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

Zhu Wang*

December 10, 2007

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 member issuers' profits as card payments become more efficient and convenient. At equilibrium, consumer rewards and card transaction volumes increase with interchange fees, while consumer surplus and merchant profits do not. The model provides a unified framework to evaluate credit card industry performance and policy interventions.

JEL classification: D40; L10; L40

Keywords: Oligopolistic market; Credit card industry; Interchange pricing

*Federal Reserve Bank of Kansas City, 925 Grand Boulevard, Kansas City, MO 64198. Email: zhu.wang@kc.frb.org. I thank Sujit Chakravorti, Michael Gort, Fumiko Hayashi, Robert Hunt, Thomas Jeitschko, Robert Lucas, James McAndrews, Marc Rysman, Randall Wright and seminar participants at FRB Kansas City, SUNY Buffalo, 2007 International Industrial Organization Conference and 2007 FRB New York "Money and Payments" conference 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.

1 Introduction

1.1 Motivation

Credit and debit cards have become increasingly prominent forms of payment.[1] From 1984 to 2005, the share of US consumer expenditures paid for with cards has increased from about 6 percent to 38 percent.[2] In 1995, credit and debit cards accounted for less than 20 percent of noncash payments; by 2003, they exceeded 40 percent.[3] Currently, the US card industry is a mature market: in 2001, an estimated 76 percent of US households had some type of credit card. Recent estimates suggest that among all households with income over $30,000, 92 percent hold at least one card, and the average for all households is 6.3 credit cards.[4]

With this growth has come increased scrutiny of both the benefits and costs of card use. In particular, the growth in card transactions has been paralleled by an accelerated trend of legal battles and regulatory actions against the credit card networks worldwide. At the heart of the controversy are the interchange fees (IFs) - the fees paid to card-issuing banks (issuers) when merchants accept their credit or debit cards for purchase. Although interchange fees seem to account for a small percentage of each transaction (e.g., interchange rates, varying by merchants' business type, average approximately 1.75 percent of transaction value in the US), they are significant in absolute terms. In 2005, American card issuers made $30.7 billion (about $270 per household) from interchange fees, an increase of 85 percent since 2001.[5]

[1] There 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 and account for 90% of total card purchase volume (Nilson Report 2003). They extend credit to card holders to some extent (at a minimum a credit equivalent to check float) and typically charge a proportional fee of the transaction value to merchants who accept them. In this paper, most discussion is in the context of credit cards, but may apply to all three types of cards.

[2] Data Source: Federal Reserve Bulletin (February 1986); Nilson Report (December 2005).

[3] Data Source: The 2004 Federal Reserve Payment Report (Federal Reserve Board, 2004).

[4] Data Source: Overview of Recent Developments in the Credit Card Industry (FDIC, 2005).

[5] The interchange revenue in the US alone is bigger than the size of the electronic game industry ($28.5 billion worldwide), Hollywood box office sales ($23 billion worldwide) or the music industry ($30 billion worldwide), and almost as big as venture capital investment worldwide ($31 billion). Source: Haddad (May 2006), http://aneace.blogspot.com.

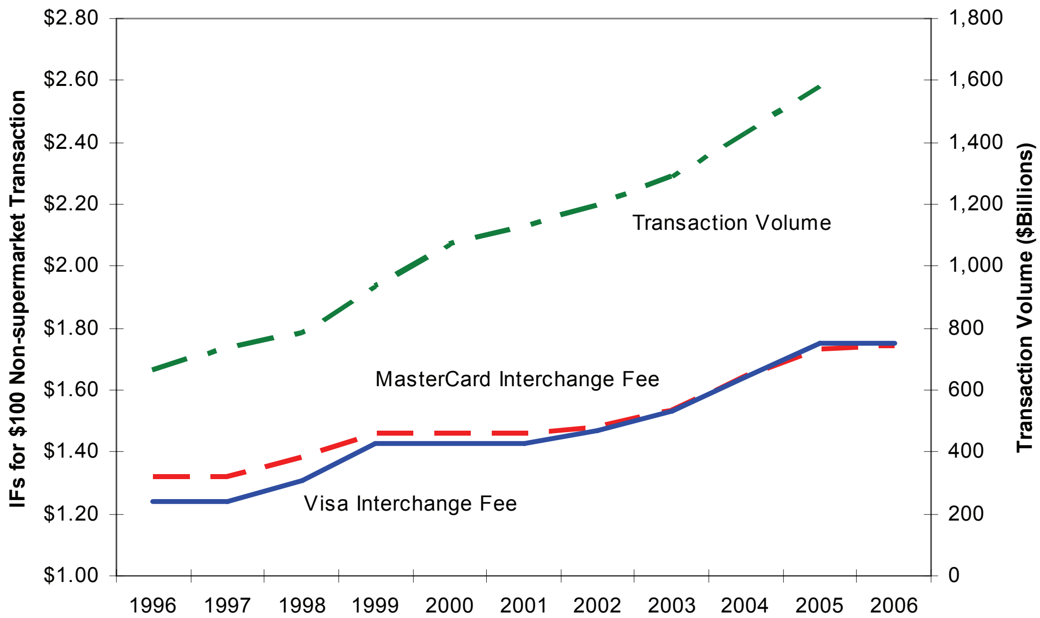

Figure 1: Credit Card Interchange Fees and Transaction Volume ($Billion) in the US

Interchange fees are set by credit card networks. The two major card networks, Visa and MasterCard, each set their interchange fees collectively for tens of thousand member financial institutions that issue and market their cards.[6] For one simple example of how interchange functions, imagine a consumer making a $100 purchase with a credit card. For that $100 item, the retailer would get approximately $98. The remaining $2, known as the merchant discount fees, gets divided up. About $1.75 would go to the card issuing bank as the interchange, and $0.25 would go to the merchant acquirer (the retailer's merchant account provider). For the credit card networks, interchange is the key of the entire enterprise, as we will show, interchange pricing increases the card

[6]Visa and MasterCard provide card services through member financial institutions (card-issuing banks and merchant-acquiring banks). They are called "four-party" systems and account for approximately 80% of the US credit card market. Amex and Discover primarily handle all card issuing and acquiring by themselves. They are called "three-party" systems and account for the remaining 20% of the market. 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.

transaction volume,[7] thus increasing interchange revenue to card issuers as well as their revenue from interest and other finance charges.[8]

Industry participants tend to agree that collectively determined interchange fees may help eliminate costly bargaining between individual card issuers and acquirers (Baxter 1983). However, they disagree on the levels of the fees. Merchants, on one hand, have become increasingly critical of interchange fees, claiming they are higher than necessary and can not be justified by the costs of card services. In fact, interchange rates in the US have been rising over the last ten years and are among the largest and fast-growing costs of doing business for many retailers (see Figure 1).[9] However, on the other hand, card networks disagree, arguing interchange fees represent an essential effort to coordinate the demands in the two-sided card market, balancing incentives to issuers to issue more cards with better rewards against the need to bring an optimum number of merchants into the card systems.

Around the world, some competition authorities and central banks have recently taken action on the interchange pricing (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 entered into an agreement with Visa in 2002 that required Visa to reduce its cross-border interchange fees through December 2007. In Australia, the Reserve Bank of Australia mandated a sizeable reduction in credit-card interchange fees in 2003, and is currently re-evaluating the regulation. Other countries, including Belgium, Israel, Poland, Portugal, Mexico, New Zealand, Netherlands, Spain and Switzerland,

[7] Throughout this paper, we use "transaction volume" to refer to the value of transactions.

[8] Note that credit cards may serve two functions: payment and credit. The payment function allows consumers to pay with cards and generate interchange revenues to card issuers. The credit function allows consumers to borrow funds and generate finance revenues to card issuers. Some consumers use both functions, but many others only use the payment function. Meanwhile, charge cards and signature debit cards offered by the credit card networks mainly provide the payment function. While our analysis in this paper focuses on the payment function and interchange revenues of cards, we may bear in mind that interchange pricing increases card transactions volumes so it may increase both interchange and finance revenues to card issuers.

[9] Data 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.

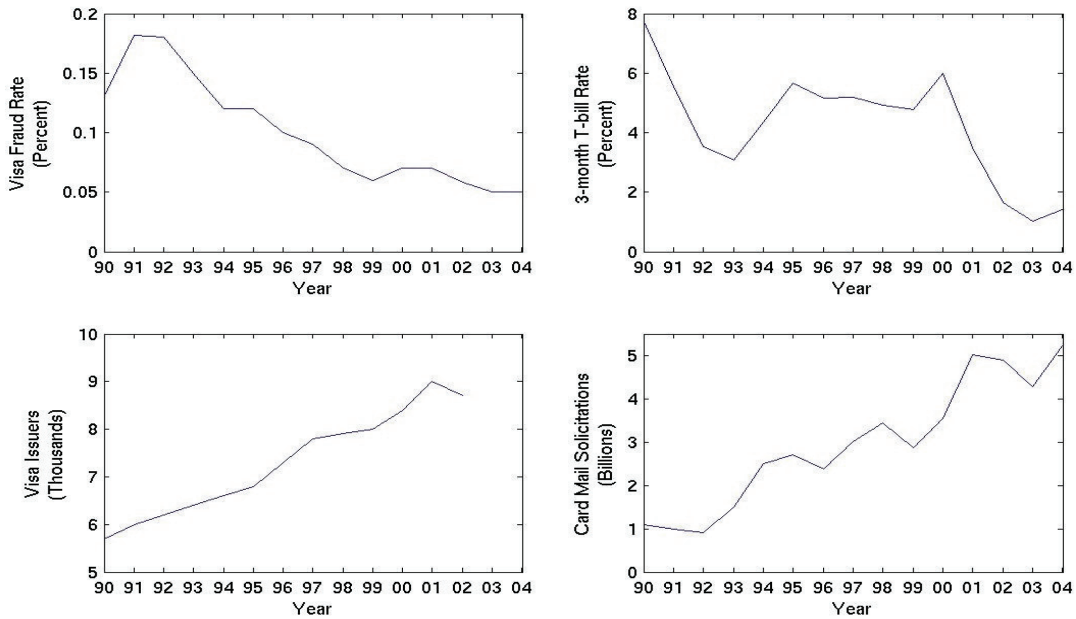

Figure 2: Credit Card Industry Trends: Costs and Competition

have made similar decisions and moves. In contrast, action on interchange fees in the US has been mainly driven by private litigation. Over recent years, merchants have launched a large number of lawsuits against Visa, MasterCard, and their members. In 2005, more than 50 antitrust cases were filed contesting interchange fees, claiming nearly $1 trillion damages. Due to the similarity of the actions, they have been consolidated into a single case which is ongoing.

The performance of the credit card industry raises challenging research questions:

- 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 rising over recent years, as shown in Figure 2)?[10]

- Given the rising interchange fees, why can't merchants refuse to accept cards? Why has card transaction volume been growing rapidly?

[10] Data 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).

- What are the causes and consequences of the increasing consumer card rewards?[11]

- What are the choices and consequences of policy interventions?

A growing literature tries to understand these issues but is far from reaching a consensus (Chakravorti 2003, Hunt 2003, Rochet 2003). Many studies emphasize the two-sided nature of payment markets and 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, 2003). 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, 2007).

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;[12] 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.

The model derives equilibrium industry dynamics consistent with empirical evidence. It suggests oligopolistic card networks are likely to collude to set monopoly interchange fees. As a result, card networks demand higher interchange fees to maximize member issuers' profits as card payments

[11] Over recent years, consumer rewards have been increasing along with the interchange fees. Rewards cards have risen from less than 25% of new card offers in 2001 to nearly 60% in 2005, and the level of rewards offered on cards has risen to as much as 5% cash back on certain purchases. In 2005, two-thirds of all cardholders owned rewards cards, up from half in 2002, and 80% of credit card transactions were made on rewards cards (Levitin 2007).

[12] Alternative payment methods may include cash, check, PIN debit cards, stored value cards, ACH and etc.

become more efficient and convenient. At equilibrium, consumer rewards and card transaction volumes increase with interchange fees, but consumer surplus and merchant profits do not.

The differences between this model and others are significant. First, this model takes a very different approach from the popular two-sided market literature. The adoption and usage externality between merchants and consumers might be crucial for an emerging card market, but has become less important as the industry matures. Therefore, we model a mature card market without network effects. Second, most previous studies (e.g., Rochet and Tirole 2002, Wright 2003, 2004, Hayashi 2006) make 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. These assumptions are relaxed in this paper. Instead, we assume competitive merchants, free entry/exit of card issuers, oligopolistic network competition, and allows for elastic demand.

The new model offers a more realistic and arguably better analytical framework, which views the credit card industry as a vertical control system with monopolistic networks on top of price taking intermediates and end users. By assuming competitive merchants and elastic consumer demand, we uncover a fundamental issue overlooked in the previous studies: Card networks have the incentive to charge high interchange fees to inflate retail prices so that they can create more demand for their services (Note for credit cards, the payment output is the total value of transactions instead of the quantity of goods sold). As the card payments become more efficient and convenient than alternatives, the card networks are able to further raise the interchange fees, inflate the value of transactions and hence extract more profits (efficiency rents) out of the system. Meanwhile, due to higher retail prices, consumer surplus and merchant profits may not improve.

It is noteworthy that our findings do not rely on excluding network effects. In fact, card networks' incentive of inflating retail prices may coexist with two-sided market externalities. McAndrews and Wang (2007) extend our analysis to an emerging card market, where a monopoly card network sets the interchange fee to internalize card adoption and usage externalities between merchants and consumers. Again, they found that the card network has a strong incentive to raise the interchange fee to inflate retail prices.

1.3 Road Map

Section 2 models the interactions among card market participants, including merchants, consumers, acquirers, issuers and card networks. The analysis suggests that a monopoly card network demands higher interchange fees to maximize member issuers' profits as card payments become more efficient or convenient. At equilibrium, consumer rewards and card transaction volumes increase with interchange fees, while consumer surplus and merchant profits do not. We then show these findings are likely to hold under oligopolistic card networks. Section 3 extends the model to evaluate policy interventions. Particularly, we show imposing a binding interchange fee ceiling allows consumers to benefit from technology progress or enhanced competition in the card industry, a result in sharp contrast with the non-intervention scenario. Meanwhile, we also point out that policy interventions may involve complicated issues and need to proceed with caution. 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.[13] The competition leads to 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., card store) charges p_e:

[13] Assuming identical merchants implies merchants always make zero profit regardless of interchange fees. However, this assumption can be easily relaxed. Under a more realistic assumption that merchants are heterogenous in costs, we show merchants' profits are indeed negatively affected by interchange fees (see Appendix B).

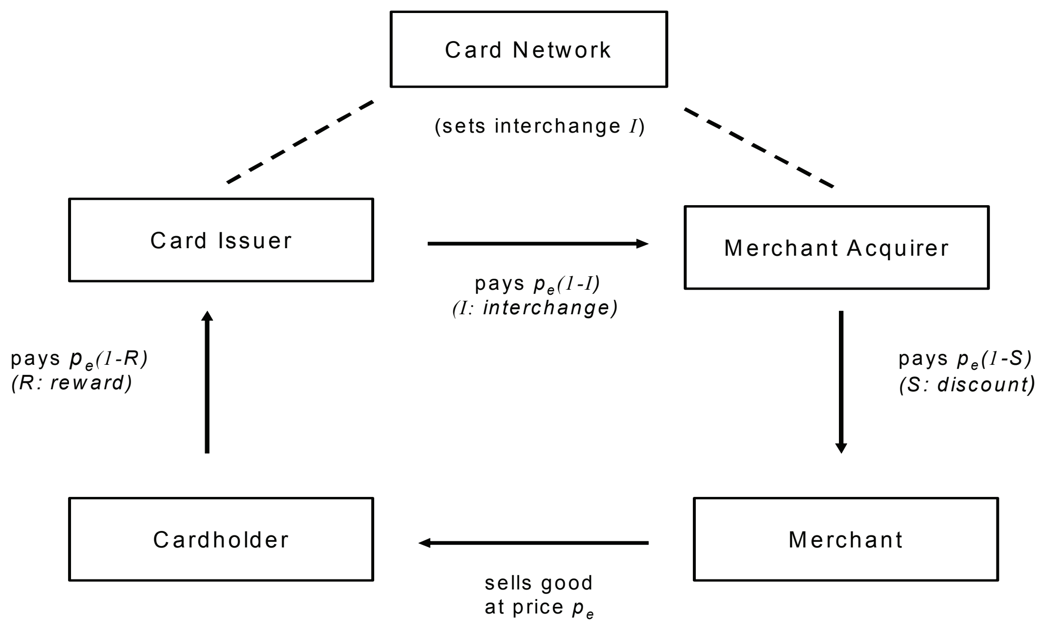

Figure 3: A Four-Party Credit Card System

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

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

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

in other words, 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 positive pricing.

Consumers: There are two types of consumers. One is cash consumers, who do not own cards and have to pay with cash. The other is card consumers, who have option to pay either with card or cash. 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.[14] Therefore, card

[14]Our analysis focuses on the payment function of credit cards, so the credit function of cards is not explicitly

consumers do not shop at cash stores if and only if

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

Meanwhile, given $p_a \leq\ p_e$, cash consumers prefer shopping at cash stores, and card consumers have no incentive to ever use cash in card stores.[15]

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:

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

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.[16] 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., $S = I$ (see Rochet and Tirole 2002 for a similar treatment).[17]

modeled. However, the reward R could be interpreted to include some benefits that consumers receive from the credit function of cards. See Chakravorti and To (2007) for related discussions.

[15] In reality, some consumers are seen using cash or PIN debit cards in stores that accept credit cards. Given the typically lower costs for merchants to accept non-credit-card payments, it is argued that these cash users are exploited. In theory, this can happen if cash stores, due to their smaller customer base, have a higher unit cost k than card stores. However, in order to focus our discussion, this issue is not explicitly explored in this paper.

[16] Although 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 in the US, and it is common for merchants to switch among acquires to get the best possible price (Evans and Schmalensee 2005).

[17] Note $C = 0$ is an innocuous assumption because C is mathematically equivalent to the network processing cost T in the following analysis. Moreover, we could instead model acquirers with heterogenous costs, but that would just duplicate our 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 for each dollar spent on card.[18] An issuer α incurs a fixed cost K each period, and faces an issuing cost V_α^β/α for its volume V_α (i.e., the value of card transactions), where $\beta > 1$.[19] 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 share their profits with the network.[20]

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

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

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

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

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

[18] Although 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) Market shares are considered less concentrated, e.g., the HHI for the Visa and MasterCard issuers was significantly below 1000 until 2003, and still below 1800 in 2005 after several major consolidations; (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 US in 2001; (5) It is easy to switch cards, and consumers do so all the time (Evans and Schmalensee 2005).

[19] The issuing costs include the costs for providing credit float, fraud protection as well as other customer services. These costs are assumed to be convex in the nominal card transaction value. Note our model does not pin down the aggregate price level, but only the price levels for those sub-markets using cards. Consequently, as nominal card transaction volumes increase, card issuers face increasing real costs to handle these volumes. For example, given that interest rates are determined at the aggregate level (hence exogenous), card issuers then incur real costs to provide credit float for the increasing nominal card transaction volumes.

[20] In 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. In addition, card networks charge their members Quarterly Service Fees 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: Frisch and Stout (2005) and Visa USA By-Laws.

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

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

Networks: Each period, a card network incurs 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 receives a share of their profits (the share is determined by bargaining between the card network and member issuers). As a result, the card network sets the interchange fee I to maximize the total profits for its member issuers.

2.2 Monopoly Outcome

The development of a card network features an enormous initial investment and a strong adoption externality. Consequently, only a small number of card networks eventually survive and they enjoy significant market power. In some countries, there is 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 maximizing its member issuers' profits Ω^m solves the following problem each period:

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

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

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

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

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

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

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

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

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

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 $\beta\gamma > 1+\gamma$;[21] the 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.[22] Therefore, the above maximization problem can be rewritten as

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

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

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

[21] The size distribution of card issuers, like firm size distribution in many other industries, is highly positively skewed. Although 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.

[22] For simplicity, we assume the consumer demand D to be a fixed function of the price p_r. Allowing the demand function to shift, e.g., by an exogenous increase of η due to income growth, would not change any of our theoretical analysis. However, it may help explain some of the increase of card transaction volumes empirically.

where

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

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

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

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

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

where A, B are defined as before. It then becomes clear that a profit maximizing card network would like to choose an optimal interchange fee I^m 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

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

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

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

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

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

Proposition 1 then characterizes the monopoly interchange fee I^m as follows.

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

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

Proof. See Appendix A. ■

Similarly, we can derive the comparative statics for the other endogenous 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: $\varepsilon \geqslant \frac{1+\tau_{c,a}}{\tau_{c,a}+\tau_{m,a}} > 1$

(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
$\tau_{c,e}$	–	±	–	–	–	–	–	–	–	0	0
T	–	–	+	–	–	–	–	–	–	0	0
K	–	–	+	±	+	–	+	–	–	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 everything else being equal, we observe the following:

- As it becomes easier for merchants to accept cards (a lower $\tau_{m,e}$), both interchange fee and consumer reward increase, but interchange fee increases more, which leads to an increase in net card price. Meanwhile, the profits and transaction volumes of individual issuers increase, the 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, and for a lower T, net card price decreases.

- As the entry barrier for card issuers declines (a lower K), both interchange fee and consumer reward increase, but consumer reward increases more, which leads to a decrease in net card price. As a result, all incumbent issuers suffer a decline in transaction volume, while large issuers see profits decrease, but small issuers see profits increase. Meanwhile, the number of issuers increases, card network profit decreases while transaction volume increases, and before-reward retail price increases. However, after-reward retail price stays the same and there is no change in card users' consumption.

- Merchants or consumers' costs of using non-card payment instruments ($\tau_{m,a}$ and $\tau_{c,a}$) have no effect on any of the endogenous variables.

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

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

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

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

Proposition 2 then 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:*

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

Proof. See Appendix A. ■

Similarly, we can derive the comparative statics for the other endogenous 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 everything 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 consumer reward, which leads to a decrease in net card price. Meanwhile, profits and transaction volumes of individual issuers decrease, the number of issuers decreases, and the card network profit and transaction volume decrease. In addition, before-and-after reward retail prices decrease and card users' consumption increases.

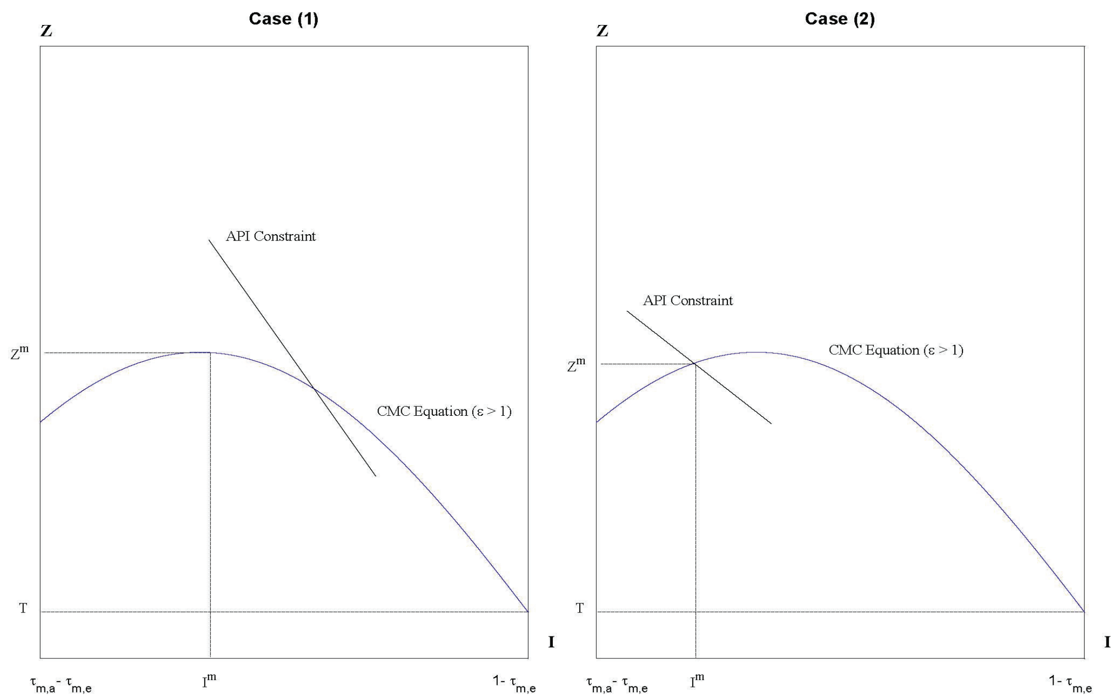

Figure 4: Monopoly Interchange Fee under Elastic Demand

- The effects of other variables are the same as Table 1.

Figure 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 (note the network profit Ω^m increases with Z) and the interchange fee $I \in [\tau_{m,a} - \tau_{m,e}, 1 - \tau_{m,e})$. 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 effect. 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 Tables 1 and 2.

2.2.2 Inelastic Demand: $\varepsilon \leqslant 1$

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

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

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

Proposition 3 then characterizes the monopoly interchange fee I^m as follows.

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

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

Proof. See Appendix A. ■

Similarly, we can derive the comparative statics for the other variables at the maximum. All the analytical results are reported in Table 3 (See Appendix A for the proofs).

Table 3. Comparative Statics: $\varepsilon \leq 1$[23]

(Signs of Partial Derivatives)

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

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

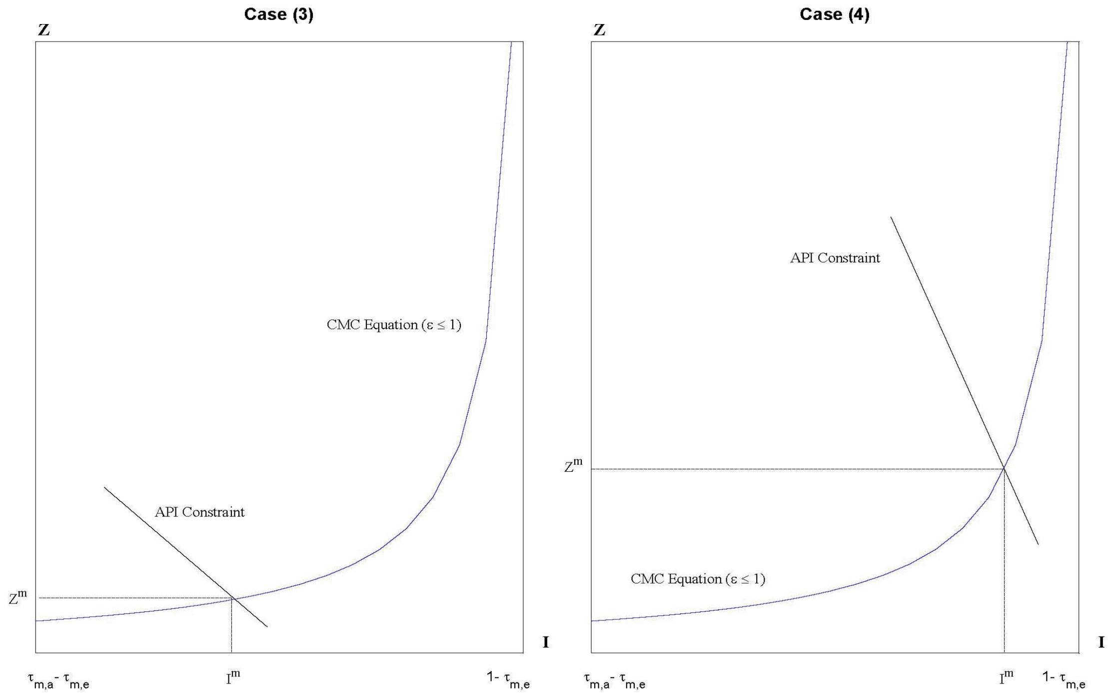

Figure 5: Monopoly Interchange Fee under Inelastic Demand

The findings suggest that everything 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 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 $I \in [\tau_{m,a} - \tau_{m,e}, 1 - \tau_{m,e})$. Therefore, the API Constraint has to bind so $\tau_{m,a}$ and $\tau_{c,a}$ 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 card payments become more efficient/convenient (a lower $\tau_{m,e}$, $\tau_{c,e}$ or T) or as the issuers' market becomes more competitive (a lower K).[24] These findings offer a consistent explanation for the empirical puzzle of rising interchange fees. Meanwhile, our analysis uncovers some major anti-competitive issues in this market. Given the market power of card networks, technology progress or enhanced competition may drive up consumer rewards and card transaction volumes, but does not necessarily improve consumer welfare.

The theory also explains other puzzling facts in the credit card market. For example, why can't merchants refuse to accept cards given the rising interchange fees? The answer is simple: As card payments become increasingly more efficient and convenient than alternative payment instruments, 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. Our analysis suggests that the answer lies on the different API constraints that card networks face in different payment environments. In an environment with higher fraud costs for cards, such as online shopping, the costs of using a non-card payment instrument are also likely to be higher, sometimes even prohibitively higher. Therefore, this allows card networks to demand higher interchange fees.

The card networks have recently been through structural changes. Allegedly not-for-profit, MasterCard changed its status to for-profit and went public in 2005, and Visa is currently exploring

[24] As mentioned, the network processing cost T is mathematically equivalent to the acquiring cost C. Hence, the recent decrease in acquiring costs in the card industry may also contribute to the increase of interchange fees.

the possibility of doing the same. However, despite the organizational changes, many industry observers feel that both networks have always been for-profit. In fact, there is little evidence that MasterCard changed its behavior after the IPO, which further confirms our assumption that card networks have always been maximizing profits.[25]

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 NO. In fact, because only a few networks coexist and they have to interact repeatedly, 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.

To formalize this idea in the credit card context, we may consider a simple duopoly model. Two card networks that produce homogenous card services have the same cost structure as specified in Section 2.1. 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}. Network i maximizes the present discounted value of its profits,

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

where δ is the discount factor.

2.3.1 Bertrand Competition

First, consider the case that the two networks engage in a simple Bertrand competition. At each period t, the networks choose their interchange fees (I_{it}, I_{jt}) simultaneously. If the two networks charge the same interchange fee $I_{it} = I_{jt} = I_t$, they share the market, that is $\Omega^i = \Omega^j = \frac{1}{2}\Omega^m(I_t)$, where $\Omega^m(I_t)$ is the monopoly network profit at the interchange level I_t. Otherwise, the lower-interchange network may capture the whole market. This is suggested by the following proposition.

[25] Note that because monopoly pricing is not by itself considered an antitrust issue, being an independent public firm may help card networks to get round the antitrust charges. According to many industry observers, this is the main reason for the networks to undertake the organizational changes (MacDonald 2006).

Proposition 4 *Everything 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, so a lower-interchange network is able to attract all the merchants and card consumers.[26] This implies that two card networks, if engaging in a Bertrand competition, should both set interchange fee at the minimum level $I = \tau_{m,a} - \tau_{m,e}$, given by Eq. (1).[27] This is the competitive equilibrium outcome.

2.3.2 Tacit Collusion

However, the outcome could be very different in a repeated game. Particularly, if each network's interchange strategy at period t is allowed to depend on the history of previous interchanges $H_t \equiv (I_{i0}, I_{j0}; ...; I_{it-1}, I_{jt-1})$, the monopoly outcome can then 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 $I^p << I^m$, 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 much lower profit for n periods. As the Folk Theorem shows, for a given n, if I^p is low and δ is large, (FT,

[26] The empirical findings of Rysman (2007) show that consumers tend to concentrate their spending on a single payment network (single-homing), but many of them maintain unused cards that allow the ability to use multiple networks (multihoming). Therefore, consumers and merchants can easily switch between networks.

[27] We reasonably assume $\tau_{m,a} > \tau_{m,e}$, so the minimum interchange fee is positive. Otherwise, consumers have to pay for the card use (i.e., the reward is negative).

FT) is a subgame perfect Nash equilibrium, and I^m can be supported at equilibrium.

This result suggests potential punishment may enforce a collusion under equilibrium, which is likely to happen in the card market for two important reasons. First, the model says 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 deviation is detected. In the card industry, the two major US credit card networks, Visa and MasterCard, share most of their member issuers and merchants (see Tables 4 and 5).[28] Therefore, there is minimal information lag in interchange pricing in the market.

Table 4: Top Eight Credit Card Issuers in 2004[29]

	Visa		MasterCard	
Issuers	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
Providen	8	10.1	11	2.5
Wells Fargo	10	7.1	9	2.8

Second, this infinitely repeated game may have multiple equilibriums, as suggested by the Folk Theorem. 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. This is consistent with the empirical observation that Visa and MasterCard have almost identical network structures and

[28] Note the overlapping membership between Visa and MasterCard may also provide direct incentive for collusion.
[29] Data Source: Frisch and Stout (2005).

market shares (see Table 5). In addition, since we assume that the two networks play FT strategies, there is little concern about renegotiation.

Table 5: Visa and MasterCard Comparison 2004[30]

	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

3 Policy and Welfare Analysis

The above analysis suggests that oligopolistic card networks are likely to collude, and set interchange fees at the monopoly level. At equilibrium, networks demand higher interchange fees to maximize member issuers' profits as card payments become more efficient or convenient. Consequently, consumer rewards and retail prices may increase, but consumer surplus may not. Furthermore, under a more realistic assumption that merchants are heterogenous in costs, we can show merchants' profits are negatively affected by interchange fees in the same way as card users' consumer surplus (see Appendix B). Based on this framework, we now proceed to discuss policy interventions.

3.1 Policy Interventions

In many countries, public authorities have chosen to regulate down interchange fees. Our theory provides a formal framework to study the implications of these policy interventions.

[30] Data Source: Frisch and Stout (2005).

3.1.1 Price Cut

As shown in Proposition 4, $\partial p_r/\partial I > 0$, which says a lower interchange fee results in a lower after-reward retail price and hence higher consumer consumption. Therefore, in order to increase consumer surplus, public authorities may have incentives to cut 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 *Everything else being equal, the CMC Equation suggests that for $I < I^m$,*

$$\begin{aligned}\partial Z/\partial I &> 0;\quad \partial \pi_\alpha/\partial I > 0;\quad \partial V_\alpha/\partial I > 0;\quad \partial N/\partial I > 0;\\ \partial \Omega/\partial I &> 0;\quad \partial p_e/\partial I > 0;\quad \partial p_r/\partial I > 0;\quad \partial D/\partial I < 0;\end{aligned}$$

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

Proof. See Appendix A. ■

Proposition 5 says everything else being equal, reducing the interchange rate below the constrained or unconstrained monopoly profit-maximizing level results in a lower net card price, lower profits and volumes for 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. The effects on consumer reward depend on the elasticity of demand. For elastic demand, consumer reward decreases, and for inelastic demand, consumer reward may either decrease or increase.

3.1.2 Price Ceiling

A one-time price cut, however, may only have temporary effects because the interchange fees can easily come back. Alternatively, public authorities may set an interchange ceiling $I^c < I^m$.[31] Given

[31] In 2003, Reserve Bank of Australia introduced a price ceiling for credit card interchange fees. At the time, the interchange fees averaged around 0.95% of the card transaction value. The regulation required that the weighted-average interchange fee for both Visa and MasterCard systems could not exceed 0.5% of the transaction value. The regulation is currently due for review, and one notable finding is that card rewards have been effectively reduced.

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	−	+	±	+	−	+	−	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 everything 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 in net card price. As a result, profits and transaction volumes of individual issuers increase, the number of issuers increases, card network profits and volumes increase, after-reward retail price decreases, and card users' consumption increases. Meanwhile, a lower $\tau_{m,e}$ results in 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 process card transactions (a lower T). Note for a lower T, consumer reward increases and net card price decreases.

- As the entry barrier for card issuers declines (a lower K), consumer reward increases, which leads to a decrease in net card price. As a result, all incumbent issuers suffer a decline in transaction volume, while large issuers see profits decrease, but small issuers see profits increase. Meanwhile, the number of issuers increases, card network profits 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 on any of the endogenous variables.

For an inelastic demand ($0 < \varepsilon \leq 1$), the analytical results are reported in Table 7 (see Appendix A for the proofs).[32]

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										

Table 7 suggests that everything 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.

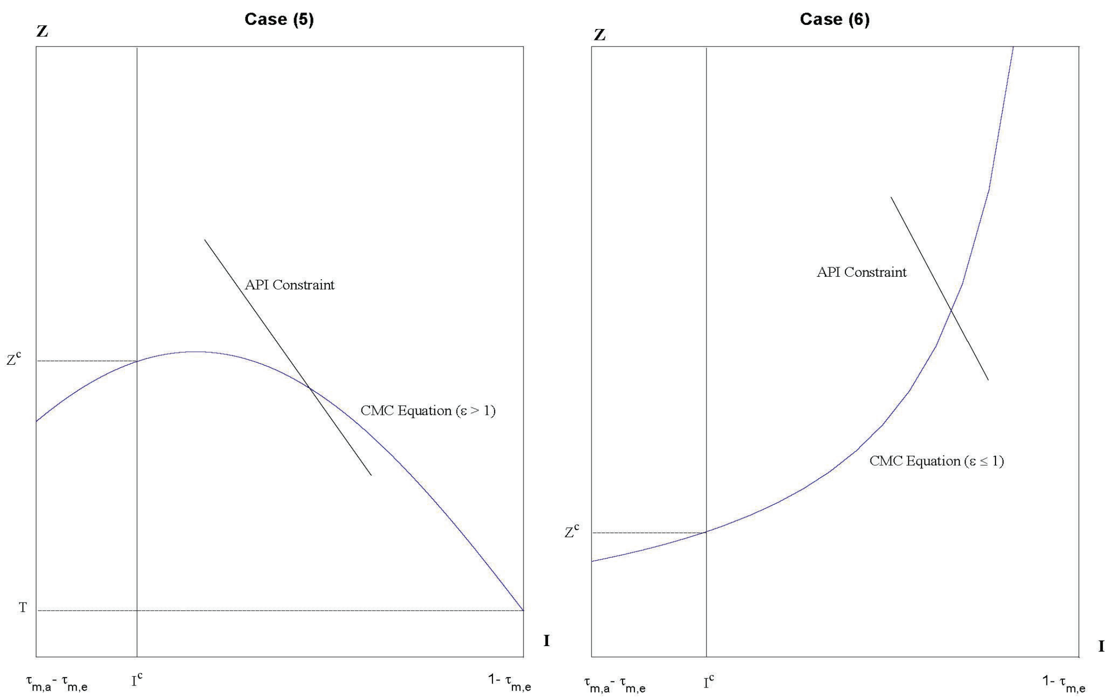

Figure 6: Interchange Fee Ceiling under Elastic/Inelastic Demand

The findings in Tables 6 and 7 suggest that a binding interchange fee 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 Tables 1, 2 and 3 for the non-intervention scenario.

Figure 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 in the 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 the interchange fee, but may affect other industry variables as described in Tables 6 and 7.

3.2 Socially Optimal Pricing

Given the structure of credit card industry, our analysis shows consumer surplus decreases with interchange fees. However, it may not be socially optimal to set the interchange fee at its minimum

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

level. In fact, the social planner aims to maximize the social surplus, the sum of producer surplus and consumer surplus. Accordingly, the social planner's problem is:

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

$$s.t. \qquad Q^* = D(\frac{k}{1-\tau_{m,e}-I}(1+\tau_{c,e}-R)) \tag{Demand of Goods}$$

$$\pi_\alpha = (\frac{\beta-1}{\beta})(\frac{\alpha}{\beta})^{\frac{1}{\beta-1}}(I-R-T)^{\frac{\beta}{\beta-1}} - K, \tag{Profit of Issuer α}$$

$$\alpha^* = \beta K^{\beta-1}(\frac{\beta}{\beta-1})^{\beta-1}(I-R-T)^{-\beta}, \tag{Marginal Issuer α^*}$$

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

$$\frac{1+\tau_{c,a}}{1-\tau_{m,a}} \geqslant \frac{1+\tau_{c,e}-R}{1-\tau_{m,e}-I}, \tag{Pricing Constraint I}$$

$$1-\tau_{m,e} > I \geq \tau_{m,a}-\tau_{m,e}, \tag{Pricing Constraint II}$$

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

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

$$TV = TD, \tag{Card Market Clearing}$$

As before, we assume that α follows a Pareto distribution $g(\alpha) = \gamma L^\gamma/(\alpha^{\gamma+1})$, the consumer demand function takes the isoelastic form $D(p_r) = \eta p_r^{-\varepsilon}$, and the pricing constraint $1-\tau_{m,e} > I \geq \tau_{m,a}-\tau_{m,e}$ are not binding.

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

$$\underset{I}{Max} \;\; \Omega^s = A(Z-T)^{\beta\gamma} + \frac{\eta}{\varepsilon - 1} p_r^{1-\varepsilon} \qquad \text{(Social Surplus)}$$

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

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

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).

Let I^s denote the socially optimal interchange fee. Note that the social surplus consists of two parts. One is card network profits, which increase with the interchange fee. The other is consumer surplus, which decreases with the interchange fee. Therefore, we expect that it requires an interchange fee I^s lower than the monopoly level I^m to maximize the social surplus. 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 the proof. ■

3.3 Further Issues

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. Meanwhile, it also provides a framework to discuss additional issues with the policy interventions.

First, our analysis has treated technological progress in payments (both cards and alternative payments) as exogenously given. Based on this, regulating down interchange fees appears to be desirable. However, it is very likely in reality that advances in card technology are also driven by intended R&D efforts by the card networks, and network profits provide major incentives and resources for these efforts. With endogenous technology progress, the social surplus calculation becomes more complicated. On one hand, regulating down interchange fees may raise 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, the extra profits in the card industry may also provide incentives for inventing and developing alternative payment products/technologies. All these endogenous and dynamic factors may make the welfare results of interchange regulation less obvious and clear.

Second, our analysis has assumed that the market costs of payment instruments reflect their social costs. In reality, this 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 payment markets.

Third, our model assumes that merchants are perfectly competitive. As shown in the model, cash and card consumers choose to shop at different stores. Because of the specialization, we then do not observe card surcharge or cash rebate (see footnote 15 for further discussions). Consequently, the no-surcharge rule imposed by the card networks plays no effective role in our analysis. 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 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, regulating interchange fee is a quick solution but might be arbitrary and less adaptable. Policy interventions may alternatively apply to the market structure (e.g., enforcing competition in the card market),[33] or competing products (e.g., encouraging technology

[33] There may be many ways to re-design the card market to enforce competition, for example, introducing multi-

progress in non-card payments). In addition, increasing public scrutiny and rising regulatory threat may also be effective policy measures (Stango 2003).

Last but not least, policy 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, including the market definition of various payment instruments, the competition between four-party systems and three-party systems, and the causes and consequents of credit card rules, just to name a few.

4 Conclusion

As credit cards become an increasingly prominent form of payment, the structure and performance of this industry has attracted intensive scrutiny. This paper presents an industry equilibrium model to better understand this market.

Our model takes a very different approach from the existing literature. First, we model a mature card market without network effects given the fact that the adoption and usage externality has become less important at this stage. Second, we relax many restrictive assumptions used in the previous studies: consumers have a fixed demand for goods irrelevant to their payment choices; merchants engage in a special form of imperfect competition; and there is no entry/exit of card issuers. Instead, we assume competitive merchants, free entry/exit of card issuers, oligopolistic network competition, and allows for elastic demand.

The new model offers a more realistic and arguably better analytical framework. By assuming competitive merchants and elastic consumer demand, we uncover a fundamental issue overlooked in the existing two-sided market literature: Card networks have the incentive to charge high interchange fees to inflate retail prices in order to create more demand for their services. As the card payments become more efficient and convenient, the card networks are able to further raise the

network cards, requiring bilateral interchange fees between issuers and merchants, or reforming the network ownership/governance structure.

interchange fees, inflate the value of transactions and hence extract more profits (efficiency rents) out of the system. Meanwhile, due to higher retail prices, consumer surplus and merchant profits may not improve.

Our analysis does not necessarily contradict with the existing two-sided market theories. In fact, card networks' incentive of inflating retail prices may coexist with two-sided market externalities. McAndrews and Wang (2007) extend our analysis to an emerging card market where card adoption and usage externalities between merchants and consumers are important. They found that, in addition to internalizing two-sided market externalities, the card network has a strong incentive to raise the interchange fee to inflate retail prices .

Based on the theoretical framework, various policy interventions are discussed in the paper. In particular, we show imposing a binding interchange fee ceiling may allow consumers to benefit from technology progress or enhanced competition in the card industry, a result in sharp contrast with the non-intervention scenario. Meanwhile, we also point out that policy interventions may involve complicated issues and need to proceed with caution.

Appendix A.

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

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

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

Therefore, Eqs. FOC and CMC imply

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

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:

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

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:

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

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

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

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

Recall

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

Therefore, we derive

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

■

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

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

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	−	+	±	+	−	+	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

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

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

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

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

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

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

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

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

or if $\varepsilon = 0$, we have

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

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 in costs, their profits are indeed negatively affected by interchange fees in the same way as the consumer surplus of card customers.

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

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

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

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

Eq. 3 then implies

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

In addition, $1-\tau_{m,e} > I$ is required for a meaningful pricing. Note all these interchange 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,

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

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

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

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 $\Psi = \varphi^{-\phi}(\frac{W\varphi}{\varphi-1})^{1+\phi-\phi\varphi}\phi J^{\phi}(\frac{1}{\phi-\frac{1}{\varphi-1}})$. 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

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

which implies the price charged in a card store is

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

Similarly, the price charged in a cash store is

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

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:

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

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

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

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

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

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

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

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

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:

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

$$s.t. \quad \frac{1+\tau_{c,a}}{1-\tau_{m,a}} \geqslant \frac{1+\tau_{c,e}-R}{1-\tau_{m,e}-I}, \qquad \text{(API Constraint)}$$

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

where

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

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 $\nu \gtreqless 1$ if and only if $\varepsilon \gtreqless 1$, 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 their 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 volumes, but may not increase consumer surplus or merchant profits. Our analysis suggests that by forcing down the 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] Chakravorti, Sujit (2003). "Theory of Credit Card Networks: A Survey of the Literature," *Review of Network Economics*, 2, 50-68.

[4] Chakravorti, Sujit and Ted To (2007). "A Theory of Credit Cards," *International Journal of Industrial Organization*, 25, (3), 583-595.

[5] Evans, David and Richard Schmalensee (2005), *Paying with Plastic: The Digital Revolution in Buying and Borrowing,* 2nd Ed., MIT Press, Cambridge, MA.

[6] Evans, David and Richard Schmalensee (2006), "The Economics of Interchange Fees and Their Regulation: An Overview," Fed Kansas City conference proceedings: *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities*? 73-120.

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

[8] Frankle, Alan (2006), "Interchange Fees in Various Countries: Commentary on Weiner and Wright," Fed Kansas City conference proceedings: *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities*? 51-61.

[9] Frankle, Alan (2007), "Towards a Competitive Card Payments Marketplace," Reserve Bank of Australia Payments System Review Conference.

[10] Frisch, Adam and Stephen Stout (2005), *Visa 101: Overview of a Payments Company.* UBS.

[11] Hayashi, Fumiko (2006), "A Puzzle of Card Payment Pricing: Why Are Merchants Still Accepting Card Payments?" *Review of Network Economics* 5: 144-174.

[12] Hayashi, Fumiko and Stuart Weiner (2006), "Interchange Fees in Australia, the UK, and the

United States: Matching Theory and Practice," *Economic Review,* Federal Reserve Bank of Kansas City, Third Quarter, 75-112.

[13] Hunt, Robert (2003), "An Introduction to the Economics of Payment Card Networks," *Review of Network Economics*, 2: 80-96.

[14] MacDonald, Duncan (2006), " What MasterCard's IPO Means for Merchant Suit," *American Banker*, June 9, 80.

[15] McAndrews, James and Zhu Wang (2007), "Microfoundations of Two-Sided Markets: The Payment Card Example," Working Paper, Federal Reserve Bank of Kansas City.

[16] Katz, Michael (2001), "Reform of Credit Card Schemes in Australia II: Commissioned Report," RBA Public Document, August.

[17] Rochet, Jean-Charles (2003), "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *Review of Network Economics* 2: 97-124.

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

[19] 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.

[20] Rochet, Jean-Charles and Jean Tirole (2006), "Externalities and Regulation in Card Payment Systems," *Review of Network Economics* 5: 1-14.

[21] Rysman, Marc (2007), "An Empirical Analysis of Payment Card Usage," *Journal of Industrial Economics* 60: 1-36.

[22] Schmalensee, Richard (2002), "Payment Systems and Interchange Fees," *Journal of Industrial Economics* 50: 103-122.

[23] 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.

[24] Stango, Victor (2003), "Strategic Responses to Regulatory Threat in the Credit Card Market," *Journal of Law and Economics*, October, 427-451.

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

[26] Weiner, Stuart and Julian Wright (2006), "Interchange Fees in Various Countries: Developments and Determinants," Fed Kansas City conference proceedings: *Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities*? 5-50.

[27] Wright, Julian. (2003), "Optimal Card Payment Systems," *European Economic Review* 47: 587-612.

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