An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems

JEAN-CHARLES ROCHET*

IDEI, Toulouse University, and LSE.

JEAN TIROLE

IDEI, Toulouse University, and MIT

Abstract

The paper investigates, in a non-technical fashion, the economic determinants of interchange fees in payment card systems and the potential need for their regulation. Among other things, it demonstrates that the proposal for a cost-based regulation of interchange fees relies on an erroneous, vertically organized, model of the payment card industry.

1   Introduction

Interchange Fees (henceforth IFs) are fees that banks that service merchants (acquirers) pay to banks that issue cards (issuers) with respect to transactions between their respective customers, i.e. merchants and cardholders. Payment card associations (like VISA or MasterCard[1]) set default values for these IFs. These default values apply to transactions within their systems in the absence of specific bilateral agreements. IFs differ across countries and across transaction types. In several regions of the world (European Union, Australia, Israel), the mode of determination of IFs has come under scrutiny by competition authorities, often at the instigation of large retailers' associations. The purpose of this article is to lay out the basic economic principles for choosing an interchange fee, and to examine the case for a public regulation of interchange fees.

In agreement with Katz (2001), we in particular explain why there is no economic rationale for cost-based regulation of IFs. The idea of a cost-based regulation of IFs erroneously relies on a model of a vertically organized market, in which an "upstream unit" supplies an intermediate input to a "downstream unit", which then serves the final consumer. The analogy is based on the idea that the issuer (the upstream unit) supplies a service (cardholder servicing, and transaction guarantee) to the acquirer (the downstream unit), who then handles the merchant. This vertical view ignores the reverse direction:

---

[1] MasterCard has recently changed its corporate structure to a for-profit organization.

Issuers also serve a category of end-users, the cardholders, and thus can in no way be treated as "upstream units".

Thus, payment card networks are fundamentally two-sided markets. The key aspect in such markets, both from a business and a social perspective, is to get both sides on board, by balancing the demands of cardholders and merchants.

This balancing act bears little relationship with accountants' notions of cost allocation. Quite generally, platforms in multi-sided markets charge price structures that are rather unrelated to standard cost allocations. In the same way airlines charge much lower prices for elastic passenger segments (students, week-end stay overs, ...) than for inelastic ones (businessmen, ...), software programs (Adobe, RealPlayer, etc.) distribute readers for free and charge the other side (writers, content providers, etc.). Portals, TV networks and newspapers charge little or nothing to "eyeballs", and tax advertisers; and so forth. These price structures are not socially inefficient; rather they are critical to the exploitation of network externalities.

To perform the balancing act in the context of the payment-card industry, proprietary systems can directly set end-user prices and use the no-discrimination rule (NDR) that prevents merchants from charging different prices for card and cash transactions.[2] In contrast, payment card associations can use only IFs and the NDR as indirect means of bringing both sides of the market "on board". Incidentally, it seems odd for competition-conscious authorities to deprive open-access cooperatives of the ability to use the instruments (IF and NDR) that are necessary to perform the balancing act, and thereby to destroy the level-playing field in their competition with closed-access, for-profit platforms that do resort to an (implicit) IF and to the NDR.

To be certain, the fact that an association's members, whether predominantly issuers or acquirers, have the socially worthwhile objective to bring on board both sides of the market does not mean that the association, whoever controls it, will pick just the socially optimal IF. But this feature is not specific to the payment card industry: no industry ever engenders the socially optimal decisions. There is no reason to believe that airlines select the perfect bundle of routes, frequencies and prices, that patent holders perfectly maximize social welfare in their licensing choices, or that biotechnology start-ups perform the socially optimal amount of R&D.

The standard approach to public intervention in industries involves two steps:

(1)   the theoretical identification of a serious market failure and the validation of its empirical relevance,

(2)   the identification of the least distortionary way of addressing the market failure and a check that the remedy will not be worse than the illness.

For example, the regulation of telecommunications, electricity and railroad industries has traditionally been based on a broad intellectual consensus that certain segments represent natural monopolies and provide their owners with incentives to charge largely inflated, distortionary prices (part 1). Concerning part 2 there has been much debate as to the proper mode of regulation as well as again a broad consensus that regulation itself introduces non-negligible distortions. Yet, most economists feel strongly enough about

---

[2] In some countries, merchants are prohibited to surcharge cards users: this rule is called the no surcharge rule (NSR).

part 1 that they are willing to accept the need for regulatory intervention in those industries, in spite of the concomitant regulatory distortions.

Proponents of a regulation of the IF must first build a theoretical paradigm that gathers broad intellectual consensus and demonstrates a clear market failure, show that the resulting distortions have a clear sign and a sizeable impact on welfare, and propose a form of regulation that is consistent with the underlying theory and is better that non-intervention. So far, no such theoretical paradigm has been achieved. On the contrary, recent academic work concurs to establishing that there is no systematic bias in the IFs selected by cooperative networks: there is no reason to think that privately optimal IFs are higher or lower than socially optimal ones. Misunderstanding the economics of the problem and imposing cost-based regulation could impose substantial distortions in the industry.

A cost-based regulation of the IF would be an unfortunate precedent for two-sided markets. The same logic would then imply that advertisers' fees paid to TV networks, newspapers and portals[3] should be regulated on a cost basis so as to stop the subsidization of eyeballs by advertisers; that videogame developers[3] be entitled to regulated royalties and development kits, and to above-cost console pricing; that the Internet[3] should be regulated so as to stop the subsidization of websites by dial-up customers through bill-and-keep; that software reader programs[3] be charged the same price as software writer programs (they cost the same); and the social gatherings should be regulated so as to prevent payments to or free entry for attractive participants (e.g., celebrities) while others pay for entry.[3] We do not think that these implications are intended by the proponents of cost-based IFs.

This article is structured as follows. Section 2 explains the fundamental two-sidedness of payment activities and the need to balance the demands of the two sides of the market (cardholders and merchants). Section 3 studies how payment card associations perform this balancing act by setting the appropriate IFs. Section 4 examines whether there is a case for a regulation of IFs by competition authorities.

2   Payment systems are two-sided markets

This section builds on the observation that payment card networks provide interdependent services to two categories of users: cardholders and merchants. These networks can function effectively only if the demands of the two sides of the market are appropriately balanced. We also show that proprietary systems perform this balancing act by setting cardholder and merchant fees having regard to the elasticities of demand on the two sides of the market. Finally we explain why this justifies charging different prices to different categories of users.

2.1   All payment card systems must balance cardholders' demand and merchants' demand

The fundamental characteristic of payment card systems is that every card transaction necessarily involves *two users*: a cardholder and a merchant. Thus it is appropriate to view payment card systems as providing *interdependent services* to cardholders and merchants.

---

[3] These examples are developed in several mini-case studies in Rochet and Tirole (2003).

Cardholders benefit from their holding a card only if their cards are accepted by a wide range of merchants, and merchants benefit from the card only if a sufficient number of consumers use it. Therefore a payment card network can only function effectively if sufficient numbers of both cardholders and merchants participate in the network. To this "membership externality" must be added a "usage externality". In particular, merchants do not benefit from their patrons' holding a card if the latter use it only sporadically.

Thus it is crucial for payment networks to find an effective method for balancing the prices on the two sides of the market. Payment card systems are not the only case of such two-sided markets. In Rochet and Tirole (2003), we provide a detailed analysis of other examples, such as the software industry, videogames, internet portals, medias, and shopping malls.[4] In all these industries as well, the crucial challenge for the platforms is to get both sides of the market "on board", while making a profit overall.

2.2   How proprietary systems perform this balancing act

In proprietary (three-party) systems (such as American Express in most countries or Discover), the network is the only intermediary in the payment card transaction between a cardholder and a merchant. The flow of funds is described in Figure 1.
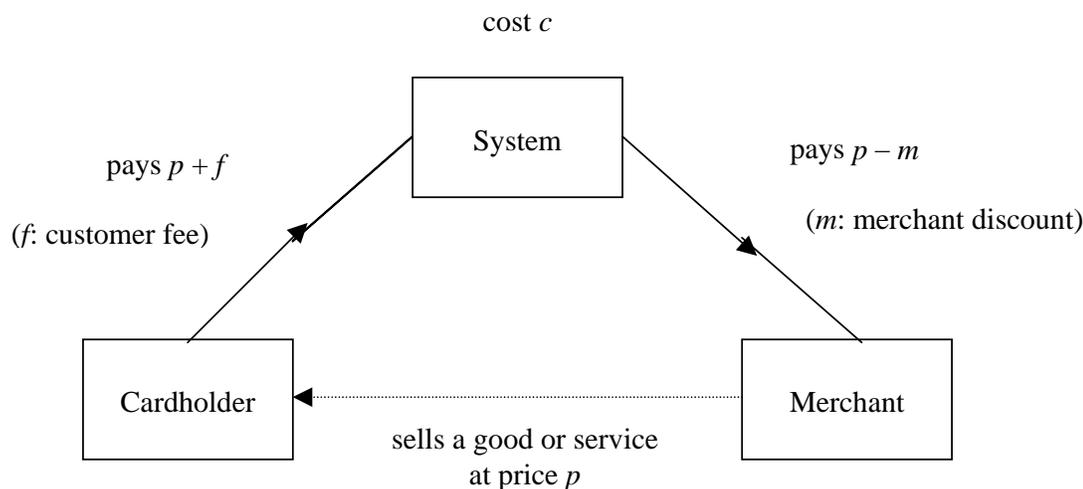
Figure 1: Flow of payments in a card transaction within a proprietary system

The levels of the customer fee f and the merchant discount m are chosen by the managers of the proprietary system so as to maximize its profit. In Rochet-Tirole (2003), we show that this business decision is quite complex. Indeed profit maximizing fees take into account many factors: the system cost c of a transaction (including both issuing and acquiring activities), the intensity of competition with other payment card networks and other payment instruments, and finally the values of demand elasticities of both cardholders and merchants.

---

[4] See also Armstrong (2002) and Evans (2002).

2.3   Is it justified to charge different prices to different users?

The determination of prices as a function of demand elasticities is familiar from other industries. Consider for instance airfares: a flight between any two destinations involves a common cost associated with the operation of the plane. The low prices granted to the elastic side (students, elderly, week-end stay over passengers) and the high fares for inelastic customers (businessmen) allocate the common cost so as to keep "everyone on board", so to speak. Charging different prices to different users, while resulting from profit maximization, is not a distortion due to market power: it is a socially efficient way of recovering the common cost while providing services to a larger number of users than would be otherwise the case.

In the case of two-sided markets, such as payment networks, where there exist positive externalities between different categories of users, it may even happen that one side of the market is left entirely free of charge. For example, cardholders are often exempt of transaction fees. This happens when cardholders are highly resistant to transaction charges.

Another example is the pricing policy of Adobe Inc. in the software industry. Adobe Inc. (and many other software programs) sells Acrobat Exchange, the software needed to transform electronic files into the PDF format. The economic value of this software comes in great part from the fact that any potential reader can download for free the complementary software, needed to read PDF files (Acrobat Reader) on Adobe's web site. This reader is a "damaged" version of Acrobat Exchange. This price structure results from the fact that readers have a lower willingness to pay for the software than writers.

Other examples include portals, free TV networks and low-price newspapers, in which "eyeballs" are highly subsidized by advertisers; and videogames, for which consoles are sold below costs to game players by platforms, which make money by charging game developers.

3   How do payment card associations perform the balancing act?

This section considers payment card associations, i.e. four-party systems, and argues that IFs are the only mechanism through which these associations can perform the balancing act. We then study the determinants of IFs, and finally identify the role of the no-discrimination rule in performing the balancing act.
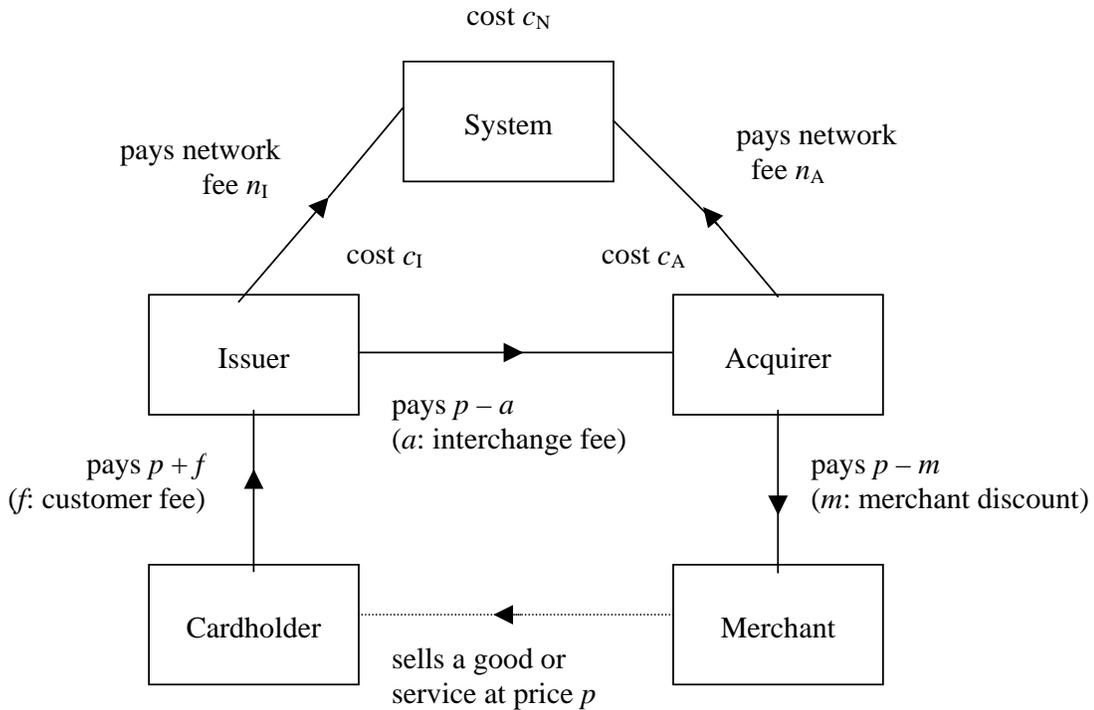
3.1   Interchange fees are the only mechanism through which associations can perform the balancing act

In payment card associations, cardholder and merchant are serviced by two different banks,[5] the issuer and the acquirer. The corresponding flow of payments is described in Figure 2.
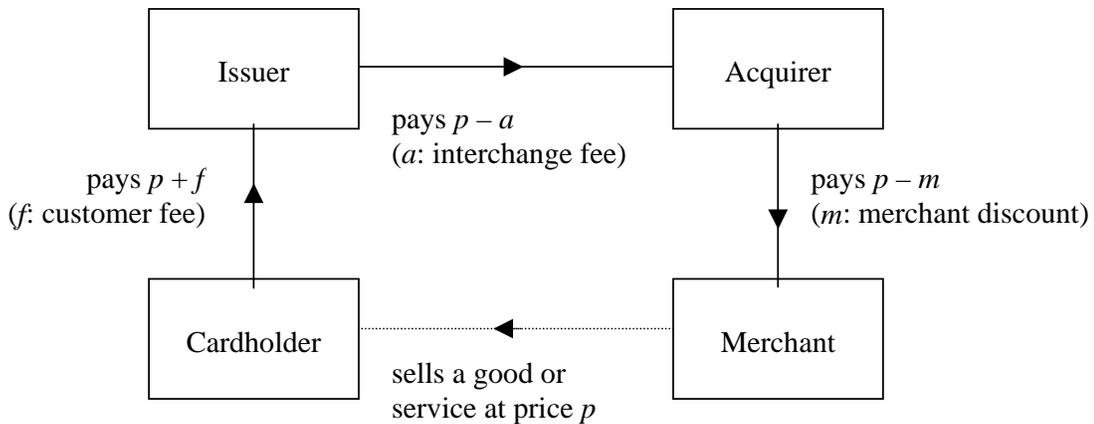
This flow of payments is complex, given that the operation involves five participants. However, since our focus is on IFs, we can simplify the analysis without any loss of relevance, by neglecting network costs c_N and network fees n_I and n_A. We obtain the simplified Figure 3.

---

[5] An exception is "on-us" transactions, where the issuer and the acquirer are the same bank.

cost c_N

System

pays network
fee n_I

cost c_I cost c_A

Issuer ⟶ Acquirer

pays network
fee n_A

pays $p - a$
(a: interchange fee)

pays $p + f$
(f: customer fee)

pays $p - m$
(m: merchant discount)

Cardholder   Merchant

sells a good or
service at price p

Figure 2: Flow of payments in a card transaction within a cooperative system

Issuer ⟶ Acquirer

pays $p - a$
(a: interchange fee)

pays $p + f$
(f: customer fee)

pays $p - m$
(m: merchant discount)

Cardholder   Merchant

sells a good or
service at price p

**Figure 3: Simplified flow of payments in a card transaction within a cooperative
system**

Comparing with Figure 1, Figure 3 shows that the network's influence in the determination of final user prices is only indirect in a cooperative system: it goes through the setting of a, the default value of the IF between member banks. Notice that a does not go to the network, but to the issuer. Also, f and m are not fixed by the network but result from competition among member banks within the network in the issuing and acquiring

markets.[6] By contrast, a proprietary system directly sets *f* and *m* and receives the associated revenues (see Figure 1).

3.2   The determination of interchange fees by cooperative systems

In a cooperative network, the member banks may have different views on the determination of IFs, depending on whether they are net issuers or net acquirers. However, the collective interest of the association is to maximize the total profit of its members. Therefore a cooperative system also has to perform the balancing act between the two sides of the market, and allocate the total cost between them in a proper fashion. However, it can only do so indirectly, through the level of IFs.

In Rochet and Tirole (2002), we analyze the determination of IFs by a cooperative system. By making several restrictive assumptions (monopoly system, fixed total number of transactions, perfect competition among acquirers and – in most of the paper – homogenous merchants), we are able to study the determinants of merchants' resistance and its impact on IFs. Our results have been extended, and some of our restrictive assumptions fruitfully relaxed by several authors: for example, Wright (2000), (2001) introduces heterogeneity also on the merchants' side, while Schwartz and Vincent (2002) introduce some elasticity in the total number of transactions. In Rochet and Tirole (2003), we build a more general model, which allows us to study the impact of platform competition on the determination of IFs. We show that although socially optimal and privately optimal IFs may sometimes differ, there is no systematic bias between them. In specific environments, (linear demands, constant margins) they actually coincide.

One of the main conclusions of this recent literature[7] is that both socially optimal and privately optimal IFs have to take several factors into account: *the split of total costs* between issuer and acquirer, the *demand elasticities* for both types of users, and the *intensity of competition* in both the issuing and the acquiring markets.

3.3   The impact of banning the no discrimination rule

It is often argued that merchants' charging different prices for cash and card transactions is an alternative mechanism to the IF for internalizing externalities between merchants and cardholders. This could justify banning the no-discrimination rule. The NDR is a rule that in some countries prevents the merchants affiliated with payment card networks from charging different prices for customers who pay by cash (or check) and customers who pay by card.

In Rochet and Tirole (2002) we provide conditions under which the IF becomes totally ineffective[8] if all merchants charge different prices for cash and card payments. Indeed, in the model, merchants fully pass-through their net cost for a card payment (merchant discount minus convenience benefit) to the customer who pays by card. If cardholders' fees to issuers are exactly proportional to transactions (in particular there is no yearly fee) and if cardholders are perfectly informed (before selecting a merchant) of card and cash prices charged by competing merchants, the choice of the payment instrument is

---

[6] Note that a change in *a* induces changes in *f* and *m*, although not necessarily one for one, except if downstream markets are perfectly competitive.

[7] See in particular Katz (2001 paragraphs 103 and 117), emphasizing that efficient pricing must be based in part on demand conditions.

[8] Gans and King (2001), (2003) show that this neutrality of IFs when the NDR is banned holds under very weak assumptions.

determined by the sum of the fees paid by cardholders and merchants, which implies that the IF no longer plays a role. Assuming issuing and acquiring markets are perfectly competitive, the externality between merchants and cardholders is perfectly internalized, resulting in an efficient usage of cards, as suggested by Katz (2001).

However, when downstream markets are not fully competitive, the reasoning no longer applies. For example, we show in Rochet and Tirole (2002) that when issuers have market power, banning the NDR would result in a systematic underprovision of card services, while with the NDR in place, the IF chosen by an association can result in an efficient card usage. Similarly, Wright (2000) shows that when merchants have market power and cardholders' payments to issuers are not perfectly proportional to transactions, merchants are able to extract consumer surplus from card usage, destroying incentives for holding payment cards.

Last, the reasoning assumes that price discrimination by merchants is costless; in practice, merchants and end-users incur transaction costs. Casual evidence from the U.K. or some U.S. States, where there is no NDR, shows that in practice such transaction costs prevent a large proportion of merchants from charging differentiated prices for cash and cards. More systematic survey evidence from the Netherlands and Sweden also points at a small amount of surcharging.[9] For all these reasons (as well as the one presented in the next subsection), banning the NDR is no substitute for an IF.

3.4   Does the no-discrimination rule generate distortions in the choice of payment instruments?

It is sometimes argued that cash (and check) users are penalized by the NDR, since undifferentiated retail prices incorporate average transaction costs of merchants (including merchant discounts for card transactions). Thus customers who pay by cash incur a fraction of the costs of card payments, which provides an additional incentive for them to switch to cards. However, this does not imply that IFs selected by card associations distort the choice of payment instruments by consumers, in favour of cards and at the expense of other payment instruments like cash or check. In fact, in most countries, the prices of these other instruments do not fully reflect their costs. In particular, in many countries, check users do not internalize the costs of these checks, because there are no check fees for the vast majority of users. As a consequence, surcharging card users, while check users do not pay for their costs, would certainly not go in the direction of economic efficiency but instead hinder the development of electronic transactions.[10]

4   Is there a case for regulating interchange fees?

In several regions of the world, associations of retailers have repeatedly lobbied competition authorities to regulate IFs. This is easy to understand: retailers have a vested interest in bringing IFs down,[11] since such reductions would be partly reflected in reduced

---

[9] See ITM (2000a,b).

[10] This argument applies to the present situation in most European countries, where the costs of checks are not reflected in their prices for users. A different reasoning would apply to countries like Canada, where check fees are high.

[11] This is only true up to some limits since a low interchange fee reduces the number of cardholders. Interestingly, a decrease in merchant discounts, compensated by an increase in cardholder fees, may be

merchant discounts. Of course, cardholders have a vested interest in high IFs since decreases in IFs would cause cardholder fees to rise. A cardholders' lobby, assuming that it existed, would have some vested interest to push for a regulation that would increase IFs and thereby raise cardholders' benefits.[12]

As a general rule, public policy should be guided by social interest rather than special interests, and should rest on a full-fledged economic analysis of market failures. This section examines the possible market failures that have been advanced, in relation with the use of IFs within payment card associations, and concludes that none of them justifies public intervention as regards IFs.

4.1   The interchange fee is not a fee for service

The proponents of a cost-based regulation of interchange fees implicitly refer to a vertical structure, considering that issuers provide an intermediation service to acquirers, who then supply the final service to merchants. This view is erroneous, simply because it ignores the role of cardholders as consumers of the payment services, on the same footing as merchants. This view fails to recognize the fundamental characteristics of the payment card industry, namely a two-sided market with network externalities, and the resulting need for balancing the demands of the two types of users. Unlike the fee for service in a vertically organized industry, the interchange fee affects not only the marginal cost of merchants, but also the size of the cardholder clientele and the usage of cards.[13]

4.2   Concerns about anti-competitive behaviour in relation with interchange fees

An important question is whether the IF – which is not received by associations but rather by competing members – could be set at the "wrong level". The recent theoretical literature (see in particular Rochet and Tirole, 2002, 2003; Wright, 2001) shows that, although the socially optimal and privately optimal levels of IFs both depend on the same factors (issuers' and acquirers' costs, issuers' and acquirers' margins, cardholders' and merchants' demand elasticities), they are not equal in general. However, given the profession's current state of knowledge, there is no reason to believe that the IFs chosen by an association are systematically too high or too low, as compared with socially optimal levels. That it is not in the interest of an association to choose IFs that deviate markedly from social optima, comes from three factors. First, network externalities imply that weakening the other side of the business reduces the demand from one's own side. Second, competition within networks implies that a reduction in issuers' marginal cost of doing business is partly or fully competed away in favour of cardholders. Finally, competition between networks implies that merchants and/or cardholders can switch providers when one network decides to increase its prices. Let us briefly examine these three factors

Network externalities: Even a monopoly issuer (respectively, a monopoly acquirer) would not benefit from a very high (respectively, very low) IF. A high IF would result in

---

detrimental to the welfare of merchants themselves. Indeed, when the elasticity of cardholders demand is high, even a small increase in cardholders fees may result in a large decrease in cardholders' usage, and thus in the economic value of cards for merchants.

[12] Again, up to a limit, since cardholders want their card to be accepted by merchants.

[13] Moreover it must be kept in mind that the IF is not retained as profit by the association, but rather goes toward lowering the net costs on the side of the system that receives the IF. Ultimately, issuers receiving interchange tend to pass on some of the reduction in net costs in the form of lower cardholder prices.

substantial merchant resistance and would induce many merchants to reject the card. Therefore, even though a very high IF would result in a low marginal cost to the issuer of offering payment services to cardholders, this would do the issuer little good if the cardholders reduce the use of the card because many merchants do not accept it. Symmetrically, even a monopoly acquirer would exercise restraint in setting the IF since a very low IF would lead to a correspondingly high cardholder fee, and so would discourage consumers from holding and using the card. Thus network externalities by themselves induce *some* restraint.

Within Network Competition: An increase in the IF does not go just into the issuers' pockets. It reduces the marginal cost of all issuers, and correspondingly reduces the price charged to cardholders. Indeed, some standard models of competition used in economic theory[14] predict that the increase in the IF is fully passed through to cardholders. While there will not necessarily be full pass-through to cardholders, it is reasonable to assume that much of the increase in the IF is competed away (passed through to consumers). In this case, issuers gain little directly from an increase in the IF and would lose if this increase in the IF induced a substantial fraction of merchants not to take the card. To sum up, the fact that an increase in the IF is partly competed away reinforces the benefits to issuers from exercising restraint.

Between Networks Competition: Competition between networks is an additional force that restrains payment associations from distorting their price structure. However, Rochet and Tirole (2002) and Guthrie and Wright (2003) show that competition between networks does not necessarily lead to welfare improvements with respect to a situation with a monopoly network. The intuition goes as follows: First, competition is usually good at bringing price *levels* down. However, this benefit of competition does not exist when systems are associations whose not-for-profit status prevent exercise of market power. Competition therefore alters the price *structure* (the allocation of cost between the two sides of the market) rather than the price level. As shown in these papers, competition may lead systems to leave a large rent to a side of the market that has a credible threat to sever its relationship with one of the systems. As this rent comes at the expense of the other side of the market, the balancing act is perturbed.

5   Concluding remark

There is no need reiterating the main messages of this paper. Let us only stress the following one. Successful regulation in other industries has built on an intellectual consensus about the existence of a clearly identified and sizeable market failure. The theory of two-sided markets in general, and its application to payment systems in particular, are still in their infancy. Substantial progress has been performed recently, and the existing corpus of knowledge does not bode well for the development of such a consensus. More research is warranted, and caution should be exerted before making sweeping claims about the social improvements that the regulation of IFs will bring about.

---

[14] These include the perfectly competitive model and the Hotelling model of product differentiation.

6   References

Armstrong, M. (2002) "Competition in Two Sided Markets," presentation at ESEM meeting, Venice.

Evans, D. (2002) "The Antitrust Economics of Two-Sided markets," mimeo, AEI-Brookings Joint Center.

Gans, J.S. and S.P. King (2001) "Regulating Interchange Fees in Payment Systems," Working Papers, University of Melbourne.

Gans, J.S. and S.P. King (2003) "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis and Policy*, Vol. 3, Article 1.

Guthrie, G. and J. Wright (2003) "Competing Payment Schemes," mimeo, University of Auckland.

ITM (2000a) *Study Regarding the Effects of the Abolition of the Non-discrimination Rule in Sweden for European Commission Competition Directorate General*, Amsterdam.

ITM (2000b) "The Abolition of the Non-discrimination Rule," Amsterdam.

Katz M.L. (2001) "Reform of Credit Card Schemes in Australia," commissioned report for the Reserve Bank of Australia.

Rochet, J.C. and J. Tirole (2002), "Cooperation among Competitors: Some Economics of Payment Card Associations," *RAND Journal of Economics*, 33, 1-22.

Rochet, J.C. and J. Tirole (2003) "Platform Competition in Two-Sided Markets," *Journal of European Economic Association,* forthcoming.

Schwartz, M. and D.R. Vincent (2002) "Same Price, Cash or Credit: Vertical Control by Payment Networks," Georgetown University, Department of Economics Working Paper 02-01.

Wright J. (2000) "An Economic Analysis of a Card Payment Network," mimeo, NECG and University of Auckland.

Wright J. (2001) "The Determinants of Optimal Interchange Fees in Payment Systems," Department of Economics Working Paper No 220, University of Auckland.