

Market Structure and Credit Card Pricing: What Drives the Interchange?

Zhu Wang*

Payments System Research, Federal Reserve Bank of Kansas City

Working Paper 06-04

December 20, 2006

Abstract

This paper presents a model for the credit card industry, where oligopolistic card networks price their products in a complex marketplace with competing payment instruments, rational consumers/merchants, and competitive card issuers/acquirers. The analysis suggests that card networks demand higher interchange fees to maximize card issuers' profits as card payments become more efficient. At equilibrium, consumer rewards and card transaction volume also increase, while consumer surplus and merchant profits may not. The model provides a unified framework to evaluate credit card industry performance and government interventions.

JEL classification: D40; L10; L40

Keywords: Credit card industry; Oligopolistic pricing; Interchange fees

* Email: zhu.wang@kc.frb.org. I thank seminar participants at Federal Reserve Bank of Kansas City for comments and Nathan Halmrast for research assistance. The views expressed herein are solely those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System. Preliminary draft, please do not cite or quote without author's permission.

1 Introduction

1.1 Motivation

Credit and debit cards have become an increasingly prominent form of payments.¹ From 1986 to 2000, the share of US consumer expenditures paid for with cards has increased from about 3 percent to 25 percent. In 1995, credit and debit cards counted less than 20 percent of noncash payments; by 2003, they exceeded 40 percent. According to recent estimates, 92 percent of US households with income over \$30,000 hold at least one credit card, with an average for all households of 6.3 cards.²

With this growth has come increased scrutiny of both the benefits and costs of card use. In particular, the growth of credit card transaction has been paralleled by an accelerated trend of legal battles and lobby for regulations against the credit card networks. At the heart of the controversy is the interchange fees (IFs) - the fees that merchant-acquiring banks (acquirers) pay to card-issuing banks (issuers) with respect to transactions between their respective customers, i.e., merchants and cardholders. It is estimated that the annual amount paid for interchange fees totals \$30 billion or \$200 per household in the United States.

Interchange fees are typically set by the credit card networks. The two major card networks, Visa and MasterCard, each set its interchange fees collectively for

¹There are four types of general purpose payment cards in the US: (1) credit cards; (2) charge cards; (3) signature debit cards; and (4) PIN debit cards. The first three types of cards are routed over credit card networks. They extend credit to card holders to some extent and generally charge a proportional fee based on transaction volume to merchants who accept them. In this paper, most discussion is in the context of credit cards, but may apply to all these three types of cards.

²Source: Lyon, James (2006). "The Interchange Fee Debate: Issues and Economics", *The Region*, Federal Reserve Bank of Minneapolis.

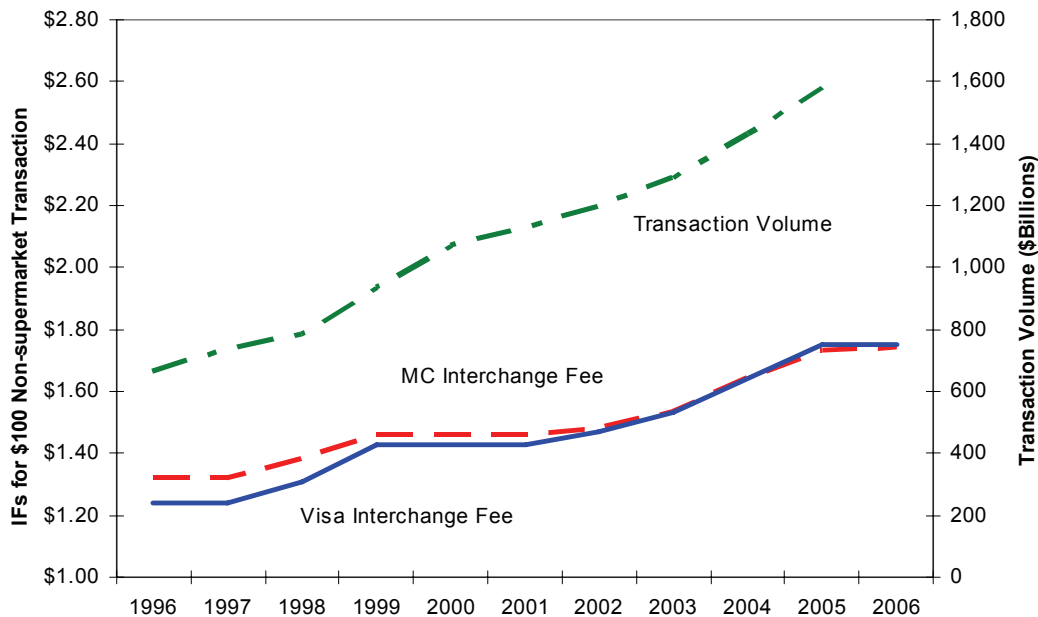


Figure 1: U.S. Credit Card Interchange Fees and Transaction Volume

tens of thousand member financial institutions that issue and market their cards.³ Industry participants tend to agree that some centrally determined interchange fees are necessary since they help eliminate costly bargaining between individual card issuers and acquirers (Baxter 1983). However, they disagree on the actual levels of interchange fees. Particularly, merchants in the US and worldwide are complaining furiously about the increasing interchange burden. Figure 1 shows that in the US,

³Visa and MasterCard provide card services through member financial institutions (card-issuing banks and merchant-acquiring banks). They are called “four-party” systems and count for 80% of the US credit card markets. Amex, Discover and Diner’s Club handle card issuing and acquiring by themselves. They are called “three-party” systems and count for the rest 20% of the US credit card markets. In a “three-party” system, interchange fees are internal transfers and hence not directly observable. This paper provides a model in the context of four-party credit card systems, but the analysis can be easily extended to three-party systems.

where no government intervention is introduced, Visa and MasterCard's interchange fees have been rising over the past several years.⁴

Around the world, some competition authorities and central banks have taken action recently (Hayashi 2006, Weiner and Wright 2006). In the UK, the Office of Fair Trading announced in 2005 its intention to regulate down MasterCard's credit card interchange fees as well as investigate Visa's. In the European Union, the European Commission pushed the Visa International to agree to reduce its cross-border interchange fees on credit and debit transactions in 2002. In Australia, the Reserve Bank of Australia mandated a sizeable reduction of credit-card interchange fees in 2003, and is considering doing the same for debit transactions. Other countries, including Israel, Spain, Portugal, Belgium, Mexico and the Netherlands, have made similar decisions and actions. The interchange fees in the US are among the highest in the world. Although the Department of Justice and the Federal Reserve so far have not heavily involved, major legal battles are taking place in the court. In 2005, there were more than 50 pending antitrust cases on interchange fees and have been consolidated into a single case. Meanwhile, the issue was the focus of a Congressional Subcommittee hearing.

The performance of the credit card industry raises many challenging research questions, for example:

- Why have interchange fees been increasing given falling costs and increased competition in the card industry (card processing, borrowing and fraud costs have all declined, while the number of issuers and card solicitations have been

⁴Data Source: Credit card transaction volume is from *Nilson Report*; interchange fees (IFs) from *American Banker*. Interchange fees for supermarket transactions (not shown in the graph) follow a similar trend.

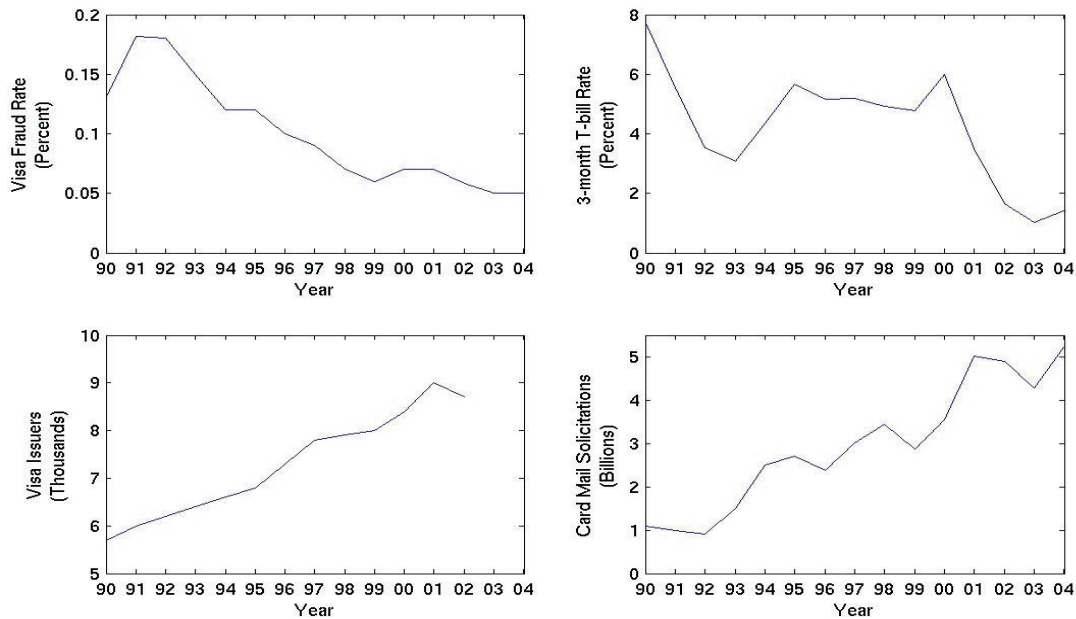


Figure 2: Credit Card Industry Trends: Costs and Competition

rising over recent years, as shown in Figure 2)?⁵

- Given the rising interchange fees, why can't merchants refuse accepting cards?
Why has card transaction volume been growing rapidly?
- What are the causes and consequences of the increasing consumer card rewards?
- What can government intervention do in the credit card industry? Is there a socially optimal card pricing?

A growing literature tries to understand these issues but is far from reaching a consensus. Many studies emphasize the two-sided nature of payment markets and

⁵Data Sources: Visa card fraud rate is from the Visa USA; interest rate (3-month treasury bill rate) from the Federal Reserve Board; number of Visa issuers from Evans and Schmalense (2005); number of card mail solicitations from Frankel (2006).

argue that interchange fees are not an ordinary market price but a balancing device for increasing the value of a payment system by shifting costs between issuers and acquirers and thus shifting charges between consumers and merchants (Schmalensee 2002, Rochet and Tirole 2002). Wright (2004) shows that when merchants compete and consumers are fully informed as to whether merchants accept cards, the profit and welfare maximizing fee coincide for a non-trivial set of cases. In contrast, other studies try to identify potential anti-competitive effects of the collective determination of interchange fees, but most of them lack a formal treatment (Carlton and Frankel 1995, Katz 2001, Frankel 2006).

1.2 A New Approach

This paper presents an industry equilibrium model to analyze the structure and performance of the credit card market. The market that we consider consists of competing payment instruments, e.g., credit cards vs. alternative payment methods;⁶ rational consumers (merchants) that always use (accept) the lowest-cost payment instruments; oligopolistic card networks that set profit-maximizing interchange fees; and competitive card issuers that join the most profitable network and compete with one another via consumer rewards.

Exploring the oligopolistic structure of this market, the model derives equilibrium industry dynamics consistent with empirical facts. It suggests that market power of credit card networks plays a critical role in determining the card pricing. In particular, card networks are likely to collude to set monopoly interchange fees under the constraints that merchants choose to accept cards; consumers choose to use cards;

⁶Competing payment instruments to credit cards may include cash, check, PIN debit cards, stored value cards, ACH and etc.

issuers choose to join the networks; and interchange fees and consumer rewards clear the market. It is found that card networks demand higher interchange fees to maximize card issuers' profits as card payments become more efficient. At equilibrium, consumer rewards and card transaction volume also increase, while consumer surplus and merchant profits may not. The model provides a unified framework to evaluate credit card industry performance and government interventions.

The differences between this model and others are significant. First, this paper offers a very different perspective on the two-sidedness of the credit card market. During the early development of the card industry, adoption and usage externality between merchants and consumers was a remarkable feature (McAndrews and Wang 2006, Rochet and Tirole 2006), but has become less important as the industry gets mature. Therefore, the interactions among participants in a mature card market are modeled in an industry equilibrium model without externality in this paper. Second, most studies in the literature (Rochet and Tirole 2002, Wright 2003, Hayashi 2006) rely on restrictive assumptions: consumers have a fixed demand for goods (irrelevant to their payment choices); merchants engage in a special form of imperfect competition (e.g., Hotelling); and there is no entry/exit of card issuers. Although that framework is handy to consider merchants' business stealing motive for accepting cards, it has ignored critical issues beyond those assumptions. Particularly, the pricing of payment does affect consumers' demand for goods; the entry and exit of card issuers are endogenous; and most important, interchange fees play a key role in network competition for attracting issuers. In contrast, this new model allows elastic demand, competitive merchants, free entry/exit of card issuers and oligopolistic network competition. As a result, it provides a more realistic and arguably better framework to understand the overall picture of the credit card market.

1.3 Road Map

Section 2 models the interactions among card market participants, including consumers, merchants, issuers, acquirers and card networks. The analysis suggests that a monopoly card network demands higher interchange fees to maximize card issuers' profits as card payments become more efficient. At equilibrium, consumer rewards and card transaction volume increase, while consumer surplus and merchant profits may not. We then show these findings are likely to hold under oligopolistic card networks. Section 3 extends the model to evaluate government intervention and discuss socially optimal card pricing. Section 4 concludes.

2 The Model

2.1 Basic Setup

A four-party card system is composed of five players: merchants, consumers, acquires, issuers, and card networks, as illustrated in Figure 3. They are modeled as follows.

Merchants: A continuum of identical merchants sell a homogenous good in the market.⁷ The competition requires zero profit. Let p and k be price and non-payment cost for the good respectively. Merchants have two options to receive payments. Accepting non-card payments, such as cash, costs merchants $\tau_{m,a}$ per dollar, which includes the handling, storage, and safekeeping expenses that merchants have to bear. Accepting card payments costs merchants $\tau_{m,e}$ per dollar plus a merchant discount rate S per dollar paid to merchant acquirers. Therefore, a merchant who does not accept cards (i.e., cash store) charges p_a , while a merchant who accepts cards (i.e.,

⁷Alternatively, we can model heterogenous merchants and get similar results (See Appendix B).

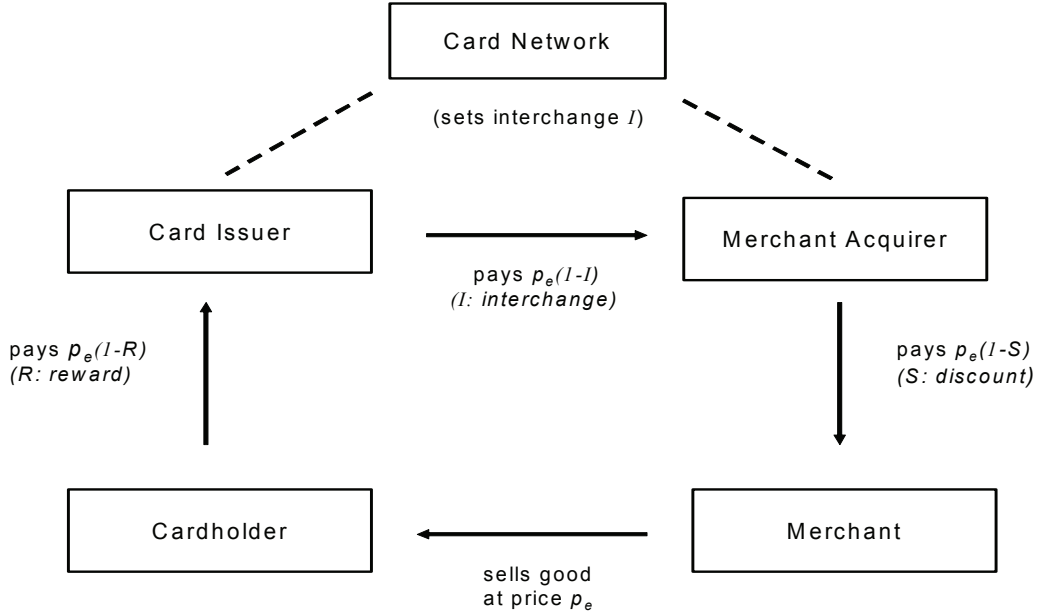


Figure 3: A Four-Party Credit Card System

card store) charges p_e :

$$p_a = \frac{k}{1 - \tau_{m,a}}; \quad p_e = \max\left(\frac{k}{1 - \tau_{m,e} - S}, p_a\right).$$

The pricing of p_e requires $p_a \leq p_e$ so that $(1 - \tau_{m,a})p_e \geq k$, which ensures card stores do not incur losses in case someone use cash for purchase. This condition implies

$$S \geq \tau_{m,a} - \tau_{m,e}; \tag{1}$$

in another word, S has no effect on card store pricing whenever $S < \tau_{m,a} - \tau_{m,e}$. Moreover, $1 - \tau_{m,e} > S$ is required for a meaningful pricing.

Consumers: All consumers have access to cash and most of them also own credit cards. To use each payment instrument, consumers also incur costs on handling, storage and safekeeping. Using cash costs consumers $\tau_{c,a}$ per dollar while using card

costs $\tau_{c,e}$. In addition, consumers receive a reward R from card issuers for each dollar spent on cards. Therefore, card consumers do not shop cash stores if and only if

$$(1 + \tau_{c,a})p_a \geq (1 + \tau_{c,e} - R) p_e \iff \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - S}. \quad (2)$$

Meanwhile, given $p_a \leq p_e$, cash consumers prefer shopping cash stores,⁸ and card consumers have no incentive to ever use cash in card stores.

When making a purchase decision, card consumers face the after-reward price

$$p_r = (1 + \tau_{c,e} - R) \frac{k}{1 - \tau_{m,e} - S},$$

and have the total demand for card transaction volume TD :

$$TD = p_e D(p_r) = \frac{k}{1 - \tau_{m,e} - S} D\left[\frac{k}{1 - \tau_{m,e} - S} (1 + \tau_{c,e} - R)\right],$$

where D is the demand function for goods.

Acquirers: The acquiring market is competitive, where each acquirer receives a merchant discount rate S from merchants and pays an interchange rate I to card issuers.⁹ Acquiring incurs a constant cost C for each dollar of transaction. For simplicity, we normalize $C = 0$ so acquirers play no role in our analysis but pass through the merchant discount as interchange fee to the issuers, i.e., $S = I$.¹⁰

⁸In reality, some consumers are seen using Pin debit cards or cash in stores that accept credit cards. Given the typically lower costs for merchants to accept non-credit-card payments, it is argued that cash users are exploited. In theory, this can happen if cash stores, due to their smaller customer base, may have a higher unit cost k than card stores. However, considering the small size of cash consumers anyway, its effect on card store retail price is negligible.

⁹Although acquirers can differentiate themselves by providing different accounting services, the business is mainly about offering reliable transaction processing services at the lowest possible prices. There is evidently intense competition for merchant accounts, and it is common for merchants to switch among acquires to get the best possible price (Evans and Schmalensee 2005).

¹⁰Alternatively, we may model heterogenous acquirers, but the results will be very similar to our following analysis of issuers.

Issuers: The issuing market is competitive, where each issuer receives an interchange rate I from acquirers and pays a reward rate R to consumers.¹¹ An issuer α incurs a fixed cost K each period and faces an issuing cost V_α^β/α for its volume V_α , where $\beta > 1$. Issuers are heterogenous in their operational efficiency α , which is distributed with pdf $g(\alpha)$ over the population. They also pay the card network a processing fee T per dollar transaction and a share c of their profits.¹²

Issuer α 's profit π_α (before sharing with the network) is determined as follows:

$$\pi_\alpha = \underset{V_\alpha}{Max}(I - R - T)V_\alpha - \frac{V_\alpha^\beta}{\alpha} - K$$

$$\implies V_\alpha = \left(\frac{\alpha}{\beta}(I - R - T)\right)^{\frac{1}{\beta-1}}; \quad \pi_\alpha = \frac{\beta - 1}{\beta} \left(\frac{\alpha}{\beta}\right)^{\frac{1}{\beta-1}} (I - R - T)^{\frac{\beta}{\beta-1}} - K.$$

Free entry condition requires that the marginal issuer α^* breaks even, so we have

$$\pi_{\alpha^*} = 0 \implies \alpha^* = \beta K^{\beta-1} \left(\frac{\beta}{\beta-1}\right)^{\beta-1} (I - R - T)^{-\beta}.$$

¹¹Although card issuers do not offer identical products, the following description of perfect competition matches the issuers' market very well: (1) There is a large number of issuers, e.g., over 8000 issuers for Visa; (2) No single issuer is large relative to the industry, e.g., the HHI for the credit and charge card industry was 816, considered unconcentrated by the DOJ merger guideline; (3) Entry and exit are fairly easy. Visa and MasterCard are open to all financial institution that qualify for FDIC deposit insurance and charge a low membership fee. Member issuers that wish to exit, for whatever reason, can easily sell their portfolios to other members. (4) Information is widely available to consumers through newspapers and internet. Issuers are extremely active in marketing, e.g., approximately 3.9 solicitations per month for each household in the U.S. in 2001; (5) It is easy to switch cards, and consumers do so all the time (Evans and Schmalensee 2005).

¹²In reality, T refers to the Transaction Processing Fees that card networks collect from their members to process each card transaction through its central system, which is typically cost-based; c refers to the Quarterly Service Fees that card networks charge their members, which are calculated based on each member's statistical contribution to the network (such as the number of card issued, total transaction and sales volume and other measures). Source: Visa USA and UBS.

As a result, the total number of issuers is

$$N = \int_{\alpha^*}^{\infty} g(\alpha) d\alpha$$

and the total supply of card transaction volume is

$$TV = \int_{\alpha^*}^{\infty} V_{\alpha} g(\alpha) d\alpha = \int_{\alpha^*}^{\infty} \left[\left(\frac{I - R - T}{\beta} \right) \alpha \right]^{\frac{1}{\beta-1}} g(\alpha) d\alpha.$$

Networks: Each period, a card network incurs a fixed cost E and a variable cost T per dollar transaction to provide the service. In return, it charges its member issuers a processing fee T to cover the variable costs and demands a proportion c of their profits, where c is determined by bargaining between the card network and issuers. As a result, the card network sets the interchange fee I to maximize its profit $\Omega = c \int_{\alpha^*}^{\infty} \pi_{\alpha} g(\alpha) d\alpha - E$, which also maximizes the total profits of its member issuers.

2.2 Monopoly Outcome

Due to scale economies, only a small number of card networks can exist in a market and they enjoy significant market power. In some countries, there is only one monopoly card network. In many others, there are a few oligopolistic networks. However, if oligopolistic networks are able to collude, as we later will discuss, they act as a monopoly. Therefore, we start our analysis with the monopoly case.

A monopoly network, whose profit Ω^m ties closely to its member issuers' profits, solves the following problem each period:

$$\underset{I}{Max} \quad \Omega^m = c \int_{\alpha^*}^{\infty} \pi_{\alpha} g(\alpha) d\alpha - E \quad (\text{Card Network Profit})$$

$$s.t. \quad \pi_{\alpha} = \left(\frac{\beta - 1}{\beta} \right) \left(\frac{\alpha}{\beta} \right)^{\frac{1}{\beta-1}} (I - R - T)^{\frac{\beta}{\beta-1}} - K, \quad (\text{Profit of Issuer } \alpha)$$

$$\alpha^* = \beta K^{\beta-1} \left(\frac{\beta}{\beta-1} \right)^{\beta-1} (I - R - T)^{-\beta}, \quad (\text{Marginal Issuer } \alpha^*)$$

$$N = \int_{\alpha^*}^{\infty} g(\alpha) d\alpha, \quad (\text{Number of Issuers})$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I}, \quad (\text{Pricing Constraint I})$$

$$1 - \tau_{m,e} > I \geq \tau_{m,a} - \tau_{m,e}, \quad (\text{Pricing Constraint II})$$

$$TV = \int_{\alpha^*}^{\infty} V_{\alpha} g(\alpha) d\alpha = \int_{\alpha^*}^{\infty} \left[\left(\frac{I - R - T}{\beta} \right) \alpha \right]^{\frac{1}{\beta-1}} g(\alpha) d\alpha, \quad (\text{Total Card Supply})$$

$$TD = \frac{k}{1 - \tau_{m,e} - I} D \left(\frac{k}{1 - \tau_{m,e} - I} (1 + \tau_{c,e} - R) \right), \quad (\text{Total Card Demand})$$

$$TV = TD. \quad (\text{Card Market Clearing})$$

To simplify the analysis, we assume that α follows a Pareto distribution so that $g(\alpha) = \gamma L^{\gamma} / (\alpha^{\gamma+1})$, where $\gamma > 1$ and $\beta\gamma > 1 + \gamma$;¹³ consumer demand function takes the isoelastic form $D = \eta p_r^{-\varepsilon}$; and the pricing constraint $1 - \tau_{m,e} > I \geq \tau_{m,a} - \tau_{m,e}$ is not binding. Therefore, the above maximization problem can be rewritten as

¹³The size distribution of card issuers, like firm size distribution in many other industries, is highly positively skewed. Although the possible candidates for this group of distributions are far from unique, Pareto distribution has typically been used as a reasonable and tractable example in the empirical IO literature.

$$\underset{I}{Max} \quad \Omega^m = A(I - R - T)^{\beta\gamma} - E \quad (\text{Card Network Profit})$$

$$s.t. \quad B(I - R - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1} (1 + \tau_{c,e} - R)^{-\varepsilon}, \quad (\text{Card Market Clearing})$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I}, \quad (\text{Pricing Constraint I})$$

where

$$A = cKL\gamma\beta^{-\gamma} \left(\frac{K\beta}{\beta-1}\right)^{(1-\beta)\gamma} \left(\frac{\gamma}{\gamma - \frac{1}{\beta-1}} - 1\right); \quad B = \frac{L^\gamma \beta^{-\gamma} k^{\varepsilon-1}}{\eta} \left(\frac{\gamma}{\gamma - \frac{1}{\beta-1}}\right) \left(\frac{K\beta}{\beta-1}\right)^{1+\gamma-\beta\gamma}.$$

To simplify notations, we thereafter refer the ‘‘Card Market Clearing Equation’’ as the ‘‘CMC Equation’’; and refer the ‘‘Pricing Constraint I’’ as the ‘‘API Constraint’’, where API stands for ‘‘Alternative Payment Instruments’’. Denote the net card price $Z = I - R$, we can further rewrite the above maximization problem into a more intuitive form:

$$\underset{I}{Max} \quad \Omega^m = A(Z - T)^{\beta\gamma} - E \quad (\text{Card Network Profit})$$

$$s.t. \quad B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1} (1 + \tau_{c,e} + Z - I)^{-\varepsilon}, \quad (\text{CMC Equation})$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I}, \quad (\text{API Constraint})$$

where A, B are defined as before. It then becomes clear that a profit maximization card network would like to choose an optimal interchange fee I to maximize the net card price Z . To fully characterize the maximization problem, we need to discuss two scenarios: elastic demand ($\varepsilon > 1$) and inelastic demand ($\varepsilon \leq 1$).

2.2.1 Elastic Demand: $\varepsilon > 1$

When demand is elastic ($\varepsilon > 1$), the CMC Equation implies there is an interior maximum Z^m where

$$\partial Z^m / \partial I^m = 0 \implies \frac{1 + \tau_{c,e} + Z^m - I}{1 - \tau_{m,e} - I} = \frac{\varepsilon}{\varepsilon - 1} \quad \text{and} \quad \partial^2(Z^m) / \partial (I^m)^2 < 0.$$

Therefore, if the API constraint is not binding, the maximum is determined by the following conditions:

$$\frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I} = \frac{\varepsilon}{\varepsilon - 1},$$

$$B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1} (1 + \tau_{c,e} + Z - I)^{-\varepsilon},$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{\varepsilon}{\varepsilon - 1} \implies \varepsilon \geq \frac{1 + \tau_{c,a}}{\tau_{c,a} + \tau_{m,a}} > 1.$$

Given the above conditions, Proposition 1 characterizes the monopoly interchange fee I^m as follows.

Proposition 1 *If demand is elastic and the API Constraint is not binding (i.e., $\varepsilon \geq \frac{1 + \tau_{c,a}}{\tau_{c,a} + \tau_{m,a}} > 1$), the monopoly profit-maximizing interchange fee I^m satisfies:*

$$\partial I^m / \partial T < 0; \quad \partial I^m / \partial \tau_{m,e} < 0; \quad \partial I^m / \partial \tau_{c,e} < 0;$$

$$\partial I^m / \partial K < 0; \quad \partial I^m / \partial \tau_{m,a} = 0; \quad \partial I^m / \partial \tau_{c,a} = 0.$$

Proof. See Appendix A. ■

Similarly, we can derive comparative statics of other variables at the monopoly maximum, including consumer reward R^m ; net card price Z^m ; issuer α 's profit π_α and

volume V_α ; number of issuers N ; card network's profit Ω^m and volume TV ; before-reward retail price p_e , after-reward retail price p_r , and card consumers' consumption D . All the analytical results are reported in Table 1 (See Appendix A for the proofs).

Table 1. Comparative Statics: $\varepsilon \geq \frac{1+\tau_{c,a}}{\tau_{c,a}+\tau_{m,a}} > 1$
(Signs of Partial Derivatives)

| | I^m | R^m | Z^m | π_α | V_α | N | Ω^m | TV | p_e | p_r | D |
|--------------|-------|-------|-------|--------------|------------|-----|------------|------|-------|-------|-----|
| $\tau_{m,e}$ | - | - | - | - | - | - | - | - | - | 0 | 0 |
| $\tau_{c,e}$ | - | \pm | - | - | - | - | - | - | - | 0 | 0 |
| T | - | - | + | - | - | - | - | - | - | 0 | 0 |
| K | - | - | + | \pm | + | - | + | - | - | 0 | 0 |
| $\tau_{m,a}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\tau_{c,a}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 1 suggests that anything else being equal, we observe the following:

- As it becomes easier for merchants to accept card (a lower $\tau_{m,e}$), both interchange fee and consumer reward increase but interchange fee increases more, which leads to an increase of net card price. Meanwhile, profits and transaction volumes of individual issuers increase; number of issuers increases; profit and transaction volume of the card network increase; and before-reward retail price increases. However, after-reward retail price and card users' consumption stay the same.
- The above effects also hold if it becomes easier for consumers to use card (a lower $\tau_{c,e}$) or it costs less for the network to provide card services (a lower T).

However, there are two noticeable differences: for a lower $\tau_{c,e}$, consumer reward can either increase or decrease; for a lower T , net card price decreases.

- As the entry barrier of card issuers declines (a lower K), both interchange fee and consumer reward increase but consumer reward increases more, which leads to a decrease of net card price. As a result, all incumbent issuers suffer a decrease of transaction volume, while large issuers see a profit decrease but small issuers see a profit increase. Meanwhile, the number of issuers increases; profit of the card network decreases while transaction volume increases; and before-reward retail price increases. However, after-reward retail price stays the same and there is no change of card users' consumption.
- Merchants or consumers' costs of using non-card payment instruments ($\tau_{m,a}$ and $\tau_{c,a}$) have no effect.

Alternatively, if the API constraint is binding, the monopoly maximum satisfies the following conditions:

$$B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1}(1 + \tau_{c,e} + Z - I)^{-\varepsilon},$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} = \frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I},$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} < \frac{\varepsilon}{\varepsilon - 1} \implies \frac{1 + \tau_{c,a}}{\tau_{c,a} + \tau_{m,a}} > \varepsilon > 1.$$

Given the above conditions, Proposition 2 characterizes the monopoly interchange fee I^m as follows.

Proposition 2 *If demand is elastic and the API Constraint is binding (i.e., $\frac{1+\tau_{c,a}}{\tau_{c,a}+\tau_{m,a}} > \varepsilon > 1$), the monopoly profit-maximizing interchange fee I^m satisfies:*

$$\begin{aligned} \partial I^m / \partial T &< 0; & \partial I^m / \partial \tau_{m,e} &< 0; & \partial I^m / \partial \tau_{c,e} &< 0; \\ \partial I^m / \partial K &< 0; & \partial I^m / \partial \tau_{m,a} &> 0; & \partial I^m / \partial \tau_{c,a} &> 0. \end{aligned}$$

Proof. See Appendix A. ■

Similarly, we can derive comparative statics of other variables at the maximum. All the analytical results are reported in Table 2 (See Appendix A for the proofs).

Table 2. Comparative Statics: $\frac{1+\tau_{c,a}}{\tau_{c,a}+\tau_{m,a}} > \varepsilon > 1$
(Signs of Partial Derivatives)

| | I^m | R^m | Z^m | π_α | V_α | N | Ω^m | TV | p_e | p_r | D |
|--------------------------------|-----------------------|-------|-------|--------------|------------|-----|------------|------|-------|-------|-----|
| $\tau_{m,a}$ | + | + | + | + | + | + | + | + | + | + | - |
| $\tau_{c,a}$ | + | + | + | + | + | + | + | + | + | + | - |
| $\tau_{m,e}, \tau_{c,e}, T, K$ | Same signs as Table 1 | | | | | | | | | | |

Table 2 suggests that anything else being equal, we observe the following:

- As it becomes easier for merchants or consumers to use non-card payment instruments (a lower $\tau_{m,a}$ or $\tau_{c,a}$), interchange fee decreases more than that of consumer reward, which leads to a decrease of net card price. Meanwhile, profits and transaction volumes of individual issuers decrease; number of issuers decreases; profit and transaction volume of the card network decrease. In addition, before-and-after reward retail prices decrease and card users' consumption increases.
- The effects of other variables are the same as Table 1.

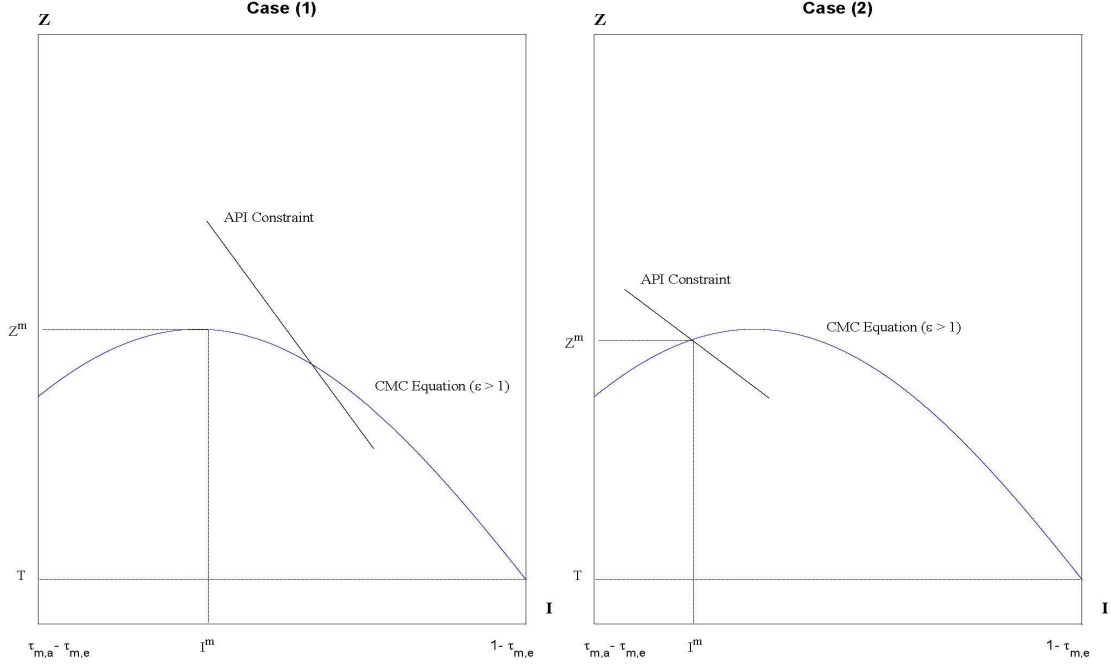


Figure 4: Monopoly Interchange Fee under Elastic Demand

Fig. 4 provides an intuitive illustration for the analysis. In the two graphs, the CMC Equation describes a concave relationship between the net card price Z (as well as the card network profit Ω^m , which increases with Z) and the interchange fee I for $I \in [\tau_{m,a} - \tau_{m,e}, 1 - \tau_{m,e})$. In Case (1), the API constraint is not binding so the monopoly card network can price at the interior maximum, on which $\tau_{m,a}$ and $\tau_{c,a}$ have no effects. Alternatively, in Case (2), the API constraint is binding so $\tau_{m,a}$ and $\tau_{c,a}$ do affect the interchange pricing. Particularly, at the constrained maximum (I^m, Z^m) , the curve of the CMC Equation has a slope less than 1. As a result, a local change of $\tau_{m,a}$ or $\tau_{c,a}$ shifts the line of the API Constraint, but Z^m changes less than I^m so that $\partial R^m / \tau_{m,a} > 0$ and $\partial R^m / \tau_{c,a} > 0$. Furthermore, in Cases (1) and (2), changes of other parameters, such as $\tau_{m,e}$, $\tau_{c,e}$, T , K , shift the curve of CMC Equation and affect the interchange pricing as described in Table 1 and 2.

2.2.2 Inelastic Demand: $\varepsilon \leq 1$

When demand is inelastic ($\varepsilon \leq 1$), the CMC Equation suggests that Z is an increasing function of I ($\partial Z/\partial I > 0$) and there is no interior maximum. Therefore, the API Constraint has to bind, The maximum satisfies the following conditions:

$$B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1}(1 + \tau_{c,e} + Z - I)^{-\varepsilon},$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} = \frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I}.$$

Given the above conditions, Proposition 3 characterizes the monopoly interchange fee I^m as follows.

Proposition 3 *If demand is inelastic (i.e., $\varepsilon \leq 1$), the API Constraint is binding and the monopoly profit-maximizing interchange fee I^m satisfies:*

$$\partial I^m/\partial T < 0; \quad \partial I^m/\partial \tau_{m,e} < 0; \quad \partial I^m/\partial \tau_{c,e} < 0;$$

$$\partial I^m/\partial K < 0; \quad \partial I^m/\partial \tau_{m,a} > 0; \quad \partial I^m/\partial \tau_{c,a} > 0.$$

Proof. See Appendix A. ■

Similarly, we can derive comparative statics of other variables at the maximum. All the analytical results are reported in Table 3 (See Appendix A for the proofs). The findings suggest that anything else being equal, we observe the following:

- The effects of $\tau_{m,a}$ and $\tau_{c,a}$ are the same as Table 2 except that consumer reward may either increase or decrease.
- The effects of other variables are the same as Tables 1 and 2.

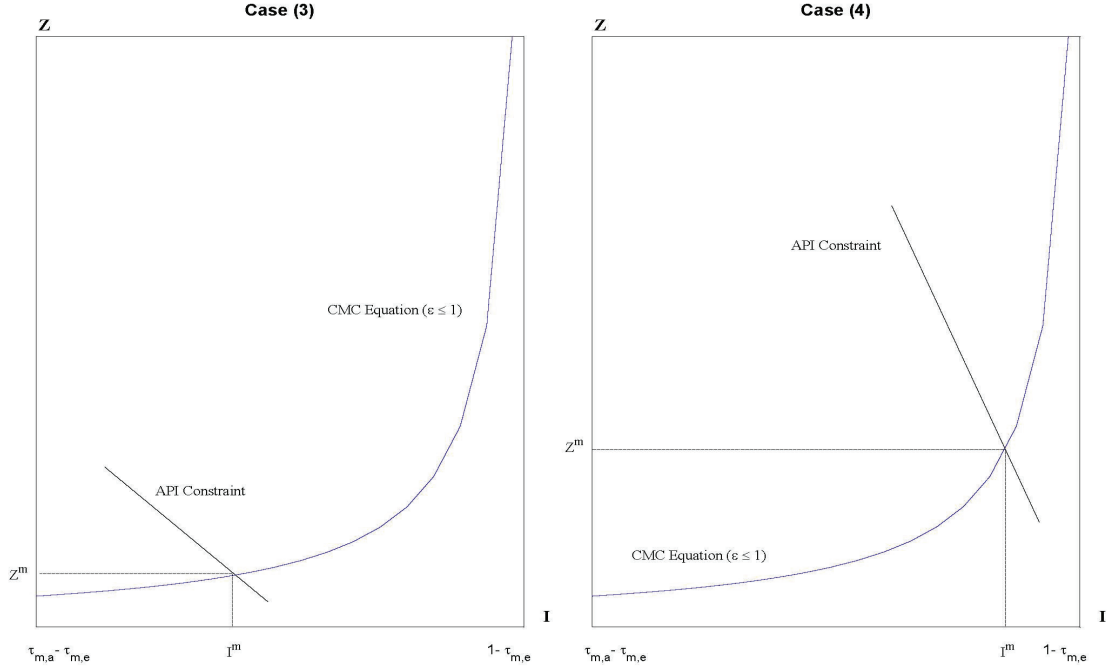


Figure 5: Monopoly Interchange Fee under Inelastic Demand

Table 3. Comparative Statics: $\varepsilon \leq 1$
(Signs of Partial Derivatives)

| | I^m | R^m | Z^m | π_α | V_α | N | Ω | TV | p_e | p_r | D^{14} |
|--------------------------------|------------------------------|-------|-------|--------------|------------|-----|----------|------|-------|-------|----------|
| $\tau_{m,a}$ | + | \pm | + | + | + | + | + | + | + | + | - |
| $\tau_{c,a}$ | + | \pm | + | + | + | + | + | + | + | + | - |
| $\tau_{m,e}, \tau_{c,e}, T, K$ | Same signs as Table 1 and 2. | | | | | | | | | | |

Fig. 5 provides an intuitive illustration of the analysis. In the two graphs, the CMC Equation describes an increasing and convex relationship between the net card price Z (as well as the card network profit Ω^m) and the interchange fee I for $I \in [\tau_{m,a} - \tau_{m,e}, 1 - \tau_{m,e}]$. Therefore, the API constraint has to bind so $\tau_{m,a}$ and $\tau_{c,a}$

¹⁴Notice that for $\varepsilon = 0$, we have $\partial D / \partial \tau_{m,a} = \partial D / \partial \tau_{c,a} = 0$.

affect the interchange pricing. In Case (3), at the constrained maximum (I^m, Z^m) , the curve of the CMC Equation has a slope less than 1. As a result, a local change of $\tau_{m,a}$ or $\tau_{c,a}$ shifts the line of the API Constraint, but Z^m changes less than I^m so that $\partial R^m / \tau_{m,a} > 0$ and $\partial R^m / \tau_{c,a} > 0$. Alternatively, in Case (4), at the constrained maximum (I^m, Z^m) , the curve of the CMC Equation has a slope greater than 1 so that $\partial R^m / \tau_{m,a} < 0$ and $\partial R^m / \tau_{c,a} < 0$. Furthermore, changes of other parameters, such as $\tau_{m,e}$, $\tau_{c,e}$, T , K , shift the curve of CMC Equation and affect the interchange pricing as described in Table 3.

2.2.3 Recap and Remarks

As shown in the above analysis, under a monopoly card network, equilibrium interchange fees tend to increase as credit cards become a more efficient payment instrument (a lower $\tau_{m,e}$, $\tau_{c,e}$ or T) or card issuers' market becomes more competitive (a lower K). These findings offer a consistent explanation for the empirical puzzle of rising interchange fees. Meanwhile, our analysis uncovers some major anti-competitive issue in this market: Given market power of card networks, technology progress or enhanced competition may drive up consumer rewards and card transaction volume, but does not necessarily improve consumer welfare.

Moreover, the theory consistently explains other puzzling facts in the credit card market. For example, why can't merchants refuse accepting cards given the rising interchange fees? The answer is simply that due to technology progress, card payments become increasingly more efficient than alternative payment instruments. Therefore, card networks can afford charging higher interchange fees but still keep cards as a competitive payment service to merchants and consumers. Another puzzle is why card networks, from a cross-section point of view, charge lower interchange fees on

transaction categories with lower fraud costs, e.g., face-to-face purchases with card present are generally charged a lower interchange rate than online purchases without card present. It might seem to contradict the time-series evidence that interchange fees increase as fraud costs decrease over time. Our analysis suggests that the answer lies on the different API (alternative payment instruments) constraints that card networks face in different environments. In an environment with higher fraud costs for cards, such as online shopping, the costs of using a non-card payment instrument is also likely to be higher, sometimes even prohibitively higher. Therefore, this allows card networks to demand higher interchange fees.

Exploring market outcomes based on a monopoly card network structure, we have offered a consistent explanation for the controversies surrounding credit card pricing. In the next section, we will show that our analysis can be readily carried over to the market with oligopolistic card networks.

2.3 Duopoly Outcome

So far, we have discussed the monopoly outcome in the credit card market. Will the competitive outcome be restored if there is more than one network in the industry? The answer is most likely a NO. In fact, given the tremendous scale economies of payment technology, only a few networks coexist in the industry and they have to interact repeatedly. Consequently, the networks would recognize their interdependence and might be able to sustain the monopoly price without explicit collusion. This is a well-known result from the literature of dynamic price competition. The intuition is as follows: In an oligopoly producing a homogeneous product, a firm must take into account not only the possible increase in current profits but also the possibility of a price war and long-run losses when deciding whether to undercut a given price.

In another word, the threat of a vigorous price war would be sufficient to deter the temptation to cut prices. Hence, the oligopolists might be able to collude in a purely noncooperative manner and the monopoly price is the most likely outcome.

To formalize this idea, let us consider a duopoly model in the credit card context: Two card networks that produce homogenous card services have the same cost structure as specified in Section 2.2. Let $\Omega^i(I_{it}, I_{jt})$ denote network i 's profit at period t when it charges interchange fee I_{it} and its rival charges I_{jt} . If the two networks charge the same interchange fee $I_{it} = I_{jt} = I_t$, they share the market, that is $\Omega^i = \Omega^j = \frac{1}{2}\Omega^m(I_t) - \frac{1}{2}E$, where $\Omega^m(I_t)$ is the monopoly network profit at the interchange fee level I_t . Otherwise, the lower-interchange network may get the whole market. This is suggested by the following proposition.

Proposition 4 *Anything else being equal, the CMC Equation implies $\partial p_r / \partial I > 0$.*

Proof. See Appendix A. ■

Proposition 4 says that a lower interchange fee results in a lower after-reward retail price for card consumers. Therefore, a lower-interchange network is able to attract all the merchants and card consumers.

In this market, each card network seeks to maximize the present discounted value of its profits; that is

$$U_i = \sum_{t=0}^{\infty} \delta^t \Omega^i(I_{it}, I_{jt}),$$

where δ is the discount factor (δ close to 1 represents low impatience or rapid price change).

At each period t , the networks choose their interchange fees (I_{it}, I_{jt}) simultaneously. There is no physical link between the periods but the interchange strategies at period t are allowed to depend on the history of previous interchanges $H_t \equiv$

$(I_{i0}, I_{j0}; \dots; I_{it-1}, I_{jt-1})$. The strategies are required to form a “perfect equilibrium”; that is, for any history H_t at date t , firm i ’s strategy from period t on maximizes the present discounted value of profits given network j ’s strategy from that period on.

Since the two networks are engaged in an infinite-horizon game, there exist many equilibrium strategies. In particular, we can show monopoly outcome can be supported at equilibrium. Consider the following symmetric strategies:

1. Phase A: set interchange fee at the monopoly level I^m and switch to Phase B;
2. Phase B: set interchange fee at I^m unless some player has deviated from I^m in the previous period, in which case switch to Phase C and set $\tau = 0$;
3. Phase C: if $\tau \leq n$, set $\tau = \tau + 1$ and charge the interchange fee at the punishment level I^p that $\Omega^i(I^p, I^p) = 0$, otherwise switch to Phase A.

This strategy, also known as Forgiving Trigger (FT), prescribes collusion in the first period, and then n periods of defection for every defection of any player, followed by reverting to cooperation no matter what has occurred during the punishment phase. Therefore, if a network undercuts the monopoly interchange fee I^m , it may earn a maximum profit $\Omega^m(I^m)$ during the period of deviation (indeed it earns approximately $\Omega^m(I^m)$ by slightly undercutting) but then it receives zero for n periods. Consequently, there will be no profitable one-shot deviation in the collusion phase if and only if

$$\frac{1}{2}\Omega^m(I^m) + \frac{1}{2}E < \frac{\delta(1 - \delta^n)}{1 - \delta} \left[\frac{1}{2}\Omega^m(I^m) - \frac{1}{2}E \right].$$

It can be shown that for a given n , if δ is large enough, (FT, FT) is a subgame perfect Nash equilibrium, and I^m can be supported at equilibrium. For example, if $n = 2$, the condition can be satisfied for any $\delta > \{[1 + (4\Omega^m(I^m) + 4E)/(\Omega^m(I^m) - E)]^{1/2} - 1\}/2$.

Moreover, as the length of punishment increases, the lower bound on δ decreases, and as $n \rightarrow \infty$, the bound converges to $(\Omega^m(I^m) + E)/(2\Omega^m(I^m))$. This corresponds to the harshest punishment, also known as Grim Trigger (GT).

This result is a formalization of tacit collusion that potential punishment enforces a collusion under equilibrium. Several things may need further clarification.

Table 4: Top Eight Credit Card Issuers in 2004

| ISSUERS | VISA | | MASTERCARD | |
|-----------------|------|-------------|------------|-------------|
| | Rank | # Cards (M) | Rank | # Cards (M) |
| JP Morgan Chase | 2 | 48.1 | 2 | 39.9 |
| Citigroup | 3 | 28.9 | 1 | 75.1 |
| MBNA | 5 | 24.4 | 3 | 32.3 |
| Bank of America | 1 | 58.1 | 8 | 3.1 |
| Capital One | 4 | 26.9 | 4 | 26.7 |
| HSBC | 7 | 10.3 | 5 | 24.4 |
| Provident | 8 | 10.1 | 11 | 2.5 |
| Wells Fargo | 10 | 7.1 | 9 | 2.8 |

First, the model shows that the collusion can be supported at equilibrium only if δ is large enough. This is because tacit collusion is enforced by the threat of punishment, but punishment can occur only when it is learned that someone has deviated. Therefore, it is necessary to sustain a collusion that any price cut by a player can be quickly observed and punished by its competitors. This condition is very likely to be met for the credit card industry. In fact, the two major US credit card networks, Visa and MasterCard, share the same group of card issuers and merchants

(see Tables 4 and 5). Therefore, there is minimal information lag of interchange pricing in this market.

Second, the assumption of infinite horizon is crucial for the results. It is known that collusion cannot be sustained even for a long but finite horizon due to backward induction. However, practically the infinite-horizon assumption need not to be taken too seriously. Suppose that at each period there is a probability θ in $(0, 1)$ that the market survives, i.e., that the networks keep competing on this market (in another word, $1 - \theta$ is the probability that the market completely changes). The game then ends in finite but stochastic time with probability 1. However, everything is as if the horizon were infinite and the network's discount factor were equal to $\delta\theta$.

Table 5: Visa and MasterCard Comparison 2004

| | VISA | MASTERCARD | TOTAL |
|--------------------|---------|------------|---------|
| Merchants(M) | 4.6 | 4.6 | 4.6 |
| Outlets(M) | 5.7 | 5.6 | 5.7 |
| Cardholders(M) | 96.2 | 96.3 | 118.5 |
| Cards(M) | 295.3 | 271.5 | 566.8 |
| Accounts(M) | 215.5 | 217.6 | 433.1 |
| Active Accts (M) | 115.2 | 120.1 | 235.3 |
| Transactions (M) | 7,286.8 | 5286.2 | 12573.0 |
| Total Volume (\$B) | 722.2 | 546.7 | 1268.9 |
| Outstandings (\$B) | 302.9 | 293.7 | 596.48 |

Last but not least, this infinitely repeated game has multiple equilibriums, as suggested by Folk Theorems. As a natural method, we assume that the networks

coordinate on an equilibrium that yields a Pareto-optimal point for the two networks, that is the monopoly outcome. Furthermore, we choose a symmetric equilibrium given the symmetric nature of the game. In fact, this result is consistent with the empirical observation that Visa and MasterCard have almost identical organizational structure and market shares (see Table 5). In addition, since we assume that the two networks play Forgiving Trigger strategies, the issue of renegotiation is not much a concern.

3 Policy and Welfare Analysis

The above analysis suggests that oligopolistic card networks are likely to collude to set monopoly interchange fees. At equilibrium, they demand higher interchange fees to maximize card issuers' profits as card payments become more efficient. Consequently, consumer rewards and retail price may increase but not the consumer surplus. Furthermore, under a more realistic assumption that merchants are heterogenous, we can show merchants' profits are affected by interchange fees in the same way as the card consumer surplus (see Appendix B). Based on this framework, we may proceed to evaluate the card industry performance and government interventions.

3.1 Policy Interventions

In many countries, public authorities have chosen to regulate down the interchange fees. It is of great interest to understand how that would affect card industry profits and consumer surplus. We now turn to this question.

As shown in Proposition 4, $\partial p_r / \partial I > 0$; that says a lower interchange fee results in a lower after-reward retail price and hence a higher consumers' consumption. Therefore, in order to increase consumer surplus, public authorities may have incentives to

suppress the interchange fees. Characterizing the CMC Equation

$$B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1}(1 + \tau_{c,e} + Z - I)^{-\varepsilon},$$

the following proposition predicts the likely effects:

Proposition 5 *Anything else being equal, the CMC Equation suggests that for $I < I^m$,*

$$\partial Z/\partial I > 0; \partial \pi_\alpha/\partial I > 0; \partial V_\alpha/\partial I > 0; \partial N/\partial I > 0;$$

$$\partial \Omega/\partial I > 0; \partial p_e/\partial I > 0; \partial p_r/\partial I > 0; \partial D/\partial I < 0;$$

$$\text{and } \partial R/\partial I > 0 \text{ for } \varepsilon > 1; \quad \partial R/\partial I \geq 0 \text{ for } \varepsilon \leq 1.$$

Proof. See Appendix A. ■

Proposition 5 says anything else being equal, reducing interchange rate below the constrained or unconstrained monopoly profit-maximizing level will result a lower net card price, lower profits and volumes for individual card issuers, a smaller number of issuers, lower profits and volumes for card networks, lower before-and-after-reward retail prices and higher card consumers' consumption. Note the effects on consumer reward depend on the elasticity of demand: for elastic demand, consumer reward decreases; for inelastic demand, consumer reward may either decrease or increases.

However, a one-time price cut may only have temporary effects since the interchange fees can easily come back. Alternatively, public authorities may set an interchange ceiling $I^c < I^m$. Given a binding interchange ceiling I^c , the market outcome is determined by the modified CMC Equation:

$$B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I^c)^{\varepsilon-1}(1 + \tau_{c,e} + Z - I^c)^{-\varepsilon},$$

where I^c is a constant. As a result, any changes of environmental parameters will then affect the industry differently from the non-intervention scenario.

For an elastic demand ($\varepsilon > 1$), the analytical results are reported in Table 6 (see Appendix A for the proofs).

Table 6. Comparative Statics: $\varepsilon > 1$ and I^c is binding
(Signs of Partial Derivatives)

| | I^c | R^c | Z^c | π_α | V_α | N | Ω^c | TV | p_e | p_r | D |
|--------------|-------|-------|-------|--------------|------------|-----|------------|------|-------|-------|-----|
| $\tau_{m,e}$ | 0 | + | - | - | - | - | - | - | + | + | - |
| $\tau_{c,e}$ | 0 | + | - | - | - | - | - | - | 0 | + | - |
| T | 0 | - | + | - | - | - | - | - | 0 | + | - |
| K | 0 | - | + | \pm | + | - | + | - | 0 | + | - |
| $\tau_{m,a}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\tau_{c,a}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 6 suggests that anything else being equal, we may observe the following under a binding interchange ceiling:

- As it becomes easier for merchants or consumers to use card (a lower $\tau_{m,e}$ or $\tau_{c,e}$), consumer reward decreases, which leads to an increase of net card price. As a result, profits and transaction volumes of individual issuers increase; number of issuers increases; profits and transaction volumes of card networks increase; after-reward retail price decreases; and card users' consumption increases. Meanwhile, a lower $\tau_{m,e}$ results a lower before-reward price, but a lower $\tau_{c,e}$ does not affect the before-reward price.

- The above effects also hold if it costs less for card networks to provide card services (a lower T). Note for a lower T , consumer reward increases and net card price decreases.
- As the entry barrier of card issuers declines (a lower K), consumer reward increases, which leads to a decrease of net card price. As a result, all incumbent issuers suffer a decrease of transaction volume, while large issuers see a profit decrease but small issuers see a profit increase. Meanwhile, the number of issuers increases; profits of card networks decrease but transaction volumes increase; after-reward retail price decreases; and card users' consumption increases. However, before-reward retail price stays the same.
- Merchants or consumers' costs of using non-card payment instruments ($\tau_{m,a}$ and $\tau_{c,a}$) have no effect.

For an inelastic demand ($0 < \varepsilon \leq 1$), the analytical results are reported in Table 7 (see Appendix A for the proofs).¹⁵

Table 7. Comparative Statics: $0 < \varepsilon \leq 1$ and I^c is binding

(Signs of Partial Derivatives)

| | | I^c | R^c | Z^c | π_α | V_α | N | Ω^c | TV | p_e | p_r | D |
|--|---------------------|------------------------|-------|-------|--------------|------------|-----|------------|------|-------|-------|-----|
| $\tau_{m,e}$ | $(\varepsilon < 1)$ | 0 | - | + | + | + | + | + | + | + | + | - |
| | $(\varepsilon = 1)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | + | + | - |
| $\tau_{c,e}, T, K, \tau_{m,a}, \tau_{c,a}$ | | Same signs as Table 6. | | | | | | | | | | |

¹⁵For a perfectly inelastic demand ($\varepsilon = 0$), the analytical results are reported in Table 8 in Appendix A.

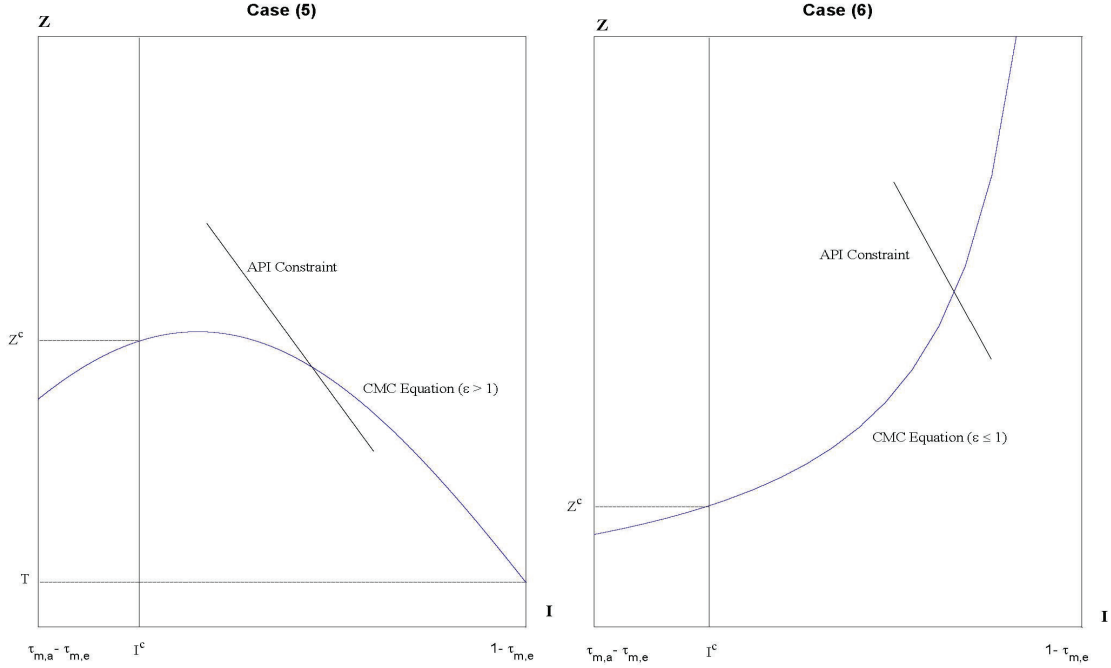


Figure 6: Interchange Fee Ceiling under Elastic/Inelastic Demand

Table 7 suggests that anything else being equal, we may observe the following under a binding interchange ceiling:

- For a unit elastic demand ($\varepsilon = 1$), a lower $\tau_{m,e}$ has no effect on card pricing, output and profits. For an inelastic demand ($\varepsilon < 1$), a lower $\tau_{m,e}$ will have opposite effects on card pricing, output and profits as the elastic demand. However, regardless of demand elasticity, a lower $\tau_{m,e}$ always lowers the before-and-after-reward retail prices and raises card users' consumption (except for a perfectly inelastic demand).
- The effects of other variables are the same as Tables 6.

The findings in Table 6 and 7 suggest that a binding interchange ceiling allows card consumers to benefit from technology progress or enhanced competition in the

credit card industry. These results are in sharp contrast with what we have seen in Table 1, 2 and 3 for the non-intervention scenario.

Fig. 6 illustrates the effects of the interchange ceiling. In the two graphs for Cases (5) and (6), the API Constraint is not binding so $\tau_{m,a}$ and $\tau_{c,a}$ have no effects. Furthermore, changes of other parameters, such as $\tau_{m,e}$, $\tau_{c,e}$, T , K , shift the curve of the CMC Equation. However, given a binding interchange ceiling, these changes can not raise the level of interchange fee but may affect other industry variables as described in Table 6 and 7.

3.2 Socially Optimal Pricing

Given the structure of credit card industry, Proposition 4 suggests consumer surplus increases as interchange fees decline. However, it may not be socially optimal to set the interchange fee at its minimum level where $\Omega^m(I) = 0$. In fact, the social planner aims to maximize the social surplus, i.e., the sum of producer surplus and consumer surplus. Accordingly, if the social planner runs the card network, he solves the following problem:

$$Max_I \Omega^s = \int_{\alpha^*}^{\infty} \pi_{\alpha} g(\alpha) d\alpha + \int_0^{Q^*} D^{-1}(Q) dQ - \frac{k(1 + \tau_{c,e} - R)}{1 - \tau_{m,e} - I} Q^* - E \quad (\text{Social Surplus})$$

$$s.t. \quad \pi_{\alpha} = \left(\frac{\beta - 1}{\beta}\right) \left(\frac{\alpha}{\beta}\right)^{\frac{1}{\beta-1}} (I - R - T)^{\frac{\beta}{\beta-1}} - K, \quad (\text{Profit of Issuer } \alpha)$$

$$\alpha^* = \beta K^{\beta-1} \left(\frac{\beta}{\beta-1}\right)^{\beta-1} (I - R - T)^{-\beta}, \quad (\text{Marginal Issuer } \alpha^*)$$

$$Q^* = D\left(\frac{k}{1 - \tau_{m,e} - I}(1 + \tau_{c,e} - R)\right) \quad (\text{Demand of Goods})$$

$$N = \int_{\alpha^*}^{\infty} g(\alpha)d\alpha, \quad (\text{Number of Issuers})$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I}, \quad (\text{Pricing Constraint I})$$

$$1 - \tau_{m,e} > I \geq \tau_{m,a} - \tau_{m,e}, \quad (\text{Pricing Constraint II})$$

$$TV = \int_{\alpha^*}^{\infty} V_{\alpha}g(\alpha)d\alpha = \int_{\alpha^*}^{\infty} \left[\left(\frac{I - R - T}{\beta}\right)\alpha\right]^{\frac{1}{\beta-1}}g(\alpha)d\alpha, \quad (\text{Total Card Supply})$$

$$TD = \frac{k}{1 - \tau_{m,e} - I}D\left(\frac{k}{1 - \tau_{m,e} - I}(1 + \tau_{c,e} - R)\right), \quad (\text{Total Card Demand})$$

$$TV = TD, \quad (\text{Card Market Clearing})$$

$$c \int_{\alpha^*}^{\infty} \pi_{\alpha}g(\alpha)d\alpha - E \geq 0. \quad (\text{Ramsey Constraint})$$

As before, we assume that α follows a Pareto distribution, i.e., $g(\alpha) = \gamma L^{\gamma}/(\alpha^{\gamma+1})$; consumer demand function takes the isoelastic form $D(p_r) = \eta p_r^{-\varepsilon}$; and the constraints $1 - \tau_{m,e} > I \geq \tau_{m,a} - \tau_{m,e}$ and $c \int_{\alpha^*}^{\infty} \pi_{\alpha}g(\alpha)d\alpha - E \geq 0$ are not binding.

For $\varepsilon > 1$, the above maximization problem then can be rewritten as

$$Max_I \Omega^s = \frac{A}{c}(Z - T)^{\beta\gamma} + \frac{\eta}{\varepsilon - 1}p_r^{1-\varepsilon} - E \quad (\text{Social Surplus})$$

$$s.t. \quad B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1}(1 + \tau_{c,e} + Z - I)^{-\varepsilon}, \quad (\text{CMC Equation})$$

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I}, \quad (\text{API Constraint})$$

where $Z = I - R$, $p_r = \frac{k(1+\tau_{c,e}+Z-I)}{(1-\tau_{m,e}-I)}$, and A, B are defined as before. Similarly, we can derive the social surplus maximization problem for $\varepsilon \leq 1$ (see Appendix A).

Denote I^s the socially optimal interchange fee. Note the social surplus consists of two parts. The first part, the card network profits, increases with the interchange fee. The second part, the consumer surplus, decreases with the interchange fee. Therefore, we expect that the maximum of social surplus requires an interchange fee I^s lower than that of monopoly I^m . This is shown in the following proposition.

Proposition 6 *The socially optimal interchange level I^s is generally lower than that of monopoly I^m , i.e., $I^s \leq I^m$.*

Proof. This result holds for both elastic and inelastic demand. See Appendix A for a detailed proof. ■

3.3 Further Considerations

The above policy and welfare analysis offers some justification for the concerns and actions that public authorities worldwide have on the credit card interchange pricing. However, it by no means implies that government interventions would be an easy task and can be implemented without care. In fact, several additional issues have made policy interventions in this market a rather difficult job.

First, our welfare analysis has treated technology progress in the credit card market as exogenously given. In reality, it is more likely that card technology is driven by intended R&D efforts of the card networks, and network profits provide major incentives and resources for these efforts. With endogenous technology progress, although it is still true that card networks and issuers may keep most of the benefits generated by technology progress as shown in our model, it has made the social surplus calculation more complicate. On one hand, regulating down interchange fees may improve consumer surplus, but on the other hand, it may hurt technology progress in the card industry and cause efficiency losses in the long run. Moreover, if the card technology has spillover effects on other payment instruments, the regulation may also affect negatively on consumer welfare.

Second, our analysis has assumed that market costs of payment instruments reflect their social costs. In reality, it may not be true. In some cases, when market costs of alternative payment instruments are lower than their social costs, the binding API constraint of card pricing may already lower interchange fees from where they otherwise would be. Therefore, adequate information on total social costs of various payment instruments is a prerequisite for designing and implementing good policy in payments markets.

Third, our model assumes that merchants are perfectly competitive. As shown in the model, card and cash consumers choose to shop at different stores (see footnote 7 for more discussions). Hence, the no-surcharge rule of credit cards does not play an important role. In fact, this is consistent with compelling empirical evidence: Even though under some circumstances merchants are allowed to surcharge credit card use, few of them have chosen to do so. However, competitive market may be a reasonable approximation for many industries but not for all. In some monopolistic markets,

there may be a monopoly merchant who serves both card and cash customers. Then, the no-surcharge rule imposed by the monopoly card network may help solve the double margin problem and can be welfare enhancing (see Schwartz and Vincent 2006 for more discussions).

Fourth, direct price regulation is not the only option or necessarily the best option for public authorities to improve market outcomes. There are always other policy mixes worthy of exploring. In the case of credit card market, policy interventions may alternatively apply to market structure, e.g., enforcing competition between card networks; or apply to competing products, e.g., encouraging technology progress in non-card payments. In addition, increasing public scrutiny and rising regulatory threat may also be effective policy measures.

Last but not least, government interventions may render unintended consequences. This is more likely to happen in a complex environment like the credit card market. Therefore a thorough study of the market structure can not be over emphasized. This paper is one of the beginning steps toward this direction, and many issues need further research: the market definition of various payment instruments, the competition between four-party systems and three-party systems, the causes and consequents of credit card rules, just to name a few.

4 Conclusion

As credit cards become an increasingly prominent form of payments, the structure and performance of this industry has attracted intensive scrutiny. This paper presents an industry equilibrium model to better understand the credit card market. The market that we consider consists of competing payment instruments, e.g., credit cards

vs. alternative payment methods; rational consumers (merchants) that always use (accept) the lowest-cost payment instruments; oligopolistic card networks that set profit-maximizing interchange fees; and competitive card issuers that join the most profitable network and compete with one another via consumer rewards.

Exploring the oligopolistic structure of this market, our model derives equilibrium industry dynamics consistent with empirical facts. It suggests that market power of credit card networks plays a critical role in determining the card pricing. In particular, credit card networks are likely to collude and demand higher interchange fees to maximize card issuers' profits as card payments become more efficient. At equilibrium, consumer rewards and card transaction volume also increase, while consumer surplus and merchant profits may not. Based on this theoretical framework, consequences and risks of government interventions in the credit card market are discussed in depth.

Appendix A.

Proof. (Proposition 1): If demand is elastic and the API Constraint is not binding (i.e., $\varepsilon \geq \frac{1+\tau_{c,a}}{\tau_{c,a}+\tau_{m,a}} > 1$), the monopoly profit-maximizing interchange fee I^m satisfies:

$$\frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I} = \frac{\varepsilon}{\varepsilon - 1}, \quad (\text{FOC})$$

$$B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1}(1 + \tau_{c,e} + Z - I)^{-\varepsilon}. \quad (\text{CMC})$$

Therefore, eqs. FOC and CMC imply

$$B\left(\frac{1}{\varepsilon - 1} - \frac{\varepsilon}{\varepsilon - 1}\tau_{m,e} - \frac{1}{\varepsilon - 1}I^m - \tau_{c,e} - T\right)^{\beta\gamma-1} = \left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon}(1 - \tau_{m,e} - I^m)^{-1}.$$

All the results then are derived by implicit differentiation. ■

Proof. (Table 1): Results in the first column are given by Proposition 1. Note eqs. FOC and CMC imply

$$B(Z - T)^{\beta\gamma-1} = (\varepsilon - 1)^{\varepsilon-1}(\varepsilon)^{-\varepsilon}(\tau_{c,e} + Z + \tau_{m,e})^{-1}.$$

The results in column 3 then are derived by implicit differentiation. Recall that all other endogenous variables are functions of Z , I and parameters:

$$\begin{aligned} R &= I - Z; & \pi_\alpha &= \left(\frac{\beta-1}{\beta}\right)\left(\frac{\alpha}{\beta}\right)^{\frac{1}{\beta-1}}(Z - T)^{\frac{\beta}{\beta-1}} - K; \\ V_\alpha &= \left(\frac{\alpha}{\beta}(Z - T)\right)^{\frac{1}{\beta-1}}; & \alpha^* &= \beta K^{\beta-1}\left(\frac{\beta}{\beta-1}\right)^{\beta-1}(Z - T)^{-\beta}; \\ N &= \int_{\alpha^*}^{\infty} g(\alpha)d\alpha = (L/\alpha^*)^\gamma; & \Omega^m &= A(Z - T)^{\beta\gamma} - E; \\ TV &= B(Z - T)^{\beta\gamma-1}k^{1-\varepsilon}; & p_e &= \frac{k}{1-\tau_{m,e}-I}; \\ p_r &= \frac{(1+\tau_{c,e}+Z-I)}{(1-\tau_{m,e}-I)}k; & D &= \eta p_r^{-\varepsilon}; \\ A &= cKL\gamma\beta^{-\gamma}\left(\frac{K\beta}{\beta-1}\right)^{(1-\beta)\gamma}\left(\frac{\gamma}{\gamma-\beta-1} - 1\right); & B &= \frac{L\gamma\beta^{-\gamma}k^{\varepsilon-1}}{\eta}\left(\frac{\gamma}{\gamma-\beta-1}\right)\left(\frac{K\beta}{\beta-1}\right)^{1+\gamma-\beta\gamma}. \end{aligned}$$

The other results in the table then are derived by differentiation. ■

Proof. (Proposition 2): If demand is elastic and the API Constraint is binding (i.e., $\frac{1+\tau_{c,a}}{\tau_{c,a}+\tau_{m,a}} > \varepsilon > 1$), the monopoly profit-maximizing interchange fee I^m satisfies:

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} = \frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I}, \quad (\text{API})$$

$$B(Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\varepsilon-1}(1 + \tau_{c,e} + Z - I)^{-\varepsilon}. \quad (\text{CMC})$$

Therefore, eqs. API and CMC imply

$$B(I - 1 - \tau_{c,e} + \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}}(1 - \tau_{m,e} - I) - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{-1}(\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}})^{-\varepsilon}.$$

All the results then are derived by implicit differentiation. ■

Proof. (Table 2): Results in the first column are given by Proposition 2. Note eqs. API and CMC imply

$$B(Z - T)^{\beta\gamma-1} = (\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} - 1)(\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}})^{-\varepsilon}(\tau_{m,e} + \tau_{c,e} + Z)^{-1}.$$

The results in column 3 then are derived by implicit differentiation. Recall that all other endogenous variables are functions of Z , I and parameters, as shown in the proof of Table 1. The other results in the table then are derived by differentiation. ■

Proof. (Proposition 3): Similar to the proof of Proposition 2. ■

Proof. (Table 3): Similar to the proof of Table 2. The different results between Table 2 and Table 3 come from their different demand elasticity ε . ■

Proof. (Proposition 4): Implicit differentiation on the CMC equation implies

$$\frac{\partial Z}{\partial I} = -\left[\frac{(\varepsilon - 1)(1 - \tau_{m,e} - I)^{-1}(1 + \tau_{c,e} + Z - I) - \varepsilon}{(\beta\gamma - 1)(Z - T)^{-1}(1 + \tau_{c,e} + Z - I) + \varepsilon}\right].$$

Recall

$$p_r = (1 + \tau_{c,e} + Z - I)k(1 - \tau_{m,e} - I)^{-1}.$$

Therefore, we derive

$$\frac{\partial p_r}{\partial I} = k(1 - \tau_{m,e} - I)^{-1} \left(\frac{\partial Z}{\partial I} - 1 \right) + (1 + \tau_{c,e} + Z - I)k(1 - \tau_{m,e} - I)^{-2} > 0.$$

■

Proof. (Proposition 5): As shown in the proof of Proposition 4, we have

$$\frac{\partial Z}{\partial I} = - \left[\frac{(\varepsilon - 1)(1 - \tau_{m,e} - I)^{-1}(1 + \tau_{c,e} + Z - I) - \varepsilon}{(\beta\gamma - 1)(Z - T)^{-1}(1 + \tau_{c,e} + Z - I) + \varepsilon} \right],$$

which implies $\partial Z/\partial I > 0$ for $I < I^m$ for both $\varepsilon > 1$ and $\varepsilon \leq 1$. Recall that all other endogenous variables are functions of Z , I and parameters, as shown in the proof of Table 1. The other results then are derived by differentiation. ■

Proof. (Table 6): Given that the interchange ceiling is binding, the CMC equation becomes

$$B(Z - T)^{\beta\gamma - 1} = (1 - \tau_{m,e} - I^c)^{\varepsilon - 1}(1 + \tau_{c,e} + Z - I^c)^{-\varepsilon}.$$

where I^c is a constant. The results in column 3 then are derived by implicit differentiation. Recall that all other endogenous variables are functions of Z , I^c and parameters, as shown in the proof of Table 1. The other results in the table then are derived by differentiation. ■

Proof. (Tables 7 and 8): Table 8 below shows the analytical results for the case of perfectly inelastic demand ($\varepsilon = 0$). The proofs of Tables 7 and 8 are similar to the proof of Table 6. The different results between Table 6 and Tables 7 & 8 come from their different demand elasticity ε . ■

Table 8. Comparative Statics: $\varepsilon = 0$ and I^c is binding
(Signs of Partial Derivatives)

| | I^c | R^c | Z^c | π_α | V_α | N | Ω^c | TV | p_e | p_r | D |
|--------------|-------|-------|-------|--------------|------------|-----|------------|------|-------|-------|-----|
| $\tau_{m,e}$ | 0 | - | + | + | + | + | + | + | + | + | 0 |
| $\tau_{c,e}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | + | 0 |
| T | 0 | - | + | 0 | 0 | 0 | 0 | 0 | 0 | + | 0 |
| K | 0 | - | + | \pm | + | - | + | 0 | 0 | + | 0 |
| $\tau_{m,a}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\tau_{c,a}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Proof. (Proposition 6): For $\varepsilon > 1$, the social surplus maximization problem is

$$Max_I \Omega^s = \frac{A}{c}(Z - T)^{\beta\gamma} + \frac{\eta}{\varepsilon - 1} p_r^{1-\varepsilon} - E$$

Consider the following two cases. First, if the API Constraint is not binding, the monopoly's problem requires $\partial Z^m / \partial I^m = 0$ for the CMC Equation. Accordingly, the social planner's problem implies

$$\frac{\partial \Omega^s}{\partial I^m} = \frac{\partial \Omega^s}{\partial Z} \frac{\partial Z}{\partial I^m} + \frac{\partial \Omega^s}{\partial p_r} \frac{\partial p_r}{\partial I^m} = -\eta p_r^{-\varepsilon} \frac{\partial p_r}{\partial I^m} < 0$$

since Proposition 4 shows $\partial p_r / \partial I > 0$. Therefore, $I^s < I^m$. Alternatively, if the API Constraint is binding, (Z^m, I^m) have to satisfy both the CMC Equation and the API Constraint, and $\partial Z^m / \partial I^m > 0$ for the CMC Equation. Accordingly, the social planner's problem implies

$$\frac{\partial \Omega^s}{\partial I^m} = \frac{\partial \Omega^s}{\partial Z} \frac{\partial Z}{\partial I^m} + \frac{\partial \Omega^s}{\partial p_r} \frac{\partial p_r}{\partial I^m} = \frac{A}{c} \beta \gamma (Z - T)^{\beta\gamma-1} \frac{\partial Z}{\partial I^m} - \eta p_r^{-\varepsilon} \frac{\partial p_r}{\partial I^m}.$$

Then, if $\partial \Omega^s / \partial I^m < 0$, we have $I^s < I^m$; otherwise, if $\partial \Omega^s / \partial I^m \geq 0$, $I^s = I^m$.

For $\varepsilon \leq 1$, the analysis would be very similar. However, we then need a technical assumption to ensure that consumer surplus is bounded, e.g., $D(p_r) = \eta p_r^{-\varepsilon}$ for $D(p_r) \geq Q_0 > 0$, and $\int_0^{Q_0} D^{-1}(Q)dQ = H < \infty$. If $\varepsilon = 1$, the social surplus maximization can be written as

$$\underset{I}{Max} \quad \Omega^s = \frac{A}{c}(Z - T)^{\beta\gamma} + H - \eta \ln Q_0 - \eta + \eta \ln \eta - \eta \ln p_r - E.$$

Alternatively if $\varepsilon < 1$, the social surplus maximization can be written as

$$\underset{I}{Max} \quad \Omega^s = \frac{A}{c}(Z - T)^{\beta\gamma} + H + \frac{\varepsilon}{1 - \varepsilon} \eta^{1/\varepsilon} Q_0^{1-1/\varepsilon} + \frac{\eta}{\varepsilon - 1} p_r^{1-\varepsilon} - E;$$

or if $\varepsilon = 0$, we have

$$\underset{I}{Max} \quad \Omega^s = \frac{A}{c}(Z - T)^{\beta\gamma} + H - p_0 Q_0 + (p_0 - p_r)\eta - E,$$

where p_0 is consumers' highest willingness to pay for $Q \in (Q_0, \eta)$. In each case, a similar proof as the elastic demand case then shows that $I^s \leq I^m$. ■

Appendix B.

In the paper, merchants are assumed to be identical. As a result, they always break even regardless of interchange fees. Although this assumption help simplify our analysis, it does not explicitly explain merchants' motivation for lowering interchange fees. In this appendix, we show that under a more realistic assumption that merchants are heterogenous, their profits are indeed affected by interchange fees in the same way as the card consumer surplus.

As before, we assume a continuum of merchants sell a homogenous good in a competitive market. A merchant θ incurs a fixed cost W each period and faces an operational cost q_θ^φ/θ for its sale q_θ , where $\varphi > 1$. Merchants are heterogenous in

their operational efficiency θ , which follows a Pareto distribution over the population with pdf $f(\theta) = \phi J^\phi / (\theta^{\phi+1})$, $\phi > 1$ and $\phi\varphi > 1 + \phi$. Merchants have two options to receive payments. Accepting non-card payments, such as cash, costs merchants $\tau_{m,a}$ per dollar. Accepting card payments costs merchants $\tau_{m,e} + I$ per dollar. Therefore, a merchant who does not accept cards (i.e., cash store) charges p_a , while a merchant who accepts cards (i.e., card store) charges p_e . The share of card merchants is λ and the share of cash merchants is $1 - \lambda$. The values of p_a , p_e , and λ are endogenously determined as follows.

A merchant θ may earn profit $\pi_{\theta,e}$ for serving the card consumers:

$$\pi_{\theta,e} = \underset{q_\theta}{Max}(1 - \tau_{m,e} - I)p_e q_\theta - \frac{q_\theta^\varphi}{\theta} - W.$$

Alternatively, it may earn profit $\pi_{\theta,a}$ for serving the cash consumers:

$$\pi_{\theta,a} = \underset{q_\theta}{Max}(1 - \tau_{m,a})p_a q_\theta - \frac{q_\theta^\varphi}{\theta} - W.$$

At equilibrium, firms of the same efficiency must earn the same for serving either card or cash consumers. Therefore, it is required that

$$(1 - \tau_{m,e} - I)p_e = (1 - \tau_{m,a})p_a. \quad (3)$$

Note that the pricing of p_e requires $p_a \leq p_e$ so that card stores do not attract cash users. Eq. 3 then implies

$$I \geq \tau_{m,a} - \tau_{m,e}.$$

Meanwhile, card consumers do not shop cash stores if and only if

$$(1 + \tau_{c,a})p_a \geq (1 + \tau_{c,e} - R) p_e.$$

Eq. 3 then implies

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I}.$$

In addition, $1 - \tau_{m,e} > I$ is required for a meaningful pricing. Note all these inter-change pricing constraints are the same as what we derived for identical merchants.

Solving the profit-maximizing problem, a merchant θ has sale q_θ and profit π_θ for serving card consumers,

$$q_\theta = \left[\frac{\theta}{\varphi} (1 - \tau_{m,e} - I) p_e \right]^{\frac{1}{\varphi-1}}; \quad \pi_\theta = \frac{\varphi - 1}{\varphi} \left(\frac{\theta}{\varphi} \right)^{\frac{1}{\varphi-1}} [(1 - \tau_{m,e} - I) p_e]^{\frac{\varphi}{\varphi-1}} - W;$$

which would be the same at the equilibrium if it serves cash consumers.

Free entry condition requires that the marginal card merchant θ^* breaks even, so we have

$$\pi_{\theta^*,e} = 0 \implies \theta^* = \varphi \left(\frac{\varphi W}{\varphi - 1} \right)^{\varphi-1} [(1 - \tau_{m,e} - I) p_e]^{-\varphi}.$$

Then, the total supply of goods by card stores is

$$Q_{s,e} = \lambda \int_{\theta^*}^{\infty} q_{\theta,e} f(\theta) d\theta = \Psi \lambda [(1 - \tau_{m,e} - I) p_e]^{\phi\varphi-1},$$

where $\Psi = \varphi^{-\phi} \left(\frac{W\varphi}{\varphi-1} \right)^{1+\phi-\phi\varphi} \phi J^\phi \left(\frac{1}{\phi-\frac{1}{\varphi-1}} \right)$. At the same time, the total demand of goods by card consumers is

$$Q_{d,e} = \eta_e [(1 + \tau_{c,e} - R) p_e]^{-\varepsilon},$$

where η_e is related to the measure of card consumers. Therefore, the good market equilibrium achieved via card payments requires

$$Q_{s,e} = Q_{d,e} \implies \Psi \lambda [(1 - \tau_{m,e} - I) p_e]^{\phi\varphi-1} = \eta_e [(1 + \tau_{c,e} - R) p_e]^{-\varepsilon},$$

which implies the price charged in a card store is

$$p_e = \left[\frac{\Psi \lambda}{\eta_e} (1 - \tau_{m,e} - I)^{\phi\varphi-1} (1 + \tau_{c,e} - R)^\varepsilon \right]^{\frac{1}{1-\phi\varphi-\varepsilon}}.$$

Similarly, the price charged in a cash store is

$$p_a = \left[\frac{\Psi(1-\lambda)}{\eta_a} (1 - \tau_{m,a})^{\phi\varphi-1} (1 + \tau_{c,a})^\varepsilon \right]^{\frac{1}{1-\phi\varphi-\varepsilon}},$$

where η_a is related to the measure of cash consumers

At equilibrium, eq. 3 can then pin down the share of merchants accepting cards
verse cash:

$$\frac{\lambda}{1-\lambda} = \frac{\eta_e}{\eta_a} \left(\frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} \right)^{-\varepsilon} \left(\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \right)^{\varepsilon}.$$

In the market, the total demand of card transaction volume now becomes

$$\begin{aligned} TD &= p_e \eta_e [(1 + \tau_{c,e} - R)p_e]^{-\varepsilon} \\ &= \Psi^{\frac{1-\varepsilon}{1-\phi\varphi-\varepsilon}} \eta_e \left[\eta_a \left(\frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} \right)^{\varepsilon} \left(\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \right)^{-\varepsilon} + \eta_e \right]^{\frac{\varepsilon-1}{1-\phi\varphi-\varepsilon}} \\ &\quad \cdot (1 - \tau_{m,e} - I)^{\frac{(1-\varepsilon)(\phi\varphi-1)}{1-\phi\varphi-\varepsilon}} (1 + \tau_{c,e} - R)^{\frac{\varepsilon\phi\varphi}{1-\phi\varphi-\varepsilon}}. \end{aligned}$$

Recall the total supply of card transaction volume derived in Section 2.2:

$$\begin{aligned} TV &= \int_{\alpha^*}^{\infty} \left[\left(\frac{I - R - T}{\beta} \right) \alpha \right]^{\frac{1}{\beta-1}} g(\alpha) d\alpha \\ &= \gamma L^{\gamma} \beta^{-\gamma} \left(\frac{1}{\gamma - \frac{1}{\beta-1}} \right) \left(\frac{K\beta}{\beta-1} \right)^{1+\gamma-\beta\gamma} (I - R - T)^{\beta\gamma-1}. \end{aligned}$$

Therefore, the card market equilibrium $TD = TV$ implies

$$\begin{aligned} \Theta (I - R - T)^{\beta\gamma-1} &= \left[\eta_a \left(\frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} \right)^{\varepsilon} \left(\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \right)^{-\varepsilon} + \eta_e \right]^{\frac{\varepsilon-1}{1-\phi\varphi-\varepsilon}} \\ &\quad \cdot (1 - \tau_{m,e} - I)^{\frac{(1-\varepsilon)(\phi\varphi-1)}{1-\phi\varphi-\varepsilon}} (1 + \tau_{c,e} - R)^{\frac{\varepsilon\phi\varphi}{1-\phi\varphi-\varepsilon}}. \end{aligned}$$

where $\Theta = \frac{\gamma}{\eta_e} L^{\gamma} \beta^{-\gamma} \left(\frac{1}{\gamma - \frac{1}{\beta-1}} \right) \left(\frac{K\beta}{\beta-1} \right)^{1+\gamma-\beta\gamma} \Psi^{\frac{\varepsilon-1}{1-\phi\varphi-\varepsilon}}$.

As before, assuming the pricing constraint $1 - \tau_{m,e} > I \geq \tau_{m,a} - \tau_{m,e}$ is not binding. The monopoly card network then solves the following problem:

$$\begin{aligned} \underset{I}{Max} \quad \Omega^m &= A(I - R - T)^{\beta\gamma} - E && \text{(Card Network Profit)} \\ \text{s.t.} \quad \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} &\geq \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I}, && \text{(API Constraint)} \end{aligned}$$

$$\begin{aligned}\Theta(I - R - T)^{\beta\gamma-1} &= [\eta_a \left(\frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I}\right)^\varepsilon \left(\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}}\right)^{-\varepsilon} + \eta_e] \frac{\varepsilon-1}{1-\phi\varphi-\varepsilon} \\ &\quad (1 - \tau_{m,e} - I)^{\frac{(1-\varepsilon)(\phi\varphi-1)}{1-\phi\varphi-\varepsilon}} (1 + \tau_{c,e} - R)^{\frac{\varepsilon\phi\varphi}{1-\phi\varphi-\varepsilon}},\end{aligned}$$

(CMC Condition)

where

$$A = cKL^\gamma\beta^{-\gamma} \left(\frac{K\beta}{\beta-1}\right)^{(1-\beta)\gamma} \left(\frac{\gamma}{\gamma - \frac{1}{\beta-1}} - 1\right); \quad \Theta = \frac{\gamma}{\eta_e} \left(\frac{L^\gamma\beta^{-\gamma}}{\gamma - \frac{1}{\beta-1}}\right) \left(\frac{K\beta}{\beta-1}\right)^{1+\gamma-\beta\gamma} \Psi \frac{\varepsilon-1}{1-\phi\varphi-\varepsilon}.$$

Following a similar analysis as for identical merchants, we then can show merchants' profits are affected by interchange fees in the same way as the card consumer surplus. Particularly, when the API Constraint is binding, the monopoly maximum satisfies the following conditions:

$$\Theta(I - R - T)^{\beta\gamma-1} = (\eta_a + \eta_e) \frac{\varepsilon-1}{1-\phi\varphi-\varepsilon} (1 - \tau_{m,e} - I)^{\frac{(1-\varepsilon)(\phi\varphi-1)}{1-\phi\varphi-\varepsilon}} (1 + \tau_{c,e} - R)^{\frac{\varepsilon\phi\varphi}{1-\phi\varphi-\varepsilon}};$$

$$\frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} = \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}}.$$

Define $Z = I - R$ and $\nu = \frac{-\varepsilon\phi\varphi}{1-\phi\varphi-\varepsilon}$. The above condition then can be rewritten as

$$\Theta(\eta_a + \eta_e) \frac{1-\varepsilon}{1-\phi\varphi-\varepsilon} (Z - T)^{\beta\gamma-1} = (1 - \tau_{m,e} - I)^{\nu-1} (1 + \tau_{c,e} - R)^{-\nu};$$

$$\frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I} = \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}}.$$

Note that $\nu \geq 1$ if and only if $\varepsilon \geq 1$, so the equilibrium conditions are indeed equivalent to what we derived for identical merchants.

Now merchants' motivation for lowering interchange fees becomes clear: Credit card networks, given market power, may charge higher interchange fees to maximize card issuers' profits as card payments become more efficient. Consequently, technology progress or enhanced competition in the card industry drives up consumer rewards and card transaction volume, but may not increase consumer surplus or merchant profits. Our analysis suggests that by forcing down interchange, after-reward retail prices may decrease and card consumer consumption may increase. This could subsequently raise market demand for merchant sales, and hence increase merchant profits.

References

- [1] Baxter, William (1983), "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics* 26: 541-588.
- [2] Carlton, Dennis W., and Alan S. Frankel (1995), "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal* 63: 228-248.
- [3] Evans, David and Richard Schmalensee (2005), *Paying with Plastic: The Digital Revolution in Buying and Borrowing* 2nd Ed. MIT Press, Cambridge, MA.
- [4] Evans, David and Richard Schmalensee (2006), "The Economics of Interchange Fees and Their Regulation: An Overview," Federal Reserve Bank of Kansas City conference proceedings: *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* 73-120.
- [5] Farrell, Joseph (2006), "Efficiency and Competition between Payment Instruments," *Review of Network Economics* 5: 26-44.

- [6] Frankle, Alan (2006), “Interchange Fees in Various Countries: Commentary on Weiner and Wright,” Federal Reserve Bank of Kansas City conference proceedings: *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* 51-61.
- [7] Frisch, Adam and Stephen Stout (2005), *Visa 101: Overview of a Payments Company*. UBS Securities LLC.
- [8] Hayashi, Fumiko (2006), “A Puzzle of Card Payment Pricing: Why Are Merchants Still Accepting Card Payments?” *Review of Network Economics* 5: 144-174.
- [9] McAndrews, James and Zhu Wang (2006), “Microfoundations of Two-Sided Markets: A Payment Card Example,” Payments System Research Working Paper, Federal Reserve Bank of Kansas City.
- [10] Katz, Michael (2001), “Reform of Credit Card Schemes in Australia II: Commissioned Report” RBA Public Document. August.
- [11] Rochet, Jean-Charles (2003), “The Theory of Interchange Fees: A Synthesis of Recent Contributions,” *Review of Network Economics*, 2: 97-124.
- [12] Rochet, Jean-Charles and Jean Tirole (2002), “Cooperation among Competitors: Some Economics of Payment Card Associations,” *RAND Journal of Economics* 33: 549-570.
- [13] Rochet, Jean-Charles and Jean Tirole (2003), “An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems,” *Review of Network Economics*, 2: 69-79.

- [14] Rochet, Jean-Charles and Jean Tirole (2006), “Externalities and Regulation in Card Payment Systems,” *Review of Network Economics* 5: 1-14.
- [15] Schmalensee, Richard (2002), “Payment Systems and Interchange Fees,” *Journal of Industrial Economics* 50: 103-122.
- [16] Schwartz, Marius and Daniel Vincent (2006), “The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network,” *Review of Network Economics* 5: 72-102.
- [17] Tirole, Jean (1988), *The Theory of Industrial Organization*, The MIT Press.
- [18] Wang, Zhu, (2006), “Learning, Diffusion and Industry Life Cycle,” Payments System Research Working Paper, Federal Reserve Bank of Kansas City.
- [19] Weiner, Stuart and Julian Wright (2006), “Interchange Fees in Various Countries: Developments and Determinants,” Federal Reserve Bank of Kansas City conference proceedings: *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities?* 5-50.
- [20] Wright, Julian (2003), “Optimal Card Payment Systems,” *European Economic Review*, 47: 587-612.
- [21] Wright, Julian (2004), “Determinants of Optimal Interchange Fees in Payment Systems,” *Journal of Industrial Economics*, 52: 1-26.