Market Structure and Payment Card Pricing:
What Drives the Interchange?*

Zhu Wang†

June 24, 2009

Abstract

This paper provides a new theory to explain empirical puzzles regarding payment card interchange fees. Our model departs from the existing two-sided market theories by arguing that the extensive margin of card usage is less important in a mature card market. Instead, we focus on card issuer entry, elastic consumer demand and the role of card transaction value. Our 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 values increase with interchange fees, while consumer surplus and merchant profits may not. Based on the theoretical framework, we discuss pros and cons of policy interventions.

*JEL classification: D4; L1; G2

Keywords: Payment cards; Market structure; Interchange fee

---

†Economic Research Department, Federal Reserve Bank of Kansas City, 1 Memorial Drive, Kansas City, MO 64198. Email: zhu.wang@kc.frb.org. 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

As credit and debit cards become an increasingly prominent form of payments, the structure and performance of the payment card industry have attracted intensive scrutiny.\(^1\) At the heart of the controversy are interchange fees - the fees paid to card issuers when merchants accept their cards for purchase.

Interchange fees are set by card networks. 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.\(^2\) For a 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 interchange fees, and $0.25 would go to the merchant acquiring bank (the retailer’s account provider). Interchange fees serve as a key element of the card business model and generate significant revenues for card issuers.\(^3\) In 2007, the US card issuers made $42 billion revenue in interchange fees.

In recent years, merchants have become increasingly critical on interchange fees, claiming the fees are excessively high. They pointed out that, despite of falling costs in the card industry, 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

---

\(^1\)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 analysis of this paper applies to the first three types of cards, which are routed over credit card networks and account for 90% of total card purchase volume. Since our analysis focuses on the payment function but not the credit function of cards, their differences are immaterial for our purpose.

\(^2\)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 for four-party systems, but the analysis can also be applied to three-party systems.

\(^3\)Note that credit cards may serve two functions: payment and credit. The payment function allows cardholders to make transaction with cards and generate interchange revenues to card issuers. The credit function allows cardholders to borrow funds and generate finance revenues. While this paper focuses on card payment function and interchange revenues, we need to note that interchange fees may help increase card transaction values, so they also contribute to finance revenues for card issuers.
Figure 1: Credit Card Interchange Fees (IFs) and Transaction Values in the US

However, card networks disagree, arguing interchange fees serve the needs of all parties in the card system, including funding better consumer reward programs that could also benefit merchants.

In the meantime, many competition authorities and central banks around the world have taken action (Weiner and Wright, 2006). In Australia, the Reserve Bank of Australia mandated a sizeable reduction in credit card interchange fees in 2003. EU, UK, Belgium, Israel, Poland, Portugal, Mexico, New Zealand, Netherlands, Spain and Switzerland have made similar decisions and moves. In the US, interchange fees have been mainly challenged by private litigation. Since 2005, more than 50 antitrust cases have been filed by merchants contesting interchange fees.

The performance of the card industry raises following challenging questions:

- Why have interchange fees been increasing given falling costs and increased competition in the card industry?5

- Given the rising interchange fees, why can’t merchants refuse to accept cards? Why

---

5 As shown in Figure 2, card processing, borrowing and fraud costs have all declined, while the number of issuers and card solicitations have been rising over recent years. Data sources: Visa USA, Federal Reserve Board, Evans and Schmalense (2005) and Frankel (2006).
Figure 2: Credit Card Industry Trends: Costs and Competition

have total card transaction values been growing rapidly?

- What are the causes and consequences of increasing consumer card rewards?
- What are the choices and consequences of policy interventions?

In order to answer these questions, a growing literature on payment card markets has been developed recently. These models, following the pioneering work of Baxter (1983), emphasize two-sided market externalities in card payment systems. For example, Rochet and Tirole (2002) consider strategic interactions of consumers and merchants. In their model, two identical Hotelling merchants make card acceptance decisions to compete for consumers who have fixed demand for goods but heterogeneous benefits from using cards. Wright (2004) extends the framework by considering heterogeneous merchants who receive different benefits from accepting cards. These models show that merchant card acceptance and consumer card usage depend on each other, and card networks need to set card fees

---


7 Payment card systems are not the only case of such two-sided markets. Rochet and Tirole (2003) provide a detailed analysis of other examples, such as the software industry, video games, internet portals, medias, and shopping malls. In all these industries as well, the platforms may price differently to each side of the markets in order to balance the demand, while making a profit overall.
to balance the demand on the two sides of the market. However, because these analyses rely on the distribution of merchant and consumer card benefits as well as the strategic competition among merchants, the results are less conclusive in terms of evaluating card market performance and explaining stylized facts (Katz 2001, Hunt 2003, Rochet 2003, Rochet and Tirole 2006).

1.2 A Different Approach

The present paper takes a different approach. First, we consider a mature card market where the extensive margin of card usage is less important. Instead, we assume the set of card-using consumers is fixed ex ante, and consumers are homogenous in terms of benefits that they receive from using cards. Second, we relax restrictive assumptions in existing theories by assuming competitive merchants, free entry and exit of heterogenous issuers, oligopolistic card networks, and allowing for elastic consumer demand.

As a result, our model views the card industry as a vertical control system with monopolistic networks on top of price taking intermediates (issuers and acquirers) and end users (merchants and consumers). Card networks, in order to pursue their profits, set interchange fees to boost the card transaction value of existing card users (i.e., through the intensive margin of card usage). And the extent to which they can raise interchange fees and affect card transaction values depends on the cost advantage of cards over alternative payment instruments as well as the consumer demand elasticity.8

The model yields equilibrium outcomes consistent with the stylized facts. Particularly, it suggests that falling costs in the card industry could have indeed driven up interchange fees. This is because as card payments become more efficient and convenient, card networks can raise interchange fees to extract efficiency gains out of the system. At equilibrium, consumer rewards and total card transaction values increase with interchange fees, but consumer surplus and merchant profits may not improve.

Our analysis and findings depart from the existing two-sided market theories in important ways. First, we assume free entry and exit of heterogeneous issuers, each incurring a convex cost that depends on the card transaction value. This allows us to pin down a unique equilibrium interchange fee under the Tiebout sorting of card users and cash users.

8 Alternative payment instruments may include cash, check, PIN debit cards, stored value cards, automated clearing houses (ACH) and etc.
This result is in contrast to previous studies, which found the neutrality of the interchange fee under the separation of card and cash payments.\(^9\) Second, we found that the market equilibrium interchange fee is higher than or equal to the socially optimal level. Rochet and Tirole (2002) obtained similar results by considering the network externalities, which requires cash users subsidize card users under the no-surcharge rule.\(^10\) They also show when the no-surcharge rule is lifted, the interchange fee then becomes undetermined and ceases to matter (neutrality). In contrast, our analysis is based on the entry and competition of heterogeneous issuers but not the card usage externalities, so our findings hold regardless of the no-surcharge rule. Third, unlike previous studies, we allow for an elastic consumer demand and show that the consumer demand elasticity is a key parameter that determines the equilibrium card fees. We found that the market equilibrium interchange fee tends to exceed the socially optimal level if the consumer demand is very elastic. Moreover, we found that the consumer demand elasticity also affects the impact of policy interventions. Particularly, under an interchange fee ceiling, the efficiency gains in the card industry could be distributed very differently depending on the consumer demand elasticity.\(^11\)

Overall, our theory provides a new perspective that complements the existing two-sided market literature. McAndrews and Wang (2008) show that these two approaches can be combined. In a study of emerging card markets, they show that card networks exploit both intensive and extensive margins of card adoption and usage, and may charge interchange fees higher than the socially optimal.

\(^9\)Previous studies (e.g., Rochet and Tirole 2002, Wright 2003, Gans and King 2003) show that when card and cash payments are separate (e.g., when merchants are perfectly competitive or when card surcharging is available), the level of the interchange fee becomes undetermined and ceases to play any role (neutrality). Their neutrality results rely on special assumptions on the cost structure of issuers, for example, assuming homogenous issuers, each incurring zero or constant cost per transaction.

\(^10\)Note that in the case of Rochet and Tirole (2002), the optimal interchange fee for the card issuers is the highest level that is consistent with the merchants' accepting the card, so the socially optimal interchange fee is either lower than or equal to that level. This result is similar to our findings when our API (alternative payment instrument) constraint is binding. In addition, we also show when the API constraint is not binding (e.g., when the consumer demand is very elastic), the market determined interchange fee is strictly higher than the socially optimal level.

\(^11\)Previous studies (e.g., Rochet and Tirole 2002, Wright 2003) assume that each consumer has a unit demand for goods, and consumers derive an aggregate demand for the payment card services from their heterogeneous benefits of using card. However, because consumer demand is assumed completely inelastic and the distribution of the consumer heterogeneity is not explicitly specified, those studies are largely silent about how the consumer demand elasticity for goods or payments would affect the card pricing and usage.
1.3 Road Map

Section 2 sets up a model of a “four-party” card system with merchants, consumers, acquirers, issuers and card networks. The model shows that a monopoly card network demands higher interchange fees to maximize member issuers’ profits as card payments become more efficient and convenient. At equilibrium, consumer rewards and card transaction values increase with interchange fees, while consumer surplus and merchant profits do not. We show these findings may also hold under oligopolistic card networks. Section 3 extends the model to study interchange regulation and socially optimal card fees. We also discuss pros and cons of policy interventions. 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 mar-
ket.\textsuperscript{12} The competition leads to zero profit. Let $p$ and $k$ be the price and the 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.\textsuperscript{13} 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 $p_e$:

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

We require $p_e \geq p_a$ 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 $S \geq \tau_{m,a} - \tau_{m,e}$. Moreover, we require $1 - \tau_{m,e} > S$ so that $p_e$ is positive.

**Consumers:** There are two types of consumers. One is cash users, who do not own cards and have to pay with cash. The other is card users, who have option to pay either with card or cash. To use each payment instrument, consumers 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, card users receive a reward $R$ from card issuers for each dollar spent on cards.\textsuperscript{14} Therefore, card users do not shop at cash stores if and only if

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

Meanwhile, given $p_a \leq p_e$, cash users prefer shopping at cash stores, and card users have no incentive to ever use cash in card stores.\textsuperscript{15}

\textsuperscript{12}Assuming identical merchants implies that merchants 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 Appendix B, we show that under a more realistic assumption that merchants are heterogenous in costs, their profits are indeed negatively affected by interchange fees.

\textsuperscript{13}Our model is different from previous studies by assuming that payment cards charge proportional fees instead of fixed per-transaction fees. This is motivated by the fact that only cards charging proportional fees have pricing controversies in reality. However, assuming fixed per-transaction fees would not change the main findings of our analysis (See Shy and Wang 2008).

\textsuperscript{14}Although our analysis focuses on the payment but not the credit function of cards, 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.

\textsuperscript{15}In reality, some consumers may use cash in stores that accept cards. In theory, this can happen if cash stores have a higher unit cost $k$ than card stores. However, to keep our analysis focused, we do not
When making a purchase decision, card users 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 values \( TD \):

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

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. 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).

**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. An issuer \( \alpha \) incurs a fixed cost \( K \) each period, and an issuing cost \( V_\alpha^{\beta / \alpha} \) to handle its card transaction value \( V_\alpha \), where \( \beta > 1 \). Issuers are heterogenous in their operational efficiency \( \alpha \), 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.

---

16 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.

17 Note that our model does not pin down the aggregate price level of the economy, but only the price levels for those sub-markets using cards. Therefore, as nominal card transaction values increase, card issuers incur increasing real costs.

18 Assuming heterogeneous \( \alpha \) is crucial for our analysis due to two reasons. First, this allows the model to capture the observed size differences among issuers in the data. Second, if issuers were homogenous in \( \alpha \), every issuer would make zero profit under the free entry equilibrium, which implies that interchange fee would be irrelevant in the analysis.

19 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 Service Fees based on each member’s contribution to the network including the number of card issued, total transaction and sales volume. (Source: Visa USA By-Laws).
Issuer $\alpha$’s profit $\pi_\alpha$ (before sharing with the network) is determined as follows:

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

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

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

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

As a result, the total number of issuers is

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

and the total supply of card transaction values is

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

Networks: Each period, a card network incurs a variable cost $T$ per dollar for processing card transactions. In return, it charges its member issuers a processing fee $T$ to cover the variable costs and receives a share of their profits. As a result, the card network sets the interchange fee $I$ to maximize the total profits for its member issuers, which also maximizes its own profit.

### 2.2 Monopoly Outcome

A monopoly network maximizing its member issuers’ profits $\Omega^m$ solves the following problem each period:

$$
\begin{align*}
\text{Max} \quad & \Omega^m = \int_{\alpha^*}^{\infty} \pi_\alpha g(\alpha)d\alpha & (\text{Card Network Profit}) \\
\text{s.t.} \quad & \pi_\alpha = \left( \frac{\beta - 1}{\beta} \right) \left( \frac{\alpha}{\beta} \right)^{\frac{1}{\beta - 1}} (I - R - T)^{\frac{\beta}{\beta - 1}} - K, & (\text{Profit of Issuer } \alpha) \\
& \alpha^* = \beta K^{\beta - 1} \left( \frac{\beta}{\beta - 1} \right)^{\beta - 1} (I - R - T)^{-\beta}, & (\text{Marginal Issuer } \alpha^*)
\end{align*}
$$
\[1 + \tau_{c,a} \geq 1 + \tau_{c,e} - R \quad \text{(Pricing Constraint I)}\]

\[1 - \tau_{m,a} \geq 1 - \tau_{m,e} - I\]

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

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

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

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

To simplify the analysis, we assume that \(\alpha\) follows a Pareto distribution so that \(g(\alpha) = \frac{\gamma}{\alpha^{\gamma+1}}, \) where \(\gamma > 1 \) and \(\beta \gamma > 1 + \gamma; \)\(^{20}\) 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.\(^{21}\) Therefore, the above maximization problem can be rewritten as

\[\max_I \Omega^m = A(I - R - T)^{\beta \gamma} \quad \text{\scriptsize (Card Network Profit)}\]

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

\[\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} > \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I}, \quad \text{\scriptsize (Pricing Constraint I)}\]

where

\[A = KL^\gamma \beta^{-\gamma} \left(\frac{K}{\beta - 1}\right)^{(1-\beta)\gamma} \left(\frac{\gamma}{\gamma - \frac{1}{\beta - 1}} - 1\right), \quad B = \frac{L^\gamma \beta^{-\gamma} k^{\varepsilon - 1}}{\eta} \left(\frac{\gamma}{\gamma - \frac{1}{\beta - 1}} \left(\frac{K}{\beta - 1}\right)^{1+\gamma-\beta \gamma}.\]

\(^{20}\)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.

\(^{21}\)For simplicity, we assume the consumer demand \(D\) to be a fixed function of price \(p_r.\) Allowing the demand function to shift, e.g., by an exogenous increase of \(\eta\) due to income growth, would not affect our theoretical analysis, though empirically it may help explain the increase of card transaction values.
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 card markup \( \tau = \pi - \pi_v \), and further rewrite the above maximization problem as:

\[
\max_{\pi} \quad \Omega^m = A(Z - T)^{\beta \gamma} \quad \text{(Card Network Profit)}
\]

s.t. \( B(Z - T)^{\beta \gamma - 1} = (1 - \tau_{m,e} - I)^{-\varepsilon - 1}(1 + \tau_{c,e} + Z - I)^{-\varepsilon} \), \( \varepsilon > 0 \)

where \( A, B \) are defined as before. Now it has become clear that a monopoly network would like to choose an interchange fee \( I^m \) to maximize the card markup \( Z \). To fully characterize the monopoly outcome, we need to discuss two scenarios: elastic demand \( (\varepsilon > 1) \) and inelastic demand \( (\varepsilon \leq 1) \).

### 2.2.1 Elastic Demand: \( \varepsilon > 1 \)

With an elastic demand \( (\varepsilon > 1) \), the CMC equation implies an interior maximum \( Z^m \):

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

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

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

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

\[
\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \geq \frac{\varepsilon}{\varepsilon - 1} \Rightarrow \varepsilon \geq \frac{1 + \tau_{c,a}}{\tau_{c,a} + \tau_{m,a}} > 1.
\]

Proposition 1 characterizes the monopoly interchange fee \( I^m \) as follows.

**Proposition 1** Given a very elastic consumer demand (i.e., \( \varepsilon \geq \frac{1 + \tau_{c,a}}{\tau_{c,a} + \tau_{m,a}} > 1 \)), the API constraint is not binding, and the monopoly interchange fee \( I^m \) increases if card services
become less costly (i.e., $\tau_{m,e}, \tau_{c,e}, T$ or $K$ is lower). However, $I^m$ is not affected by costs of using non-card payments $\tau_{m,a}$ or $\tau_{c,a}$.

**Proof.** Equations (1)-(3) suggest that $\partial I^m / \partial T < 0$, $\partial I^m / \partial \tau_{m,e} < 0$, $\partial I^m / \partial \tau_{c,e} < 0$, $\partial I^m / \partial K < 0$, but $\partial I^m / \partial \tau_{m,a} = 0$, $\partial I^m / \partial \tau_{c,a} = 0$. ■

Similarly, we can derive comparative statics for the other endogenous variables at the monopoly maximum, including the card markup $Z^m$, the consumer reward $R^m = I^m - Z^m$, the issuer $\alpha$’s profit $\pi_\alpha$ and transaction value $V_\alpha$, the number of issuers $N$, the card network’s profit $\Omega^m$ and transaction value $TV$, the before-reward retail price $p_e$, the after-reward retail price $p_r$, and card users’ consumption $D$. All the analytical results are reported in Table 1 (See Appendix A for proofs).

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

<table>
<thead>
<tr>
<th></th>
<th>$I^m$</th>
<th>$R^m$</th>
<th>$Z^m$</th>
<th>$\pi_\alpha$</th>
<th>$V_\alpha$</th>
<th>$N$</th>
<th>$\Omega^m$</th>
<th>$TV$</th>
<th>$p_e$</th>
<th>$p_r$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{m,e}$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$\tau_{c,e}$</td>
<td>$-$</td>
<td>$\pm$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$T$</td>
<td>$-$</td>
<td>$-$</td>
<td>$+$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$K$</td>
<td>$-$</td>
<td>$-$</td>
<td>$+$</td>
<td>$\pm$</td>
<td>$+$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$\tau_{m,a}$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$\tau_{c,a}$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

Table 1 suggests that everything else being equal, we have the following findings:

- As it becomes less costly for merchants to accept cards (i.e., a lower $\tau_{m,e}$), both the interchange fee and the consumer reward increase, but the interchange fee increases more and leads to a higher card markup. Meanwhile, the profit and transaction value of individual issuers increase, the number of issuers increases, the total profit and transaction value of the card network increase, and the before-reward retail price increases. However, the after-reward retail price and card users’ consumption stay the same.
• Similar effects hold if it becomes less costly for consumers to use cards (i.e., a lower \( \tau_{c,e} \)) or it costs less for the network to provide card services (i.e., a lower \( T \) or \( K \)).

• 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:

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

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

\[
\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} < \frac{\varepsilon}{\varepsilon - 1} \Rightarrow \frac{1 + \tau_{c,a}}{\tau_{c,a} + \tau_{m,a}} > \varepsilon > 1.
\]

Proposition 2 characterizes the monopoly interchange fee \( I^m \) as follows.

**Proposition 2** Given a less elastic consumer demand (i.e., \( \frac{1 + \tau_{c,a}}{\tau_{c,a} + \tau_{m,a}} > \varepsilon > 1 \)), the API constraint is binding. The monopoly interchange fee \( I^m \) increases if card services become less costly (i.e., \( \tau_{m,e}, \tau_{c,e}, T \) or \( K \) is lower), but decreases if it costs less to use non-card payments (i.e., \( \tau_{m,a} \) or \( \tau_{c,a} \) is lower).

**Proof.** Equations (4)-(6) suggest that \( \partial I^m / \partial T < 0 \), \( \partial I^m / \partial \tau_{m,e} < 0 \), \( \partial I^m / \partial \tau_{c,e} < 0 \), \( \partial I^m / \partial K < 0 \), but \( \partial I^m / \partial \tau_{m,a} > 0 \), \( \partial I^m / \partial \tau_{c,a} > 0 \). □

<table>
<thead>
<tr>
<th>( \tau_{m,a} )</th>
<th>( I^m )</th>
<th>( R^m )</th>
<th>( Z^m )</th>
<th>( \pi_\alpha )</th>
<th>( V_\alpha )</th>
<th>( N^m )</th>
<th>( TV )</th>
<th>( p_e )</th>
<th>( p_r )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{c,a} )</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_{m,e}, \tau_{c,e}, T, K )</td>
<td>Same signs as Table 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[22\]Note that for a lower \( \tau_{c,e} \), the consumer reward can either increase or decrease; and for a lower \( T \) or \( K \), the card markup decreases. Moreover, for a lower \( K \), all incumbent issuers suffer a decline in transaction value, while large issuers’ profits decrease, but small issuers’ profits increase. Meanwhile, the number of issuers increases, the profit of the card network decreases but the total card transaction value increases.
Similarly, we can derive comparative statics for the other endogenous variables at the maximum. As shown in Table 2, we have the following findings:

- As it becomes less costly for merchants or consumers to use non-card payment instruments (i.e., a lower $\tau_{m,a}$ or $\tau_{c,a}$), the interchange fee decreases more than the consumer reward, which leads to a decrease in card markup. Meanwhile, the profit and transaction value of individual issuers decrease, the number of issuers decreases, and the total profit and transaction value of the card network decrease. In addition, the 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 provides an intuitive illustration for the analysis. In the two graphs, the CMC equation describes a concave relationship between the card markup $Z$ (Note the network profit $\Omega^m$ 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, and $\tau_{m,a}$ or $\tau_{c,a}$ has no effect. Meanwhile, a decrease of $T$ or $K$ would shift down the CMC curve to the right so $Z^m$ decreases and $I^m$ increases, which implies $R^m$ increases. In contrast, a decrease of $\tau_{m,e}$ or $\tau_{c,e}$ would shift up the CMC curve to the right so both $Z^m$ and $I^m$ increase, and $R^m$ may either increase or decrease. The results are summarized in Table 1.

In Case (2), the API constraint is binding so $\tau_{m,a}$ or $\tau_{c,a}$ does affect the interchange pricing. Particularly, at the constrained maximum $(I^m, Z^m)$, the CMC curve has a slope less than 1. As a result, a local change of $\tau_{m,a}$ or $\tau_{c,a}$ shifts the API line, but $Z^m$ changes less than $I^m$ so that $\partial R^m/\tau_{m,a} > 0$ and $\partial R^m/\tau_{c,a} > 0$. Meanwhile, a decrease of $T$ or $K$ would shift down the CMC curve along the API line so $Z^m$ decreases and $I^m$ increases, which implies $R^m$ increases. In contrast, a decrease of $\tau_{m,e}$ or $\tau_{c,e}$ would shift both the CMC curve and the API line so both $Z^m$ and $I^m$ end up increasing, and $R^m$ may either increase or decrease. The results are summarized in Table 2.

Intuitively speaking, our model views the card industry as a vertical control system with monopolistic networks on top of price taking intermediates and end users. Card networks set interchange fees to boost the card transaction value and compete against alternative payment instruments. As card payments become more efficient and convenient
Figure 4: Monopoly Interchange Fee under Elastic Demand

(e.g., $\tau_{m,e}, \tau_{c,e}, T$ or $K$ is lower), card networks then raise interchange fees to extract efficiency gains out of the system. Meanwhile, consumer rewards and total card transaction values increase with interchange fees. However, consumer surplus and merchant profits are fixed by the consumer demand elasticity or costs of using alternative payment instruments, and they do not change with interchange fees.

2.2.2 Inelastic Demand: $\varepsilon \leq 1$

With an inelastic demand ($\varepsilon \leq 1$), 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 is binding. The maximum satisfies the following conditions:

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

$$\frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} = \frac{1 + \tau_{c,e} + Z - I}{1 - \tau_{m,e} - I}.$$  \hspace{1cm} (8)$$

Proposition 3 then characterizes the monopoly interchange fee $I^m$ as follows.
Proposition 3  Given an inelastic consumer demand (i.e., \( \varepsilon \leq 1 \)), the API constraint is binding. The monopoly interchange fee \( I^m \) increases if card services become less costly (i.e., \( \tau_{m,e}, \tau_{e,e}, T \) or \( K \) is lower), but decreases if it costs less to use non-card payments (i.e., \( \tau_{m,a} \) or \( \tau_{c,a} \) is lower).

Proof. Equations (7)-(8) suggest that 
\[
\begin{align*}
\partial I^m / \partial T &< 0, \\
\partial I^m / \partial \tau_{m,e} &< 0, \\
\partial I^m / \partial \tau_{e,e} &< 0, \\
\partial I^m / \partial K &< 0, \\
\text{but } \partial I^m / \partial \tau_{m,a} &> 0, \\
\partial I^m / \partial \tau_{c,a} &> 0.
\end{align*}
\]

Similarly, we can derive comparative statics for the other endogenous variables at the maximum. As shown in Table 3, we have the following findings:

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

Table 3. Comparative Statics: \( \varepsilon \leq 1^{23} \)
(Signs of Partial Derivatives)

<table>
<thead>
<tr>
<th>( I^m )</th>
<th>( R^m )</th>
<th>( Z^m )</th>
<th>( \pi_{\alpha} )</th>
<th>( V_{\alpha} )</th>
<th>( N )</th>
<th>( \Omega )</th>
<th>( TV )</th>
<th>( p_e )</th>
<th>( p_r )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{m,a} )</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>( \tau_{c,a} )</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>( \tau_{m,e}, \tau_{e,e}, T, K )</td>
<td>Same signs as Tables 1 and 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

23Note that for \( \varepsilon = 0 \), we have \( \partial D / \partial \tau_{m,a} = \partial D / \partial \tau_{c,a} = 0 \).
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 and convenient (e.g., a lower \( \tau_{m,e}, \tau_{c,e}, T \) or \( K \)).\(^{24}\) These findings offer a consistent explanation for the puzzle of rising interchange fees. Meanwhile, we show that consumer rewards and card transaction values increase with interchange fees, but consumer welfare may not improve.\(^ {25}\)

The theory also explains other puzzles in the payment 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

\(^{24}\)As mentioned, the network processing cost \( T \) is mathematically equivalent to the acquiring cost \( C \). Hence, a decrease of acquiring costs may also contribute to the increase of interchange fees.

\(^{25}\)Our theory also suggests that the number of issuers increase with interchange fees, which is consistent with the evidence shown in Figure 2. However, our theory does not directly explain the increasing concentration among issuers. This is because we have assumed fixed parameters for the issuer size distribution. In fact, if we relax that assumption (e.g., allowing \( \beta \) to decrease), the theory then suggests that both the number and the concentration of issuers may increase.
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. This 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, which allows card networks to demand higher interchange fees.

The card networks underwent structural changes recently. Both Visa and MasterCard used to be legally organized as non-profit organizations and now they are registered as for-profit.\textsuperscript{26} However, there does not appear to be an important change in market strategy as a result of the change.\textsuperscript{27}

\subsection*{2.3 Duopoly Outcome}

So far, we have discussed the monopoly outcome in the payment card market. To extend our analysis to a more realistic setting, we may consider a duopoly card market where two card networks (e.g., Visa and MasterCard) 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 $\delta$ is the discount factor.

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.

\textsuperscript{26}MasterCard and Visa changed their status to for-profit and went public in 2005 and 2008 respectively.\textsuperscript{27} Note that because monopoly pricing is not by itself considered an antitrust issue, being an independent public firm may help card networks get around 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 after-reward retail price \( p_r \) increases with the interchange fee \( I \).

Proof. The CMC equation suggests that \( \partial p_r / \partial I > 0 \). □

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.\(^{28}\) 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 the equation of Pricing Constraint II.\(^{29}\) This is the competitive equilibrium outcome.

However, the competitive outcome does not seem to explain why interchange fees increase as costs of card services decline. Moreover, we may suspect that the monopoly outcome hold in the duopoly card market for several important reasons. First, MasterCard and Visa have duality structure, which means most card issuers are members of both networks. This may deter competition because the profits of member banks are interdependent. Second, even if MasterCard and Visa are competing, the shared ownership may still make agreement and observing defections very easy which would support collusion in a repeated game model.\(^{30}\) Hence, we consider policy analysis in the monopoly model.

3  Policy and Welfare Analysis

3.1  Interchange Regulation

In many countries, public authorities have chosen to regulate down interchange fees.\(^{31}\) Our theory provides a formal framework to study how this might affect the market outcome.

\(^{28}\)Rysman (2007) found that consumers tend to concentrate their spending on a single payment network (single-homing), but many of them maintain unused cards that allow for the ability to use multiple networks (multihoming). Therefore, consumers and merchants can easily switch between networks.

\(^{29}\)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).

\(^{30}\)Note that two-sided market models predict that competition between platforms benefits the single-homing side, typically the consumers. For that theory, it is easy to see why an oligopolistic structure favors consumers over firms but hard to understand why positive interchange existed before there was competition in the card market - something that is explained by this paper.

\(^{31}\)For example, Reserve Bank of Australia introduced a price ceiling for credit card interchange fees in 2003. 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.
3.1.1 Short-run Effects

According to our theory, lowering interchange fees has immediate effects. As shown in Proposition 4, everything else being fixed, a lower interchange fee results in a lower after-reward retail price and hence higher card users’ consumption. Recall 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, reducing the interchange fee results in a lower card markup, lower profits and transaction values for card issuers, fewer issuers, lower before-and-after-reward retail prices, and higher card users’ consumption. For an elastic demand, the consumer reward decreases; and for an inelastic demand, the consumer reward may either decrease or increase.

**Proof.** The CMC equation suggests that for \( I < I^m \), we have \( \partial Z / \partial I > 0, \partial \pi_a / \partial I > 0, \partial V_a / \partial I > 0, \partial N / \partial I > 0, \partial \Omega / \partial I > 0, \partial p_e / \partial I > 0, \partial p_r / \partial I > 0, \partial D / \partial I < 0, \) and \( \partial R / \partial I > 0 \) for \( \varepsilon > 1 \), \( \partial R / \partial I \geq 0 \) for \( \varepsilon \leq 1 \). ■

3.1.2 Long-run Effects

In the long run, maintaining an interchange ceiling \( I^c (< I^m) \) may have additional effects. Particularly, this allows card users to share efficiency gains with the card networks as card costs decline. To see this, note that the CMC equation need to be modified to introduce a binding interchange ceiling \( I^c \):

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

where \( I^c \) is now a constant. As a result, changes of environmental parameters (e.g., a lower \( \tau_{m,e}, \tau_{c,e}, T \) or \( K \)) would affect the market outcome differently from the non-intervention scenario.

Table 4 reports comparative statics of endogenous variables for an elastic demand (\( \varepsilon > 1 \)), which suggests that a binding interchange ceiling yields the following results:
Table 4. Comparative Statics: \( \varepsilon > 1 \) and \( I^c \) is binding
(Signs of Partial Derivatives)

<table>
<thead>
<tr>
<th>( I^c )</th>
<th>( R^c )</th>
<th>( Z^c )</th>
<th>( \pi_\alpha )</th>
<th>( V_\alpha )</th>
<th>( N )</th>
<th>( \Omega^c )</th>
<th>( TV )</th>
<th>( p_e )</th>
<th>( p_r )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{m,e} )</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_{c,e} )</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( T )</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( K )</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_{m,a} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \tau_{c,a} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- As it becomes less costly for merchants or consumers to use cards (i.e., a lower \( \tau_{m,e} \) or \( \tau_{c,e} \)), the consumer reward decreases, which leads to an increase in card markup. As a result, the profit and transaction value of individual issuers increase, the number of issuers increases, the total profit and transaction value of the card network increase, the 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.

- Similar effects hold if it costs less for card networks to provide card services (i.e., a lower \( T \) or \( K \)).\(^{32}\)

- 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.

Table 5. Comparative Statics: \( 0 < \varepsilon \leq 1 \) and \( I^c \) is binding
(Signs of Partial Derivatives)

<table>
<thead>
<tr>
<th>( I^c )</th>
<th>( R^c )</th>
<th>( Z^c )</th>
<th>( \pi_\alpha )</th>
<th>( V_\alpha )</th>
<th>( N )</th>
<th>( \Omega^c )</th>
<th>( TV )</th>
<th>( p_e )</th>
<th>( p_r )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{m,e} )</td>
<td>((\varepsilon &lt; 1))</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \tau_{c,e} )</td>
<td>((\varepsilon = 1))</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>( \tau_{m,a}, T, K, \tau_{m,a}, \tau_{c,a} )</td>
<td>Same signs as Table 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{32}\)Note that for a lower \( T \) or \( K \), the consumer reward increases and the card markup decreases.
Table 5 reports comparative statics of endogenous variables for an inelastic demand \((0 < \varepsilon \leq 1)\), which suggests that a binding interchange ceiling yields the following results:\(^{33}\)

- 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 effects on card pricing, output and profits opposite to the elastic demand. However, regardless of demand elasticity, a lower \(\tau_{m,e}\) always lowers 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 Table 4.

The findings in Tables 4 and 5 suggest that a binding interchange fee ceiling allows card users to share efficiency gains in the card industry. These results are in sharp contrast with what we have seen in Tables 1, 2 and 3 for the non-intervention scenarios.

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. Meanwhile, \(^{33}\) For a perfectly inelastic demand \((\varepsilon = 0)\), the results are reported in Table 6 in Appendix A.
changes in the other parameters, such as \( \tau_{m,e}, \tau_{c,e}, T, K \), shift the CMC curve along the line of \( I^c \). Therefore, the interchange fee is fixed at \( I^c \) but other industry variables are affected as described in Tables 4 and 5.

### 3.2 Socially Optimal Pricing

Given the structure of payment 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 level. In fact, the social planner aims to maximize the social surplus, which is the sum of issuers’ profits and consumer surplus. Accordingly, the social planner’s problem is

\[
\max_{I} \Omega^s = \int_{0}^{Q^*} D^{-1}(Q)dQ = \frac{k(1+\tau_{c,e}-R)}{1-\tau_{m,e}-I}Q^* + \int_{a^*}^{\infty} \pi_q g(\alpha)d\alpha,
\]

where \( Q^* = D\left(\frac{k}{1-\tau_{m,e}-I}(1+\tau_{c,e}-R)\right) \), subject to the CMC and API constraints as before.

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

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

\[
\max_{I} \Omega^s = A(Z - T)^{\beta\gamma} + \frac{\eta}{\varepsilon - 1}P_r^{1-\varepsilon}
\]

(social surplus)

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

(CMC equation)

\[
\frac{1+\tau_{c,a}}{1-\tau_{m,a}} \geq \frac{1+\tau_{c,e} + Z - I}{1-\tau_{m,e} - I},
\]

(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 planner’s 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 issuers’ profits, which increase with the interchange fee. The other is consumer surplus, which decreases with the interchange fee. Therefore, we expect that the socially optimal interchange fee \( I^s \) to be lower than or equal to the monopoly level \( I^m \), as shown in the following proposition.
Proposition 6 The socially optimal interchange fee $I^s$ is lower than or equal to the monopoly interchange fee $I^m$, i.e., $I^s \leq I^m$.

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

Note that Rochet and Tirole (2002) obtained similar results under network externalities. In their case, the optimal interchange fee for the card issuers is the highest level that is consistent with the merchants’ accepting the card, so the socially optimal interchange fee is either lower than or equal to that level. This result is similar to our findings when our API constraint is binding. In addition, our proof shows when the API constraint is not binding (e.g., when the consumer demand is very elastic), the market determined interchange fee would be strictly higher than the socially optimal level.

3.3 Further Issues

Our policy and welfare analysis offers some justification for the concerns and actions that public authorities worldwide have on the payment card interchange fees. Meanwhile, our analysis also provides a framework to discuss additional issues of policy interventions.

First, we treated technological progress in payments (both cards and non-card payments) as exogenous in the model. Based on this, regulating down interchange fees appears to be desirable. However, it is likely in reality that advances in card technology are driven by intended R&D efforts. And the overall profitability of cards may affect whether member banks support the networks to invest in new card technology. Moreover, the extra profits in the card industry may also provide incentives for inventing and developing alternative payment products and technologies. All these endogenous and dynamic factors may make the welfare results of interchange regulation less clear and obvious.

Second, our analysis assumed that the market costs of payment instruments reflect their social costs. In reality, this may not be true.³⁴ 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

³⁴For example, in the US, the production of cash is a government activity, subsidized through the federal budget. The check system is run by the Federal Reserve, which essentially forces banks to exchange checks at par - that is, to have a zero interchange fee. Therefore, these payment systems are not fully market-driven, and social costs may diverge from private costs.
would be. Therefore, learning about total social costs of various payment instruments is a prerequisite for designing and implementing good policy in payment markets.

Third, we abstracted from some potentially important issues in our analysis. For example, we assume that merchants are perfectly competitive, so we do not consider their strategic motives of accepting cards. And the no-surcharge rule does not play a role in our model because competitive merchants specialize on serving either card users or cash users. Assuming competitive merchants might be reasonable for many markets, but certainly not for all. It would be interesting to relax this assumption and investigate the implications.

Fourth, direct price regulation is not the only option or necessarily the best option for public authorities to improve market outcomes. There are other policy mixes worthy of exploring. In the case of payment card industry, regulating interchange fee is a quick solution but might be arbitrary and less adaptable. Policy makers may consider alternative approaches that target the market structure (e.g., enforcing competition between card networks)\textsuperscript{35} or competing products (e.g., encouraging technology progress in non-card payments). In addition, raising public scrutiny and 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 payment card industry. 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 payment card rules, just to name a few.

\section{Conclusion}

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

\textsuperscript{35}There may be many ways to re-design the card market to enforce competition, for example, introducing multi-network cards, requiring bilateral interchange fees between issuers and merchants, or reforming the network ownership and governance structure.
Our model takes a different approach from the existing two-sided market literature. First, we model a mature card market where the extensive margin of card usage is less important. Second, we relax many restrictive assumptions in previous studies by assuming competitive merchants, free entry and exit of heterogenous issuers, oligopolistic card networks, and allowing for elastic consumer demand.

The new model offers a more realistic and arguably better framework, which views the card industry as a vertical control system with monopolistic networks on top of price taking intermediates and end users. Card networks set interchange fees to boost the card transaction value and compete against alternative payment instruments. As card payments become more efficient and convenient, card networks then raise interchange fees to extract efficiency gains out of the system. At equilibrium, consumer rewards and total card transaction values increase with interchange fees, but consumer surplus and merchant profits may not improve. Based on the theoretical framework, the pros and cons of policy interventions are discussed.

Acknowledgments


Appendix A.

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

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

The results in column 3 then are derived by implicit differentiation. Recall that all other endogenous variables are functions of \( Z, I \) and parameters:
\[ R = I - Z, \]
\[ V_\alpha = \left( \frac{\alpha}{\beta} (Z - T) \right) \frac{1}{1}, \]
\[ N = \int g(\alpha)d\alpha = (L/\alpha^\gamma), \]
\[ TV = B(Z - T)^{\beta\gamma - 1}k^{1-\varepsilon}, \]
\[ p_\varepsilon = \frac{(1 + \tau_{\varepsilon} + Z - 1)}{(1 - \tau_{\varepsilon}m - 1)}k, \]
\[ A = KL^{\gamma \beta - \gamma - 1}\left( \frac{K^\beta}{\beta - 1} \right)^{1-\beta}(\gamma - \frac{\gamma}{\beta - 1} - 1), \]
\[ p = \frac{\eta^\varepsilon}{1 - \tau_{\varepsilon}m - 1}, \]
\[ D = \eta^\varepsilon, \]

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

**Proof.** Table 6 below reports comparative statics for the case of perfectly inelastic demand (\( \varepsilon = 0 \)). ■

### Table 6. Comparative Statics: \( \varepsilon = 0 \) and \( I^c \) is binding

<table>
<thead>
<tr>
<th>( I^c )</th>
<th>( R^c )</th>
<th>( Z^c )</th>
<th>( \pi_\alpha )</th>
<th>( V_\alpha )</th>
<th>( N )</th>
<th>( \Omega^c )</th>
<th>( TV )</th>
<th>( p_\varepsilon )</th>
<th>( p_r )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{\varepsilon,m} )</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( \tau_{\varepsilon,c} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>( T )</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>( K )</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>( \tau_{\varepsilon,m,a} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \tau_{\varepsilon,c,a} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Proof.** (Proposition 6): For \( \varepsilon > 1 \), the social planner’s problem is

\[ \text{Max } \Omega^s = A(Z - T)^{\beta\gamma} + \frac{\eta}{\varepsilon - 1} p_r^{1-\varepsilon}. \]

Consider the following two cases. First, if the API constraint is not binding (i.e., \( \varepsilon > 1 + \tau_{\varepsilon,a}/\tau_{\varepsilon,a} \tau_{m,a} > 1 \)), the monopoly’s solution 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 (i.e., \( 1 + \tau_{\varepsilon,a}/\tau_{\varepsilon,a} \tau_{m,a} > \varepsilon > 1 \)), \( (Z^m, I^m) \) 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

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

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 planner’s problem can be written as

\[
Max \quad \Omega^s = A(Z - T)^{\beta \gamma} + H - \eta \ln Q_0 - \eta \ln \eta - \eta \ln p_r.
\]

Alternatively if \( \varepsilon < 1 \), the social planner’s problem can be written as

\[
Max \quad \Omega^s = A(Z - T)^{\beta \gamma} + H + \frac{\varepsilon}{1 - \varepsilon} \eta^{1/\varepsilon} Q_0^{1-1/\varepsilon} + \eta \frac{1 - p_r^{1-\varepsilon}}{\varepsilon - 1},
\]

or if \( \varepsilon = 0 \), we have

\[
Max \quad \Omega^s = A(Z - T)^{\beta \gamma} + H - p_0 Q_0 + (p_0 - p_r) \eta,
\]

where \( p_0 \) is consumers’ highest willingness to pay for \( Q \in (Q_0, \eta) \). In each case, a proof similar to 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 heterogeneous in costs, their profits are indeed negatively affected by interchange fees in the same way as the consumer surplus of card users.

As before, we assume a continuum of merchants sell a homogenous good in a competitive market. A merchant \( \theta \) incurs a fixed cost \( W \) each period and faces an oper-
ational cost $q^\varphi_0/\theta$ for its sale $q_\theta$, where $\varphi > 1$. Merchants are heterogeneous in their operational efficiency $\theta$, which follows a Pareto distribution over the population with pdf $f(\theta) = \phi J^\theta/(\theta^{\phi+1})$, $\phi > 1$ 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 $\lambda$ and the share of cash merchants is $1 - \lambda$. The values of $p_a$, $p_e$, and $\lambda$ are endogenously determined as follows.

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

$$
\pi_{\theta,e} = \max_{q_\theta} \left(1 - \tau_{m,e} - I\right) p_e q_\theta - \frac{q^\varphi_0}{\theta} - W.
$$

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

$$
\pi_{\theta,a} = \max_{q_\theta} \left(1 - \tau_{m,a}\right) p_a q_\theta - \frac{q^\varphi_0}{\theta} - W.
$$

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

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

(9)

Note that the pricing of $p_e$ requires $p_a \leq p_e$ so that card stores do not attract cash users. Eq. (9) then implies

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

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

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

Eq. (9) 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, $1 - \tau_{m,e} > I$ so that $p_e$ is positive. Note all these interchange pricing constraints are the same as what we derived for identical merchants.
Solving the profit-maximizing problem, a merchant $\theta$ has sale $q_\theta$ and profit $\pi_\theta$ for serving card users,

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

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

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

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

Then, the total supply of goods by card stores is

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

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

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

where $\eta_e$ is related to the measure of card users. Therefore, the good market equilibrium achieved via card payments requires

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

which implies the price charged in a card store is

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

Similarly, the price charged in a cash store is

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

where $\eta_a$ is related to the measure of cash users.

At equilibrium, Eq. (9) can then pin down the share of merchants accepting cards
versus cash:

\[
\frac{\lambda}{1-\lambda} = \frac{\eta_c}{\eta_a} \left( \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} \right)^{-\epsilon} \left( \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \right)^{\epsilon}.
\]

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

\[
TD = p_c \eta_c [(1 + \tau_{c,e} - R)p_e]^{-\epsilon} = \Psi \left[ \frac{1}{\eta_a} \left( \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} \right)^{\epsilon} \left( \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \right)^{-\epsilon} + \eta_e \right]^{\epsilon - \frac{1}{\epsilon'}}, \\
(1 - \tau_{m,e} - I)^{(1-\epsilon)(\phi_{p-1})}(1 + \tau_{c,e} - R)^{\epsilon'_{p-1}}.
\]

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

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

Therefore, the card market equilibrium \( TD = TV \) implies

\[
\Theta(I - R - T)^{\beta\gamma - 1} = \left[ \eta_a \left( \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} \right)^{\epsilon} \left( \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \right)^{-\epsilon} + \eta_e \right]^{\epsilon - \frac{1}{\epsilon'}} \\
(1 - \tau_{m,e} - I)^{(1-\epsilon)(\phi_{p-1})}(1 + \tau_{c,e} - R)^{\epsilon'_{p-1}}
\]

where \( \Theta = \frac{\gamma L^\gamma \beta^{-\gamma} \left( \frac{1}{\gamma - 1} \frac{\beta}{\beta - 1} \right)^{1+\gamma-\gamma}}{\psi^{\epsilon - \frac{1}{\epsilon'}}} \).

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:

\[
\max_I \quad \Omega^m = A(I - R - T)^{\beta\gamma} \quad \text{(Card Network Profit)}
\]

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

\[
\Theta(I - R - T)^{\beta\gamma - 1} = \left[ \eta_a \left( \frac{1 + \tau_{c,e} - R}{1 - \tau_{m,e} - I} \right)^{\epsilon} \left( \frac{1 + \tau_{c,a}}{1 - \tau_{m,a}} \right)^{-\epsilon} + \eta_e \right]^{\epsilon - \frac{1}{\epsilon'}}(1 - \tau_{m,e} - I)^{(1-\epsilon)(\phi_{p-1})}(1 + \tau_{c,e} - R)^{\epsilon'_{p-1}}, \quad \text{(CMC Equation)}
\]
where

$$A = K L \beta^{-\gamma} \left( \frac{K \beta}{\beta - 1} \right)^{1 - \beta} \gamma \left( \frac{\gamma - 1}{\beta - 1} - 1 \right), \quad \Theta = \frac{\gamma}{\eta e} \left( \frac{L \gamma \beta^{-\gamma}}{\gamma - 1} \right) \left( \frac{K \beta}{\beta - 1} \right)^{1 + \gamma - \beta} \gamma \Psi \frac{\epsilon}{\phi - \epsilon}.$$

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 users’ consumer surplus. Particularly, when the API constraint is binding, the monopoly maximum satisfies the following conditions:

$$\Theta(I - R - T)^{\beta \gamma - 1} = (\eta_a + \eta_e) \frac{1 - \tau_c - e}{1 - \tau_m} (1 - \tau_m) \left( 1 - \tau_{m,e} - I \right)^{1 - \gamma - \tau} \left( 1 + \tau_c - e \right) \frac{1 + \tau_c - e}{1 - \tau_m},$$

Define $Z = I - R$ and $\nu = \frac{-e \phi e}{1 - \phi - \epsilon}$. The above conditions then can be rewritten as

$$\Theta(\eta_a + \eta_e) \frac{1 - \tau_c - e}{1 - \phi - \epsilon} (Z - T)^{\beta \gamma - 1} = (1 - \tau_m) \left( 1 - \tau_{m,e} - I \right)^{1 - \gamma - \tau} \left( 1 + \tau_c - e \right)^{-\nu},$$

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

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

Now merchants’ motive for lowering interchange fees becomes clear. Card networks, given their market power, may charge higher interchange fees to maximize card issuers’ profits as card payments become more efficient. Consequently, efficiency gains in the card industry drive up consumer rewards and card transaction values, but may not increase consumer surplus or merchant profits. Our analysis suggests that by forcing down the interchange fee, after-reward retail prices may decrease and card users’ consumption may increase. This could subsequently raise market demand for merchant sales, and hence increase merchant profits.
References


