the determinants of the upper-bound for the gap between the socially efficient and privately optimal interchange fee are summarized by the following expression:

[5] BC (2010) show that under log-concavity of demand functions, a standard regularity condition, the fee that maximizes merchants' surplus, i^S, is well-defined and higher than the fee that maximizes buyers' surplus.

$$(4) \qquad i^M - i^* = \eta_S^M v_B^M - \frac{\eta_S^* v_B^*}{\eta_B^* v_S^*}(c_I + i^*)$$

In order to better understand the underlying principles and the order of magnitudes involved, we now specialize to the simplest case with linear demand functions to permit calibration. That is, we assume that $D_B(f) := l_B - k_B f$ and $D_S(m) := l_S - k_S m$. This implies:

$$(5) \qquad i^M = i^* + \frac{1}{12}(v_B + v_S)$$

The advantage of this expression is that we can readily interpret it in terms of average surpluses for a typical $38.58 debit card transaction. Specifically, in this special case the gap is proportional to the sum of the average per-transaction benefits of buyers and merchants on an typical debit card transaction. This is the amount, computed at the profit maximizing fees, that buyers and merchants would have to be paid to switch from using a debit card to the next most preferred means of payment.

In this framework, if the market-set fee is $0.44, a fee of $0.12 would be socially optimal only if the average benefit per transaction were $3.84:

$$(6) \qquad (44 - 12)12 = v_B + v_S$$

As discussed in more detail in Evans, Litan and Schmalensee (2011), this number, almost 10% of the value of a typical transaction, is implausibly high. Merchants' per-transactions costs are measured in dimes, not dollars. And, with ATMs widely deployed and credit cards in the hands of most consumers, it seems implausible that a typical consumer

---

Privately as well as socially optimal interchange fees always lie in the interval $[i^S, i^M]$, so if there is not full internalization, the privately optimal fee is lower than i^M.

would suffer more than a few pennies worth of inconvenience on a transaction of this size if they if had to use cash or credit instead of debit. But if, as seems plausible, the average total surplus on a typical debit transaction is well below $3.84, this analysis implies that the Board has set an interchange fee cap well below the socially efficient level. While I cannot assert that there is empirical support for this particular functional form assumption, I know of no reason why this assumption would over-state the average total benefit necessary to rationalize the Board's proposal. It could, in fact, understate that amount.

In the above setup, a necessary premise for the market-set interchange fee to be above the socially efficient level is the exclusion from the system of those merchants whose participation would be valuable to the system as a whole but who chose not to participate at profit maximizing prices. This occurs because different merchants have different valuations for joining the card system that depend on the kind and frequency of transactions that they would expect to process if they joined. However merchants' transaction-specific benefits differ systematically by size and type of store. In order to capture these differences, networks set different fees for different types of store (e.g., service stations vs. supermarkets) and different sizes of storerent sectors of the same industry (convenience store vs. supermarket). MasterCard's published rate schedule, for instance, lists 41 different debit card interchange fees.[6] Within merchant categories there are also individually negotiated rates for the larger merchants that account for much of the volume.

The observed third-degree price discrimination vis-à-vis merchants helps the networks to capture more of the surplus that merchants would otherwise enjoy from accepting cards. In

[6]See Worldwide U.S. and Interregional Interchange Rates," MasterCard, 2010, available at http://www.mastercard.com/us/merchant/pdf/MasterCard_Interchange_Rates_and_Criteria.pdf, at pp. 74-89.

the limit, if most of the variation between merchants' benefits were explained by the sector in which these merchants operate, then price discrimination could allow banks to fully internalize the merchants' surplus. In fact several papers use these observations to support an assumption of homogeneous merchants.[7]

An important consequence of price discrimination is that it reduces the scope for efficient interchange fee regulation. That is, the lower the degree of heterogeneity within merchants the lower the absolute level of the gap between profit-maximizing and welfare-maximizing interchange fees. In the limit, BC (2010, Proposition 5) show that if merchants are homogenous (in general or within categories for which different fees can be charged), the socially optimal and market-set interchange fees are identical. Only empirical analysis can establish how close real debit card systems are to this limit, of course.

IV. Conclusions

In this note I have shown that there is no support in the economic literature for the proposition that welfare-enhancing regulation can be based only on cost data, a proposition the Board's proposal implicitly accepts. I have also shown that the existing theoretical literature indicates some of what must be known about consumer and merchant demands in order to give regulation a sound basis.[8] Finally, the calibration exercises discussed in the preceding section, while not a substitute for serious empirical research and analysis do suggest that the Board's proposed interchange fee cap would likely reduce debit card interchange fees below the socially efficient level.

[7] See, among others, Rochet and Tirole (2003, 2010), Chakravorti and To (2007), and Wang (2010).

[8] It is worth noting again that the existing literature does not deal with debit cards as part of the bundle of services provided with depository accounts. Theoretical work that focuses on the unique features of debit cards may point to other empirical questions that should be addressed.

Technical Appendix

Subscripts "B," "S," stand for Buyers and Merchants (sellers) respectively. m and f denote the merchant discount and the buyer usage charge respectively. c_I is the constant marginal cost of issuing. c_A is the constant marginal cost of acquiring.

$v_B(f) = \int_f^\infty D_B(b)db / D_B(f)$ is the per transaction, net cardholder usage surplus.

$v_S(m) = \int_m^\infty D_S(b)db / D_S$ is the average per transaction net card usage surplus.

Note that $\partial v_B D_B / \partial f = -D_B$ and $\partial v_S D_S / \partial m = -D_S$. Rearranging the first order condition of problem (1) give us characterization (2). The cardholder's surplus maximizing fee solves $i^M = \arg\max_a v_B D_B D_S N$.

Suppose that $D_B(f) = l_B - k_B f$ and $D_S(m) = l_S - k_S f$ for some arbitrary $l_B, l_S, k_B, k_S \geq 0$. Then we have: $\eta_i(x) = \left(\frac{l_i}{k_i} - x\right)^{-1} x$. and $v_i(x) = \frac{1}{2}\left(\frac{l_i}{k_i} - x\right)$.

Plugging these in the above characterizations we get:

$i^* = \frac{1}{2}\left(\frac{l_S}{k_S} - \frac{l_B}{k_B} + c_I - c_A\right)$ and $i^M = \frac{1}{3}\left(\frac{2l_S}{k_S} - \frac{l_B}{k_B} + c_I - 2c_A\right)$. Collecting yields equation (5).

References

Armstrong, Mark (2006), "Competition in Two-Sided Markets," The RAND Journal of Economics, 37:3, pp. 668-691.

Bedre-Defolie, Özlem and Emilio Calvano (2010), "Pricing Payment Cards," *European Central Bank WP series (no 1139).* Latest version available at *http://dl.dropbox.com/u/123685/Website/ppc.pdf*

Caillaud, Bernard and Bruno Jullien (2003), "Chicken & Egg: Competition among Intermediation Service Providers," *The RAND Journal of Economics*, 34:2, pp. 309-328.

Chakravorti, Sujit and Ted To (2007), "A Theory of Credit Cards," *International Journal of Industrial Organization,* 25, pp. 583-595.

Evans, David S., Robert E. Litan, and Richard Schmalensee, "The Economic Principles for Establishing Reasonable Regulation of Debit-Card Interchange Fees that Could Improve Consumer Welfare," Submission to the Board of Governors of the Federal Reserve System, February 22, 2011.

Farrell, Joseph (2006), "Efficiency and Competition between Payment Instruments," *Review of Network Economics,* 5:1, pp. 26-44.

Guthrie, Graeme and Julian Wright (2007), "Competing Payment Schemes," *The Journal of Industrial Economics,* LV (1), pp. 37-67.

Prager, Robin A., Mark D. Manuszak, Elizabeth K. Kiser, and Ron Borzekowski (2009), "Interchange Fees and Payment Card Networks: Economics, Industry Developments, and Policy Issues," *Finance and Economics Discussion Series*, Board of Governors of the Federal Reserve System, 2009-03.

Rochet, Jean-Charles and Jean Tirole (2002), "Cooperation among Competitors: Some Economics of Payment Card Associations," *The RAND Journal of Economics,* 33:4, pp. 549-570.

Rochet, Jean-Charles and Jean Tirole (2003), "Platform Competition in Two-sided Markets," *Journal of the European Economic Association,* 1:4, pp. 990-1029.

Rochet, Jean-Charles and Jean Tirole (2006), "Two-Sided Markets: A Progress Report," *The RAND Journal of Economics,* 37:3, pp. 645-667.

Rochet, Jean-Charles and Jean Tirole (2010), "Must-Take Cards: Merchant Discounts and Avoided Cost," *Journal of the European Economic Association*, forthcoming.

Rysman, Marc (2007), "An Empirical Analysis of Payment Card Usage," *The Journal of Industrial Economics*, LV (1), pp. 1-36.

Rysman, Marc (2009), "The Economics of Two-Sided Markets," *Journal of Economic Perspectives,* 23:3, pp. 125-143.

Schmalensee, Richard (2002), "Payment Systems and Interchange Fees," *The Journal of Industrial Economics*, 50:2, pp. 103-122.

Vickers, John (2005), "Public Policy and the Invisible Price: Competition Law, Regulation, and the Interchange Fee," *Proceedings – Payments System Research Conferences*, May, pp. 231-247.

Wang, Zhu (2010), "Market Structure and Payment Card Pricing: What Drives the Interchange?," *International Journal of Industrial Organization*, Elsevier, 28:1, pp. 86-98.

Weyl, E. Glen (2010), "A Price Theory of Multi-Sided Platforms," *American Economic Review,* 100:4, pp. 1642-72.

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

Wright, Julian (2010), "Why Do Merchants Accept Payment Cards?," *Review of Network Economics*, 9:3.