Working Paper Series

The Economics of Two-Sided Payment Card Markets: Pricing, Adoption and Usage

WP 12-06

James McAndrews
Federal Reserve Bank of New York

Zhu Wang
Federal Reserve Bank of Richmond

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

The Economics of Two-Sided Payment Card Markets: Pricing, Adoption and Usage*

James McAndrews† and Zhu Wang‡

September 2012
Working Paper No. 12-06

Abstract

This paper provides a new theory for two-sided payment card markets. Adopting payment cards requires consumers and merchants to pay a fixed cost, but yields a lower marginal cost of making payments. Analyzing adoption and usage externalities among heterogeneous consumers and merchants, our theory derives the equilibrium card adoption and usage pattern consistent with empirical evidence. Our analysis also helps explain the card pricing puzzles, particularly the high and rising merchant (interchange) fees. Based on the theoretical framework, we discuss socially desirable payment card fees as well as the interchange fee cap regulation.

Keywords: Payment card, Two-sided market, Interchange fee
JEL Classification: L10, D40, O30

*This paper is a substantially revised version of the Federal Reserve Bank of Kansas City Research Working Paper 08-12. We thank seminar participants at the Bank of Canada, the Kansas City Fed, the New York Fed, the Richmond Fed and numerous conferences for helpful comments. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York, the Federal Reserve Bank of Richmond or the Federal Reserve System.

†Research Department, Federal Reserve Bank of New York. Email: jamie.mcandrews@ny.frb.org.
‡Research Department, Federal Reserve Bank of Richmond. Email: zhu.wang@rich.frb.org.

1 Introduction

1.1 Motivation

With payments migrating from paper to electronic forms, credit and debit cards have become an increasingly important part of the payments system in many countries. Taking the U.S. as an example, the payment share of credit and debit cards in personal consumption expenditures rose from 23 percent in 1997 to 45 percent in 2010, while the share of cash and check dropped from 70 percent to 38 percent (Figure 1). In 2010, debit cards were used in 44 billion transactions for a total value of \$1.6 trillion, and credit cards were used in 23 billion transactions for a total value of \$1.9 trillion.[1]

Along with this development has come some controversy. Merchants are critical of the fees that they pay to accept cards. These fees are often referred to as the "merchant discount," and a major portion derives from the interchange fee, paid by merchants to card issuers through merchant acquirers.[2] Merchants criticize interchange fees for being excessively high and increasing over time (Figure 2).[3] This is in sharp contrast to the falling card processing and fraud costs during the same period.[4] However, card networks disagree, arguing that interchange fees are properly set to serve the needs of all parties in the card system, including funding better consumer reward programs that could also benefit merchants.

The interchange fee controversy is a worldwide phenomenon. Regulators and com-

[1] The data are drawn from various issues of the *Nilson Report*. Payment shares not shown in Figure 1 include the automated clearing house (ACH) and some other miscellaneous types. Note that the temporary drop of the credit card share in 2008 and 2009 was due to the recession.

[2] Interchange fees are set by credit or debit card networks. Some networks provide card services through card issuers and merchant acquirers. They are called "four-party" systems, such as Visa, MasterCard and most PIN debit networks in the United States. Some other networks handle card issuing and merchant acquiring by themselves. They are called "three-party" systems, such as American Express and Discover. For a "three-party" system, interchange fees are internal transfers.

[3] Figure 2 plots the interchange fee for a \$50 non-supermarket transaction for Visa and MasterCard credit cards as well as the four largest PIN debit card networks in the United States. Data source: *American Banker* (various issues).

[4] Payment cards is primarily an information processing industry. As the information technology progresses, the relative prices of computers, communications and software have been declining rapidly, which should have driven down the card processing costs. Meanwhile, industry statistics show that the card fraud rates also have been declining steadily. For the U.S. credit card industry as a whole, the net fraud losses as a percent of total transaction volume has dropped from roughly 16 basis points in 1992 to about 7 basis points in 2009. Data source: *Nilson Report* (various issues).

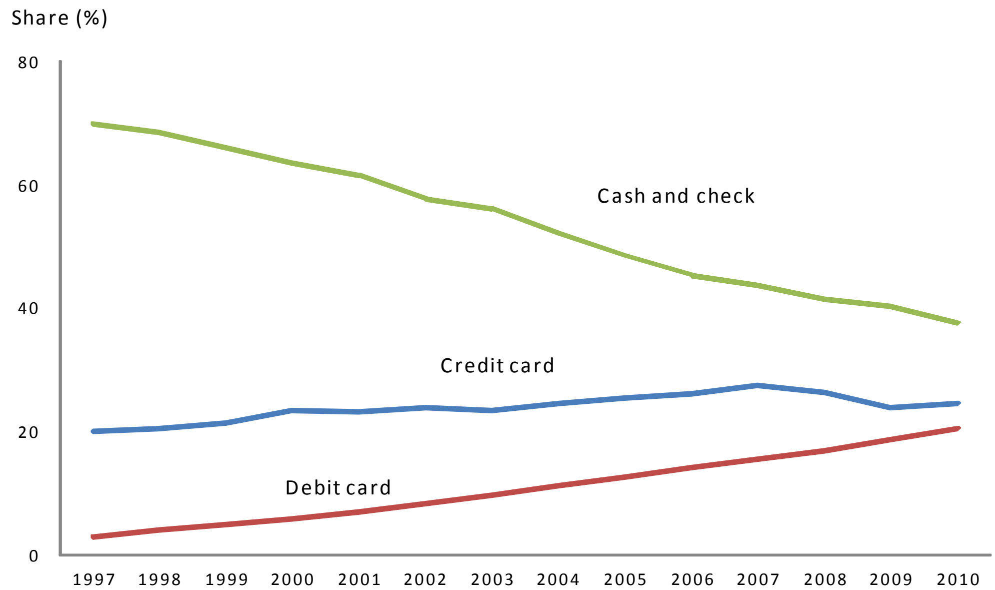


Figure 1: Payment Shares of Personal Consumption Expenditures in the U.S.

petition authorities in many countries have expressed the concern that merchant (interchange) fees inflate the cost of card acceptance without leading to proven efficiency. As a result, more than 20 countries have regulated or investigated these fees.[5] In the U.S., Congress recently passed a law that requires the Federal Reserve Board to regulate debit card interchange fees.[6] In addition, Visa and MasterCard recently agreed to a $7.25 billion settlement with U.S. retailers in a lawsuit over the fixing of credit and debit card interchange fees, which could be the largest antitrust settlement in U.S. history.[7]

To address the controversy surrounding merchant (interchange) fees, a sizable body of literature, called "two-sided market theory," has been developed in recent years.[8] Most of

[5]The countries and areas that have imposed regulations on merchant (interchange) fees include Argentina, Australia, Austria, Canada, Chile, Colombia, Denmark, the European Union, France, Israel, Mexico, Norway, Panama, Poland, Portugal, South Korea, Spain, Switzerland and Turkey. Other countries that have investigated merchant (interchange) fees include Brazil, Hungary, New Zealand, Norway, South Africa and United Kingdom (Bradford and Hayashi, 2008).

[6]The Durbin Amendment of the Dodd–Frank Act, which requires the Federal Reserve Board to regulate and oversee debit card interchange fees, took effect on October 1, 2011.

[7]Visa, MasterCard and their major issuers reached the settlement agreement with merchants in July 2012. The settlement is currently pending for final court approval.

[8]For example: Katz (2001), Schmalensee (2002), Rochet and Tirole (2002, 2006, 2011), Hunt (2003), Wright (2003, 2004), Armstrong (2006), Rysman (2007, 2009), Bolt and Chakravorti (2008), Prager et al. (2009), Bedre and Calvano (2009), Schuh et al. (2010), Rochet and Wright (2010), Wang (2010), Weyl (2010), and Shy and Wang (2011).

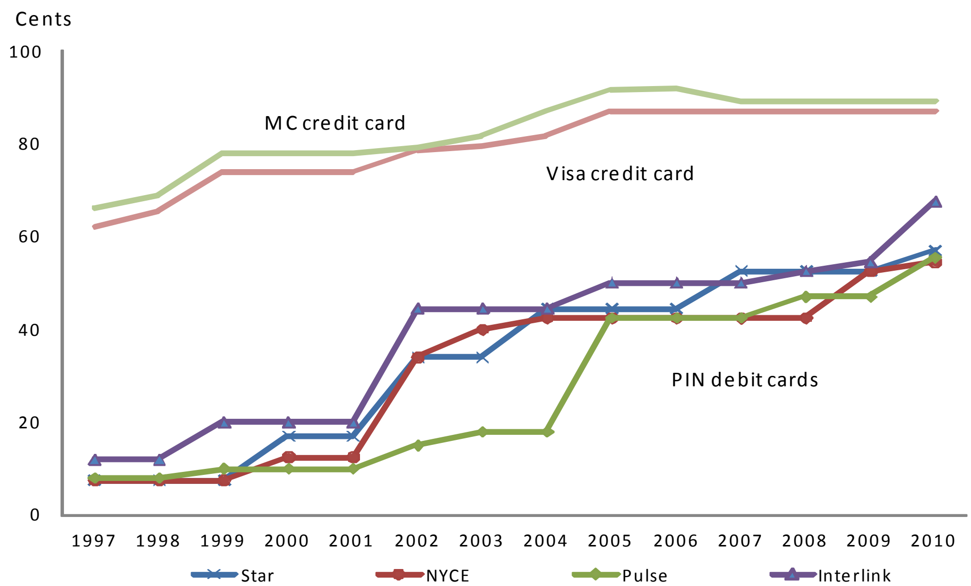


Figure 2: Interchange Fees for a $50 Transaction in the U.S.

these models, following the pioneering work of Baxter (1983), emphasize the card usage externalities between consumers and merchants. Particularly, Rochet and Tirole (2002) and Wright (2004) identify an important source of inefficiently high interchange fees: Merchants are eager to obtain a competitive edge over competitors, so they are willing to accept cards even if the fee they have to pay exceeds their direct benefit of taking cards. As a result, card consumers are over-subsidized and cards get used even for transactions where the joint merchant and consumer costs exceed the benefits. Rochet and Tirole (2011) point out that the widely used issuer cost-based cap regulation does not restore the efficiency in two-sided card markets. Instead, policymakers should consider regulating the interchange fee so that the resulting merchant discount equals the merchant direct, or convenience, benefit of accepting cards, a criterion they call "avoided-cost test" or "tourist test."[9]

While the existing two-sided market theories have shed a great deal of light on the workings of the industry, they are limited in several aspects. First, these theories typically

[9]The criterion proposed by Rochet and Tirole (2011) is adopted by the European Commission and renamed the "merchant indifference test," while some other countries, including the U.S. and Australia, adopt the issuer cost-based cap regulation.

assume imperfect competition among merchants. While this is handy for introducing merchants' business-stealing motive for accepting payment cards, it understates merchants' cost-saving motive for accepting cards. Second, these theories often do not consider card adoption decisions. However, when it comes to payment choices, consumers and merchants first have to decide whether to hold or accept payment cards (adoption), and then consumers have to decide whether to use the card in each purchase (usage). Without modeling card adoption decisions, their analyses only focus on card usage externalities but ignore the adoption externalities. Third, these theories typically assume a distribution of "convenience benefits" from using a payment card for consumers or merchants but do not identify the source of the benefits. As a result, those benefits are often referred to in nonpecuniary terms, and consumers have a fixed demand for goods invariant of their payment choices. This is not satisfactory because consumers' purchasing power is certainly affected by their payment choices. Due to these limitations, the existing theories do not fully explain the pattern of card adoption and usage, and their welfare and policy analyses are also restricted.

1.2 A new theory

In this paper, we provide a new analysis of two-sided payment card markets by introducing a number of novel features. First, we assume perfectly contestable markets for a continuum of merchants who each sell a differentiated good. This eliminates merchants' business-stealing motive of accepting cards, instead emphasizing their motive of using a more efficient payment means. Second, we model card adoption decisions under the natural assumption that adopting payment cards requires both merchants and consumers to pay a fixed cost, thereby benefiting from a lower marginal cost of making payments. Third, we introduce heterogeneity among consumers and merchants that explains the intensity of their reliance on card-based payments. This heterogeneity is modeled in terms of an income distribution for consumers and the size distribution of merchants. And we set aside the typical assumption of inelastic demand for final goods by consumers.

Our approach yields clear implications on the card adoption and usage pattern. Consider the introduction of payment cards with a high fixed but low marginal cost of use, as compared with paper payments. More affluent consumers, with higher levels of consumption and purchases, would be more willing to adopt cards than less affluent consumers. Similarly, larger merchants, or those who sell a higher-valued good, would be more likely to accept cards than other merchants. Over time, as card adoption and usage costs fall and consumer incomes rise, payment cards would then be adopted by lower-income consumers and smaller merchants. These predictions are consistent with the empirical evidence, as documented by Evans and Schmalensee (2005). In contrast, the literature that assumes an unspecified "convenience benefit" of using payment cards does not yield such straightforward empirical implications.

It has previously been pointed out (e.g., Wright, 2003) that a merchant serving both cash and card consumers would be competed out of business in a competitive environment. However, we find that in the presence of a fixed adoption cost, large merchants who serve both cash and card customers do survive the threat of entry from specialized merchants. Indeed, our equilibrium is characterized by three categories of merchants based on their size: Large merchants accept both cash and cards, and set a price that is lower than cash-only merchants. They attract both cash users and card users. Medium-size merchants are specialized. Some of them accept only cash. The others accept both cash and cards, but set a price that is higher than the competing cash-only merchants. They attract only card users, because cards are cost-saving to them overall. Finally, small merchants are all cash-only merchants. These predictions are consistent with what we observe in reality, but are not implied by the existing theories.

More important, our analysis provides a new explanation for high merchant (interchange) fees. We find that merchant (interchange) fees can be too high from a social point of view because of rent extraction by the monopoly card network. By charging a high merchant fee, the card network inflates retail prices and extracts the rents produced by replacing more costly payment means (e.g., cash or check) that would have otherwise gone to consumers in the form of lower retail prices. Moreover, our model suggests that the monopoly card network may even raise merchant (interchange) fees as card service costs decline, a pattern we observe in the data. In addition, we find payment cards, being a more efficient payment means, are not overused but rather underused because of the high merchant (interchange) fee. Meanwhile, cash users are disadvantaged by the monopoly card network. However, this is not because they have to subsidize card users as suggested by previous theories (rather, cash users are subsidized by card uses in our model). Instead, the card network sets a merchant fee that is too high so that fewer stores serve both card and cash users, hence fewer cash users get subsidized by card users.

Our paper also extends the existing literature by offering additional policy insights. We show that reducing the merchant (interchange) fee through regulation could be welfare enhancing. While the regulation may discourage consumer card adoption (because card service providers may raise consumer fees to make up for the lost interchange revenue), it increases card acceptance on the merchant side. Overall, the total card transaction value could fall, but consumers as a whole may benefit from paying less for card services. However, the interchange fee regulation may have its own limitation, and we discuss additional challenging issues that policymakers may need to consider.

1.3 Road map

In the next section we lay out our model environment and define the equilibrium. In Section 3 we characterize the equilibrium and compare the outcomes under three regimes: (1) a monopoly card network, (2) a card network run by a Ramsey regulator, and (3) a monopoly network under a merchant (interchange) fee cap regulation. In Section 4 we offer concluding remarks.

2 The model

Our model studies pricing, adoption and usage of payment devices. We first lay out a market environment in which only a paper payment device is in use. We refer to it as a cash economy for shorthand.[10] We then introduce an electronic payment device, which we refer to as a payment card. In our analysis, we treat certain market conditions, such as the variety of goods and the consumer income distribution, as exogenously given.

[10] Note that the interpretation of paper payment in our model could also include check, money order and other traditional payment means.

2.1 A cash economy

The market is composed of a continuous distribution of merchants of measure unity. Each merchant sells a distinct good characterized by two parameters: consumer preference α and unit cost μ. For ease of notation, we assume goods with the same value of α share the same value of unit cost denoted as μ_α.[11]

Cash is the sole payment device in the economy. Merchants incur a transaction cost τ_m per dollar for accepting cash, which includes handling, safekeeping and fraud expenses. The market is contestable so merchants earn zero profits. As a result, the cash price for a good α is determined as $p_{\alpha,h}$, where

$$p_{\alpha,h} = \frac{\mu_\alpha}{1-\tau_m}. \tag{1}$$

A consumer, indexed by her income I, has Cobb-Douglas preferences over a variety of goods $\alpha \in (0, \overline{\alpha})$. The index of goods α reflects the consumer's preference, which is distributed according to a cumulative distribution function $G(\alpha)$. The consumer maximizes her utility subject to the budget constraint:

$$U = Max \int_0^{\overline{\alpha}} \alpha \ln x_{\alpha,I} dG(\alpha) \qquad s.t. \qquad \int_0^{\overline{\alpha}} (1+\tau_c) p_{\alpha,h} x_{\alpha,I} dG(\alpha) = I, \tag{2}$$

where $x_{\alpha,I}$ is her quantity of demand for good α, and τ_c is the consumer's transaction cost of using cash. Solving (2) yields consumer I's demand for good α:

$$x_{\alpha,I} = \frac{\alpha I}{(1+\tau_c) p_{\alpha,h} E(\alpha)}.$$

Income I varies among consumers and is distributed according to a cumulative distribution function $F(I)$ on the support $(0, \overline{I})$. We denote $E(I)$ as the mean income and normalize the measure of consumers to be unity. At equilibrium, the supply of each good α equals its total demand:

[11] This is an innocuous assumption. In fact, the unit cost μ plays no role in our analysis. See Appendix 1 for a more general treatment where goods with the same value of α may have different values of μ.

$$x_\alpha = \int_0^{\overline{I}} x_{\alpha,I} dF(I) = \frac{\alpha E(I)}{(1+\tau_c)p_{\alpha,h}E(\alpha)}. \tag{3}$$

It follows from Eq (3) that the size of a merchant, measured by the value of sales $p_{\alpha,h}x_\alpha$, increases with the preference index, α, of the good that the merchant sells.

2.2 Introducing the payment card

We now introduce a payment innovation, referred to as a payment card. The card service is provided by a monopoly card network (or alternatively by a Ramsey regulator). The costs of providing the card service to merchants and consumers are d_m and d_c per dollar transaction, respectively. In return, the card service provider charges merchants and consumers a percentage fee f_m and f_c, respectively.[12] Figure 3 describes the transaction flow in a card system in which consumers use a payment card to pay merchants. Merchants submit charges to the card network which then bills consumers.[13]

We adopt the convention that the transaction costs for merchants and consumers to use payment cards are normalized at 0. Therefore, the avoided costs of handling cash, τ_m and τ_c, are the convenience benefits of card payments to merchants and consumers, respectively. For payment cards being a more efficient payment means, we impose the following condition:

$$\tau_m + \tau_c > d_m + d_c. \tag{4}$$

In order to adopt the payment card, a merchant and a consumer each incur a fixed cost k_m and k_c, respectively. For example, a merchant may incur a fixed cost renting card-processing terminals, while a consumer may incur a fixed cost maintaining her bank balance or credit score. Therefore, in deciding whether to accept or hold the payment card, merchants and consumers need to weigh the benefit of avoiding handling cash against their card adoption and usage costs. For merchants who decide to accept payment cards,

[12] Assuming payment cards charge percentage fees is consistent with reality. In most countries, credit cards charge fees proportional to the transaction value. In the U.S., most debit cards also charge percentage fees (Shy and Wang, 2011). In theory, card service providers can also impose lump-sum membership fees to merchants and consumers, but those are not commonly seen in the U.S. market.

[13] For simplicity, we model a "three-party" system in this paper, but our analysis and findings can similarly apply to "four-party" systems.

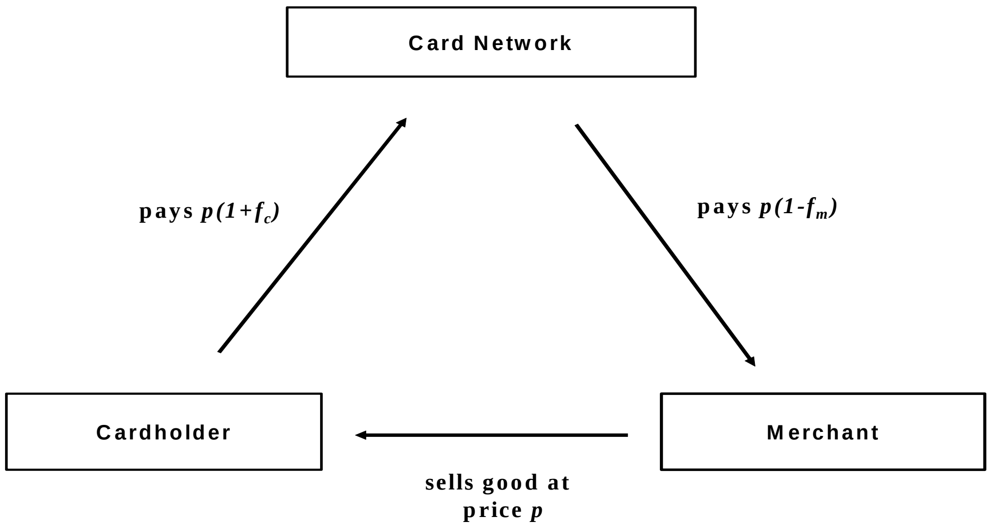


Figure 3: A Payment Card System

they still have to accept cash and charge a single price to card and cash customers.[14]

In the following analysis, we solve for a subgame perfect Nash equilibrium of the following three-stage game:

Stage I. The payment card network (alternatively a Ramsey regulator) sets the card fees f_m and f_c.

Stage II. After observing f_m and f_c, merchants and consumers decide simultaneously whether to accept and hold payment cards, respectively.

Stage III. Merchants set retail prices. Consumers decide whether to purchase, which merchants to purchase from, and what payment device to use.

Consumers and merchants each maximize their utility or profits. We assume the card network sets card fees to maximize its profit while the Ramsey regulator maximizes the consumer surplus.[15]

[14] This is referred to as "price coherence" in the literature, and has been commonly seen in reality. Price coherence may result either from network or state regulation (e.g., In the U.S., card network rules and some state laws explicitly prohibit surcharging on payment cards), or from high transaction costs for merchants to price discriminate based on payment means (e.g., In some countries where surcharging payment cards is allowed, few merchants choose to do that). Note that dropping the "price coherence" assumption would not substantially affect our analysis and findings (See Section 3.1 for more discussion).

[15] Note that consumer surplus equals total user surplus in our model since merchants earn zero profit. Maximizing total user surplus also is the criterion Rochet and Tirole (2011) used to derive the optimal

2.3 Market equilibrium

In our model, merchants and consumers each decide their own card adoption and usage, taking others' decisions as given. The interactions between the two sides of the market can easily yield multiple equilibria. For example, an equilibrium may exist at which neither merchants nor consumers adopt cards. In the following analysis, we will focus on an equilibrium with positive card adoption and usage. We show that there exist threshold values of merchant size and consumer income, above which merchants and consumers accept and hold cards. We will first construct such an equilibrium and then characterize its properties in Section 3.

2.3.1 Merchants' choices

Merchants take card fees as given and expect that consumers $I \geq I_0$ would hold cards when making their card acceptance decision. Recall that the index α indicates merchant size. It can be shown that there exist two threshold values α_1 and α_0, where $\alpha_1 > \alpha_0$, so that merchants fall into three categories: (1) Large merchants ($\alpha \geq \alpha_1$) accept both cash and cards, and charge price $p_{\alpha,d} \leq p_{\alpha,h}$ so they are patronized by both cash and card customers. (2) Intermediate merchants ($\alpha_0 \leq \alpha < \alpha_1$) specialize. Some accept both cash and cards, and charge $p_{\alpha,d}$ where $\frac{1+\tau_c}{1+f_c} p_{\alpha,h} \geq p_{\alpha,d} > p_{\alpha,h}$, so they are patronized only by card customers. The others do not accept cards and charge $p_{\alpha,h}$, so they only serve cash customers. (3) Small merchants ($\alpha < \alpha_0$) only accept cash, so all customers make purchases there using cash regardless of whether they have adopted a card.

As we will show next, the thresholds α_1 and α_0 are endogenously determined under card fees f_m and f_c. Note that because merchants who accept payment cards still have to accept cash and charge a single price to card and cash users, the consumer card fee has to be lower than the consumer cost of handling cash, i.e., $f_c \leq \tau_c$. This condition can be thought as a card pricing constraint to provide consumers "incentive at the counter" to use payment cards. As we show in the following analysis, card fees also need to satisfy

regulation based on the "avoided-cost test." Focusing on consumer surplus is legitimate as long as card network profits are not considered or weighed much less by competition authorities, which is typically true in reality.

additional pricing constraints which make card-accepting stores attractive to consumers, in other words, providing consumers "incentive at the door."

Category (1): $\alpha \geq \alpha_1$ Merchants in this category charge $p_{\alpha,d} \leq p_{\alpha,h}$ and receive revenues from both card customers ($I \geq I_0$) and cash customers ($I < I_0$):

$$p_{\alpha,d} x_{\alpha,d}^{card} = \frac{\alpha[E_{I \geq I_0}(I - k_c)]}{E(\alpha)(1 + f_c)}, \quad p_{\alpha,d} x_{\alpha,d}^{cash} = \frac{\alpha[E_{I < I_0}(I)]}{E(\alpha)(1 + \tau_c)}, \tag{5}$$

where $E_{I \geq I_0}(I) \equiv \int_{I_0}^{\overline{I}} I dF(I)$.

Contestability requires zero profit so that a merchant's total revenue equals total cost,

$$(1 - f_m) p_{\alpha,d} x_{\alpha,d}^{card} + (1 - \tau_m) p_{\alpha,d} x_{\alpha,d}^{cash} = \mu_\alpha x_{\alpha,d}^{card} + \mu_\alpha x_{\alpha,d}^{cash} + k_m. \tag{6}$$

Equations (5) and (6) pin down the price $p_{\alpha,d}$:

$$p_{\alpha,d} = \frac{\mu_\alpha \frac{\alpha[E_{I \geq I_0}(I - k_c)]}{(1+f_c)} + \mu_\alpha \frac{\alpha[E_{I < I_0}(I)]}{(1+\tau_c)}}{(1 - f_m)\frac{\alpha[E_{I \geq I_0}(I - k_c)]}{1+f_c} + (1 - \tau_m)\frac{\alpha[E_{I < I_0}(I)]}{1+\tau_c} - k_m E(\alpha)}. \tag{7}$$

Recall Eq (1) $p_{\alpha,h} = \mu_\alpha/(1 - \tau_m)$. Hence, $p_{\alpha,d} \leq p_{\alpha,h}$ implies that there exists a threshold

$$\alpha_1 = \frac{E(\alpha) k_m}{[E_{I \geq I_0}(I - k_c)](\frac{1-f_m}{1+f_c} - \frac{1-\tau_m}{1+f_c})} \quad \text{if} \quad f_m \leq \tau_m, \tag{8}$$

so that merchants whose size $\alpha \geq \alpha_1$ fall into this category. Note Eq (8) suggests that no merchant would belong to this category if $f_m > \tau_m$, because the prices they offer would not attract cash users.

Category (2): $\alpha_0 \leq \alpha < \alpha_1$ Merchants in this category specialize. For each good, there are two merchants. One accepts both cash and cards, and charges $p_{\alpha,d}$ where $\frac{1+\tau_c}{1+f_c} p_{\alpha,h} \geq p_{\alpha,d} > p_{\alpha,h}$, so it is patronized only by card customers ($I \geq I_0$). The other does not accept cards and charges $p_{\alpha,h}$, so it only serves cash customers ($I < I_0$).

A card-accepting merchant in this category receives revenues only from card customers and earns zero profit, which implies

$$p_{\alpha,d} = \frac{\mu_\alpha \frac{\alpha[E_{I\geq I_0}(I-k_c)]}{(1+f_c)}}{(1-f_m)\frac{\alpha[E_{I\geq I_0}(I-k_c)]}{1+f_c} - k_m E(\alpha)}. \tag{9}$$

Therefore, $\frac{1+\tau_c}{1+f_c}p_{\alpha,h} \geq p_{\alpha,d} > p_{\alpha,h}$ implies that merchants $\alpha_0 \leq \alpha < \alpha_1$ are in this group, where α_1 is given in Eq (8) and α_0 is determined by

$$\alpha_0 = \frac{E(\alpha)k_m}{[E_{I\geq I_0}(I-k_c)](\frac{1-f_m}{1+f_c} - \frac{1-\tau_m}{1+\tau_c})}. \tag{10}$$

Eqs (8) and (10) suggest that $\frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}$ has to hold for any merchants in the market ever to accept cards. In fact, if $\frac{1-f_m}{1+f_c} < \frac{1-\tau_m}{1+\tau_c}$, there would be no merchant in either category (1) or (2).[16]

Category (3): $\alpha < \alpha_0$ Given $\frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}$, small merchants $\alpha < \alpha_0$ are in the third category. Due to small transaction values, these merchants cannot afford paying the fixed adoption cost to accept cards. Otherwise, it would result in $p_{\alpha,d} > \frac{1+\tau_c}{1+f_c}p_{\alpha,h}$. Therefore, they only accept cash and all consumers, regardless of holding a card or not, make purchases there using cash.

2.3.2 Consumers' choices

An individual consumer takes market prices, card fees and merchants' card acceptance as given and decides whether to hold a payment card or not. Recall that merchants in the market fall into three categories according to their sizes. A consumer I, if not adopting the card, may use cash and derive utility $V_{I,h}$ from shopping at card-acceptance stores in category (1) and at cash-only stores in categories (2) and (3).

$$V_{I,h} = \int_0^{\min(\alpha_1,\overline{\alpha})} \alpha \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,h}E(\alpha)} dG(\alpha) + \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \alpha \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,d}E(\alpha)} dG(\alpha).$$

In contrast, if she holds a payment card, she can use the card to shop at card-accepting stores in categories (1) and (2), though she still needs to use cash at stores in category

[16] Because $\tau_c \geq f_c$ has to hold for any consumers to use cards, $\frac{1-f_m}{1+f_c} < \frac{1-\tau_m}{1+\tau_c}$ implies $f_m > \tau_m$.

(3). As a result, she derives utility $V_{I,d}$:

$$V_{I,d} = \int_0^{\alpha_0} \alpha \ln \frac{\alpha(I-k_c)}{(1+\tau_c)p_{\alpha,h}E(\alpha)} dG(\alpha) + \int_{\alpha_0}^{\overline{\alpha}} \alpha \ln \frac{\alpha(I-k_c)}{(1+f_c)p_{\alpha,d}E(\alpha)} dG(\alpha).$$

Therefore, consumer card adoption requires $V_{I,d} \geq V_{I,h}$, which implies

$$E_{\alpha\geq\alpha_0}(\alpha)\ln(\frac{1+\tau_c}{1+f_c}) \geq E(\alpha)\ln(\frac{I}{I-k_c}) + \int_{\alpha_0}^{\min(\alpha_1,\overline{\alpha})} \alpha\ln(\frac{p_{\alpha,d}}{p_{\alpha,h}})dG(\alpha), \tag{11}$$

where $E_{\alpha\geq\alpha_0}(\alpha) \equiv \int_{\alpha_0}^{\overline{\alpha}} \alpha dG(\alpha)$.

Equation (11) suggests that an adopter's income has to satisfy $I \geq I_0$:

$$I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{E_{\alpha\geq\alpha_0}(\alpha)/E(\alpha)}k_c}{(\frac{1+\tau_c}{1+f_c})^{E_{\alpha\geq\alpha_0}(\alpha)/E(\alpha)} - \exp(\int_{\alpha_0}^{\min(\alpha_1,\overline{\alpha})} \frac{\alpha}{E(\alpha)}\ln(\frac{p_{\alpha,d}}{p_{\alpha,h}})dG(\alpha))}, \tag{12}$$

where

$$(\frac{p_{\alpha,d}}{p_{\alpha,h}})_{\alpha_0\leq\alpha<\min(\alpha_1,\overline{\alpha})} = \frac{\frac{1-\tau_m}{1+f_c}[E_{I\geq I_0}(I-k_c)]}{\frac{1-f_m}{1+f_c}[E_{I\geq I_0}(I-k_c)] - \frac{E(\alpha)}{\alpha}k_m} \tag{13}$$

follows Eqs (1) and (9).

2.3.3 Two-sided market interactions

As shown above, the interactions between consumer and merchant card adoption are described by the threshold equations (8), (10), (12), (13).

To simplify the notations, we denote:

$$Z_1 = (\frac{1-f_m}{1+f_c} - \frac{1-\tau_m}{1+f_c}), \quad Z_0 = (\frac{1-f_m}{1+f_c} - \frac{1-\tau_m}{1+\tau_c}). \tag{14}$$

And we can summarize our findings into the following proposition.

Proposition 1 *Given card fees (f_c and f_m) that satisfy*

$$\tau_c \geq f_c \quad \textit{and} \quad \frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}, \tag{15}$$

there exist threshold values of merchant size (α_0 and α_1) and consumer income(I_0), above *which merchants and consumers accept and hold cards, where*

$$\alpha_0 = \frac{E(\alpha)k_m}{[E_{I\geq I_0}(I-k_c)]Z_0}, \tag{16}$$

$$\alpha_1 = \frac{Z_0}{Z_1}\alpha_0 \quad if \quad f_m \leq \tau_m, \tag{17}$$

$$I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{E_{\alpha\geq\alpha_0}(\alpha)/E(\alpha)}k_c}{(\frac{1+\tau_c}{1+f_c})^{E_{\alpha\geq\alpha_0}(\alpha)/E(\alpha)} - \exp(\int_{\alpha_0}^{\min(\alpha_1,\overline{\alpha})}\frac{\alpha}{E(\alpha)}\ln\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0}dG(\alpha))}. \tag{18}$$

2.3.4 Monopoly network

A monopoly card network, anticipating merchant and consumer card adoption and usage decisions in stages II and III, would set card fees (f_c, f_m) at stage I to maximize its profit. Accordingly, the card network solves the following problem:

$$\underset{f_c,f_m}{Max}\frac{E_{\alpha\geq\alpha_0}(\alpha)E_{I\geq I_0}(I-k_c)}{E(\alpha)(1+f_c)}(f_c+f_m-d_m-d_c) \tag{19}$$

$$s.t. \quad Eqs\ (15),(16),(17),(18).$$

As shown, the network's profit equals its markup times consumers' total spending on cards. In order to maximize its profit, the network internalizes the two-sided market externalities by setting its card fees (f_c, f_m) which affect the card adoption thresholds for merchants and consumers.

2.3.5 Ramsey regulator

Introducing the payment card improves consumer welfare. This can be shown in the following welfare comparison between a cash economy and a card economy.

Recall in a cash economy, an individual consumer I enjoys the utility level $U_{I,h}$:

$$U_{I,h} = \int_0^{\overline{\alpha}} \alpha \ln\frac{\alpha I}{(1+\tau_c)p_{\alpha,h}E(\alpha)}dG(\alpha).$$

After the payment card is introduced, a consumer then decides whether to adopt the card based on her income. For a card consumer $I \geq I_0$, her utility is

$$(U_{I,d})_{I\geq I_0} = \int_0^{\alpha_0} \alpha \ln \frac{\alpha(I-k_c)}{(1+\tau_c)p_{\alpha,h}E(\alpha)} dG(\alpha) + \int_{\alpha_0}^{\overline{\alpha}} \alpha \ln \frac{\alpha(I-k_c)}{(1+f_c)p_{\alpha,d}E(\alpha)} dG(\alpha),$$

while for a cash consumer $I < I_0$, her utility is

$$(U_{I,d})_{I<I_0} = \int_0^{\min(\alpha_1,\overline{\alpha})} \alpha \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,h}E(\alpha)} dG(\alpha) + \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \alpha \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,d}E(\alpha)} dG(\alpha).$$

Therefore, an individual receives different welfare gain depending on her income. For a card consumer $I \geq I_0$, her welfare gain is

$$\begin{aligned}(U_{I,d} - U_{I,h})_{I\geq I_0} &= \int_{\alpha_0}^{\min(\alpha_1,\overline{\alpha})} \alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) dG(\alpha) + \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) dG(\alpha) \\ &\quad + E_{\alpha \geqslant \alpha_0}(\alpha) \ln(\frac{1+\tau_c}{1+f_c}) + E(\alpha)\ln(\frac{I-k_c}{I}),\end{aligned} \tag{20}$$

while for a cash consumer $I < I_0$, her welfare gain is

$$(U_{I,d} - U_{I,h})_{I<I_0} = \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) dG(\alpha). \tag{21}$$

Equations (20) and (21) are intuitive: A card consumer enjoys utility gains from shopping at card-accepting stores in both categories (1) and (2), subject to the card adoption and usage costs, while a cash consumer only benefits from lower retail prices charged by merchants in category (1).

Given the above utility measures, a Ramsey regulator would choose card fees (f_c, f_m) to maximize consumer welfare gains subject to the adoption incentive constraints of merchants and consumers as well as the network balanced-budget constraint:

$$\underset{f_c,f_m}{Max} \quad \int_0^{\overline{I}} (U_{I,d} - U_{I,h}) dF(I) \tag{22}$$

$$s.t. \;\; Eqs \; (7), (13), (15), (16), (17), (18), (20), (21),$$

$$f_m + f_c \geq d_c + d_m.$$

3 Model characterizations

In this section, we characterize the equilibrium of our model. We first analyze a limiting case, in which there is no fixed cost for adopting payment cards ($k_m = k_c = 0$). This example helps illustrate the fundamental differences between a monopoly network and the Ramsey regulator. We then study a more general setting, where adoption costs are positive ($k_m > 0$, $k_c > 0$). With the general setting, we compare the market outcomes under three regimes: (1) a monopoly card network, (2) a card network run by a Ramsey regulator, and (3) a monopoly network under a merchant (interchange) fee cap regulation.

3.1 The limiting case: $k_m = k_c = 0$

In the case where no fixed cost is required ($k_m = k_c = 0$), payment cards would be adopted and used by all merchants and consumers at equilibrium ($\alpha_0 = I_0 = 0$).

Accordingly, the monopoly network's problem (19) can be simplified as follows:

$$\underset{f_c, f_m}{Max} \frac{E(I)}{(1+f_c)}(f_c + f_m - d_m - d_c) \tag{23}$$

$$s.t. \quad \tau_c \geq f_c, \quad \frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c} \ .$$

This says that the monopoly network maximizes its profit, which equals its markup times consumers' total spending on cards, provided that card fees are attractive to consumers and merchants compared with using cash.

In contrast, the Ramsey regulator's problem (22) can be simplified as follows:

$$\underset{f_c, f_m}{Max} \quad \left(\ln \frac{1-f_m}{1+f_c} - \ln \frac{1-\tau_m}{1+\tau_c} \right) E(\alpha) \tag{24}$$

$$s.t. \quad \tau_c \geq f_c, \quad \frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}, \qquad f_m + f_c \geq d_c + d_m.$$

Unlike the monopoly network, the Ramsey regulator would instead maximize consumers' total cost saving of using cards, which is in the form of lower prices of purchasing goods using cards relative to using cash.

Solving the monopoly network's problem (23) and the Ramsey regulator's problem (24), we obtain the following proposition.

Proposition 2 *In the absence of fixed card adoption cost ($k_m = k_c = 0$), a monopoly card network charges a higher merchant fee than the Ramsey regulator.*

Proof. The solution to the monopoly network's problem (23) requires $f_m = \tau_m$ and $f_c = \tau_c$. In contrast, the solution to the Ramsey regulator's problem (24) requires $f_m = d_c + d_m - \tau_c$ and $f_c = \tau_c$. Given the assumption (4) that the card is a more efficient payment means, i.e., $\tau_m + \tau_c > d_m + d_c$, the merchant fee charged by the monopoly network is higher than that charged by the Ramsey regulator. ■

The limiting case offers some simple but useful insights. First, a monopoly network charges a higher merchant fee than the Ramsey regulator. In this case, the network actually extracts all rents produced by replacing more costly cash payments that would have otherwise gone to consumers in the form of lower prices for purchasing goods. Second, a cost-based merchant fee regulation, $f_m = d_c + d_m$, does not maximize consumer welfare. This finding supports some previous studies' criticisms on issuer cost-based interchange fee regulation. However, in contrast to Rochet and Tirole (2011), we find capping the merchant fee by merchant convenience benefit of accepting cards, $f_m = \tau_m$, does not maximize consumer welfare, either. Rather, it replicates the monopoly solution in this example. Finally, in the limiting case, the regulator may want to cap the merchant fee at $f_m = d_c + d_m - \tau_c$. Under the cap, the network will then choose $f_c = \tau_c$, which coincides with the Ramsey regulator's choice. However, this is a special result. In general, as our following analysis will show, regulating the merchant fee only would not achieve the maximal consumer welfare when card adoption costs are positive.

The limiting case example also provides intuition on dropping the "price coherence" assumption in our model. Note that without the "price coherence" assumption, merchants can price discriminate their customers based on payment means. Therefore, there would be no merchants in category (1) through which card users subsidize cash users. Our limiting case above generates such a "payment separation" pattern by assuming away adoption costs (and so the cash users). The finding suggests that in an environment with "payment separation," a monopoly card network would charge a higher merchant fee than the Ramsey regulator. In our following analysis, we will show this result also holds in a more general setting where positive adoption costs and "price coherence" lead to category (1) merchants. In fact, in that environment, a monopoly card network would have additional incentives to set a higher merchant fee than the regulator. This is because the monopoly network does not care about cash users' welfare as the regulator does, so it prefers a higher merchant fee to discourage category (1) merchants and restrict the subsidy to cash users.

3.2 The general setting: $k_m > 0$, $k_c > 0$

We now switch our discussion to a more general setting, where card adoption costs are positive ($k_m > 0$, $k_c > 0$). To facilitate our discussion, we make explicit distributional assumptions in the following analysis: Assume $\alpha \in (0, 1)$ is uniformly distributed where $E(\alpha) = 1/2$; and $I \in (0, \infty)$ is exponentially distributed where $F(I) = 1 - e^{(-\lambda I)}$ and $E(I) = 1/\lambda$.[17] Note that $E_{\alpha \geq \alpha_0}(\alpha) = \frac{1-\alpha_0^2}{2}$ and $E_{I \geq I_0}(I - k_c) = e^{-\lambda I_0}(\frac{1}{\lambda} + I_0 - k_c)$.

3.2.1 Two-sided market interactions

As expected, in the presence of adoption costs, two-sided market interactions yield multiple equilibria. Assume $\alpha_1 < \bar{\alpha} = 1$. Equations (16) and (18) can be rewritten into

$$\alpha_0 = \frac{k_m}{2e^{(-\lambda I_0)}(\frac{1}{\lambda} + I_0 - k_c)Z_0}, \tag{L1}$$

$$I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} k_c}{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} - \exp(s\alpha_0^2)}, \tag{L2}$$

where $s = \ln \frac{1+f_c}{1+\tau_c} + \frac{Z_0(1+f_c)}{(1-f_m)}(\frac{Z_0}{Z_1} - 1) + \frac{Z_0^2(1+f_c)^2}{(1-f_m)^2} \ln(\frac{\frac{Z_0}{Z_1} - \frac{Z_0(1+f_c)}{(1-f_m)}}{1 - \frac{Z_0(1+f_c)}{(1-f_m)}})$ (See Appendix 2).

[17] These distributional assumptions are made for illustration purpose. We also explored some other distributional assumptions and the findings are similar.

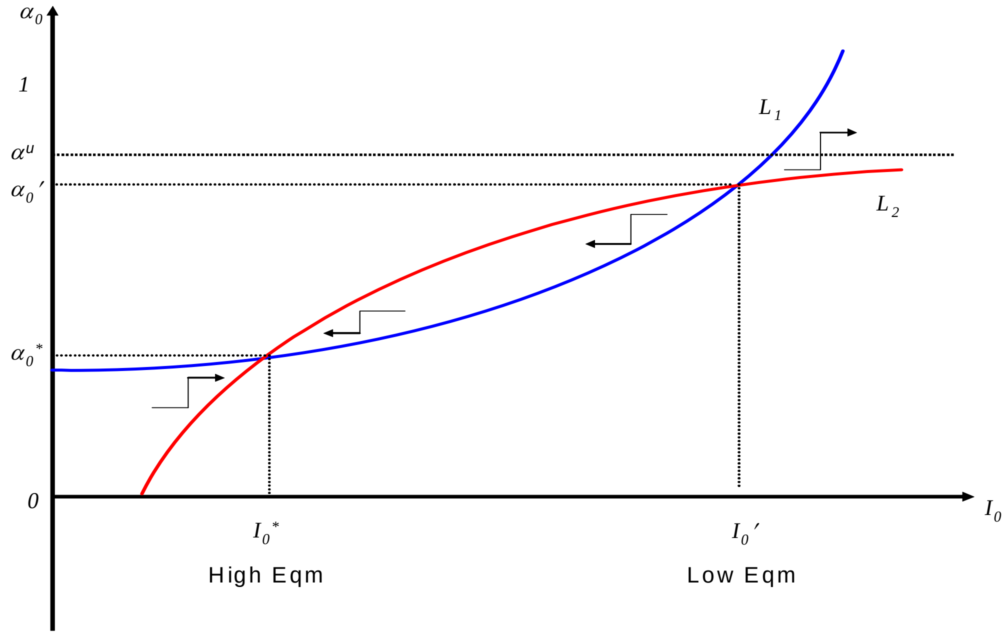


Figure 4: Interaction of Merchants and Consumers in Card Adoption

For plausible parameter values, Eq (L1) can be characterized as follows:

$$\alpha_0|_{I_0 \longrightarrow 0} \longrightarrow \frac{k_m}{2(\frac{1}{\lambda} - k_c)Z_0} > 0, \quad \alpha_0|_{I_0 \rightarrow \infty} \longrightarrow \infty,$$

$$\frac{d\alpha_0}{dI_0} > 0, \quad \frac{d^2\alpha_0}{dI_0^2} > 0.$$

Characterizing Eq (L2) yields

$$I_0|_{\alpha_0 \rightarrow 0} \rightarrow \frac{(\frac{1+\tau_c}{1+f_c})k_c}{(\frac{1+\tau_c}{1+f_c}) - 1} > 0, \quad \alpha_0|_{I_0 \longrightarrow \infty} \longrightarrow \alpha^u = (\frac{\ln(\frac{1+\tau_c}{1+f_c})}{\ln(\frac{1+\tau_c}{1+f_c}) + s})^{1/2} < 1,$$

$$\frac{d\alpha_0}{dI_0} > 0, \quad \frac{d^2\alpha_0}{dI_0^2} < 0.$$

Based on the above characterizations, Figure 4 illustrates the interactions between merchant and consumer card adoption and the resulting equilibria. For a given pair of card fees (f_m, f_c), there exist two equilibria with positive levels of card adoption (Note that no adoption is an equilibrium as well): a high-adoption equilibrium (I_0^*,α_0^*) and a low-adoption equilibrium (I_0',α_0'). The high equilibrium is stable, while the low equilibrium is not. We will then focus on the stable equilibrium in our following analysis.[18]

3.2.2 Three regimes

As shown above, there is a unique stable equilibrium associated with a given fee pair (f_m, f_c) and model parameters ($d_c, d_m, k_c, k_m, 1/\lambda, \tau_c, \tau_m$). Presumably, the equilibrium market outcome depends on how the card network is run. This provides a natural framework to compare three regimes: (1) a monopoly card network, (2) a card network run by a Ramsey regulator, and (3) a monopoly network under a merchant (interchange) fee cap regulation.

Moreover, our framework also allows us to study how payment markets (both cash and card) evolve, which can be viewed as a series of comparative statics of the market equilibrium. Due to the complexity of the problem, our analysis will rely on numerical simulations in addition to analytical results. In the benchmark simulation, we first show how payment markets evolve as the card service costs $d_c + d_m$ fall.[19] We then adjust other model parameters (e.g., $k_c, k_m, 1/\lambda, \tau_c, \tau_m$) to see their effects.

Monopoly network The card network solves its profit maximizing problem (19) (See Appendix 3 for the explicit formulation under our distributional assumptions). The simulation results, shown in Fig. A1 and A2 in the Appendix, are summarized as follows.

In the benchmark simulation (Case 1), we set $\tau_m = 0.05$, $\tau_c = 0.05$, $k_m = 160$, $k_c = 160$, $E(I) = 1/\lambda = 10,000$.[20] The results show that as the card service costs $d_m + d_c$ fall, the merchant card fee f_m increases, the consumer card fee f_c decreases, and the card pricing markup $f_c + f_m - d_c - d_m$ increases. More merchants accept cards (α_0 decreases), more consumers use cards (I_0 decreases), but fewer card-accepting merchants are patronized by cash users (α_1 increases). As a result, card users gain more surplus, cash users gain less surplus, and the total consumer welfare increases. In terms of expenditures,

[18] The idea of using stability to select an equilibrium goes back to Cournot (1838), who provided a dynamic adjustment argument by imagining sequential play by each agent myopically best-responding to all rivals. This is referred to as best-reply dynamics, and when the process converges, the solution is termed stable (Fudenberg and Tirole, 1991).

[19] Note that the effects of d_c and d_m are not separable from each other in our model.

[20] These parameter values are chosen for illustration purpose. We also tried other parameter values and the findings are similar.

merchants and consumers spend more on total card adoption costs and usage fees to pay for an increasing card transaction value, but the card-service-spending-to-card-sales ratio decreases. Meanwhile, the cash transaction value declines and the cash-cost-to-sales ratio is fixed at $\tau_m + \tau_c$. In total, society pays less for payment services, and the total-payment-spending-to-total-sales ratio decreases.

We then study the effects of jointly changing k_m and k_c. In the simulation of Case 2, we reduce both k_m and k_c by an equal proportion relative to the benchmark case, setting $k'_m = k'_c = 128$ so $k'_m/k_m = k'_c/k_c = 0.8$. The results show, compared with the benchmark case, the network now charges higher card fees f_m and f_c, so the card markup is higher. In spite of higher card fees, the lower adoption costs induce more merchants and consumers to adopt cards. Consequently, card users enjoy more welfare gains compared with the benchmark case. In terms of expenditures, merchants and consumers now spend more on total card adoption costs and usage fees to pay for a higher card transaction value, and the card-payment-spending-to-card-sales ratio is higher than the benchmark case, while cash costs and cash transaction value are lower. The total-payment-spending-to-sales ratio could be higher or lower than the benchmark case depending on the value of card service costs $d_m + d_c$.

In Case 3, we adjust the benchmark simulation by raising the mean consumer income $E(I) = 1/\lambda' = 12500$ so that $\lambda'/\lambda = 0.8$. As it turns out, the simulation results coincide with Case 2 in terms of card pricing, adoption and consumer welfare. Meanwhile, a higher income leads to more spending on card and cash payment services, and more card and cash transaction value than Case 2, while the payment-spending-to-sales ratios are the same. In fact, it is not surprising that the increase of consumer mean income $(1/\lambda)$ and the joint decline of adoption costs (k_c, k_m) should have equivalent effects on card pricing and adoption. A formal proof can be established as follows.

Proposition 3 *The increase of consumer mean income $(1/\lambda)$ and the joint decline of adoption costs (k_c, k_m) have equivalent effects on card pricing and adoption. Formally, the network profit maximization yields the same card fees and adoption rates under the parameter values $(1/\lambda, k_c, k_m)$ and $(\theta/\lambda, \theta k_c, \theta k_m)$.*

Proof. See Appendix 4 for the proof. ■

In Case 4, we adjust the card adoption costs between merchants and consumers relative to the benchmark (Case 1) by setting $k_m = 300, k_c = 20$. As the merchants now have a higher adoption cost, merchants are charged for a lower fee compared with the benchmark case (to partly offset the higher adoption cost), but the consumer card fee becomes higher. Because of the higher adoption cost, fewer merchants accept cards (α_0 increases), but more card-accepting merchants attract cash users (α_1 decreases). Meanwhile, because consumers face a lower adoption cost, their card adoption rate becomes higher. As an overall result, both card and cash users gain a higher surplus. The finding that merchants are charged a lower card fee to partly offset their higher adoption cost is consistent with the early experience of the card industry: When PIN debit cards were initially introduced in the U.S. market, merchants were indeed charged a negative fee for accepting them. One main explanation was that merchants needed to be subsidized to install the PIN compatible terminals.

In Case 5, we adjust the card usage benefits between merchants and consumers relative to the benchmark (Case 1) by setting $\tau_m = 0.08, \tau_c = 0.02$. As the merchants now have a higher card acceptance benefit, they pay a higher fee compared with the benchmark case. Meanwhile, consumers are charged a lower fee to accommodate their lower card usage benefit. In fact, the consumer fee becomes negative in this case, which means consumers actually receive rewards each time they use payment cards.

Ramsey regulator The Ramsey regulator maximizes the consumer surplus (22) (See Appendix 5 for the explicit formulation under our distributional assumptions). Our simulation, shown in Fig. A3 in the Appendix, compares the outcomes between the monopoly network and the Ramsey regulator using our benchmark parameter values (Case 1).

Several important differences stand out by comparing the monopoly network and the Ramsey regulator. With the same model parameter values, the monopoly network charges a much higher merchant fee f_m than the Ramsey regulator. As a result, both merchant and consumer card adoption are lower (higher α_0 and I_0), and fewer card-accepting merchants serve cash customers (α_1 is higher). These lead to lower welfare gains for both card users and cash users. In terms of expenditures, the monopoly network requires merchants and consumers to spend more on total card adoption costs and usage fees to pay for a lower card transaction value than the Ramsey regulator would do, so the card-service-spending-to-card-sales ratio is higher. Meanwhile, cash costs and cash transaction value are also higher under the monopoly network.

As the card service costs $d_m + d_c$ fall, the monopoly card network raises the merchant fee f_m but lowers the consumer fee f_c, and raises the card markup $f_c + f_m - d_c - d_m$. Meanwhile, fewer card-accepting merchants serve cash users (α_1 increases) so cash users gain less welfare. In contrast, the Ramsey regulator lowers card fees for both merchants and consumers and keeps the card markup at zero. Meanwhile, more card-accepting merchants serve cash users (α_1 decreases), and both card and cash users gain more welfare.

Our analysis sheds new light on the card pricing controversies. First, why does the card network charge a high merchant fee? The answer lies in its different objective than the Ramsey regulator. The card network aims to maximize its profit. By charging a high merchant fee, the network earns a positive markup and extracts the rents produced by replacing more costly cash payments that would have otherwise gone to consumers in the form of lower retail prices. Moreover, because the card network does not earn profit from cash users, charging a high merchant fee also reduces subsidies from card users to cash users through merchants in category (1). In contrast, the Ramsey regulator cares about consumers' real purchases rather than their nominal card spending, and cares about the welfare of both card users and cash users. Accordingly, the Ramsey regulator sets a zero markup and charges a much lower merchant fee.

Second, why have merchant card fees been rising over time? From a one-sided market perspective, this seems hard to explain given the card service cost $d_m + d_c$ has been falling during the same period. Our model shows that this could indeed happen in a two-sided market. As the card service cost $d_m + d_c$ falls, payment cards penetrate the consumer income distribution as well as the merchant size distribution. Under our distributional assumptions, the demand for card services by merchants becomes less elastic relative to consumers during the process. Therefore, the card network maximizes its profit by raising the merchant fee but reducing the consumer fee.

Merchant (interchange) fee cap regulation Our model provides a useful framework for conducting policy experiments, such as the interchange fee cap regulation. (Note that in our model, the interchange fee is equivalent to the merchant fee f_m under the assumption of costless acquiring). As an example, we conduct a simulation to compare the market outcomes under a monopoly card network with and without a merchant (interchange) fee cap. The simulation uses the same benchmark parameter values as Case 1 and the cap is set as $\bar{f}_m = 0.03$.

The simulation results, shown in Fig. A4 in the Appendix, are what we expect. Compared with Case 1 under no regulation, once a merchant fee cap is imposed, the card network then has to charge a higher fee to the consumer side to make up for the lost revenue from merchants, but the overall card markup is suppressed. As a result, merchant card adoption is higher, but consumer card adoption is lower. Meanwhile, the percentage of category (1) merchants becomes higher, and both card users and cash users enjoy higher welfare gains. In terms of expenditures, merchants and consumers now spend less on card adoption costs and usage fees to pay for a lower card transaction value, and the card-payment-spending-to-card-sales ratio decreases. At the same time, cash transaction value rises while the cash-cost-to-sales ratio is fixed at $\tau_m + \tau_c$. In total, society pays less for payment services, and the total-payment-spending-to-total-sales ratio decreases.

Our policy experiment suggests that regulating down the merchant (interchange) fee could improve welfare for both card users and cash users. While the regulation may discourage consumer card adoption, it increases card acceptance on the merchant side. Overall, the total card transaction value may decrease, but consumers as a whole benefit from paying less for card services.

The simulation results also show that simply capping the merchant fee by the card service cost $(d_m + d_c)$ or by merchant card benefit (τ_m) does not maximize consumer surplus. In general, in the two-sided card market, regulating the merchant fee only would not fully achieve the welfare goal because the consumer card fee is left unattended. Policymakers, in theory, could fix the problem by also regulating the consumer card fee. However, this may not be an easy task. On one hand, more regulations may come with more unintended consequences. On the other hand, there are additional challenges to set the optimal card fee levels. First, policymakers may need more information to correctly evaluate the social costs and benefits associated with card adoption and usage. Second, card service providers may face investment decisions which endogenously influence the costs of providing card services. These should be taken into account by policymakers. Third, our model assumes contestable merchant markets, which may not fit certain market segments where merchants can have substantial market power. Finally, regulating card fees may not be the best option for improving the market outcome. Policymakers may also want to consider reforming the card market structure and network rules.

4 Concluding remarks

This paper provides a new theory to study the pricing, adoption and usage of payment devices. The theory explains empirically relevant payment market patterns and also sheds light on related policy issues.

Adopting payment cards requires consumers and merchants to pay a fixed cost, but yields a lower marginal cost of making payments. Considering adoption and usage externalities among heterogeneous consumers and merchants, our theory yields some unique insights distinct from the existing literature. One of our first results shows that in equilibrium, wealthy consumers and large merchants are more likely to adopt card devices. Meanwhile, three types of merchants exist in the market: Some only serve cash customers, some only serve card customers, and some serve both. Our model also predicts that payment cards are adopted by lower-income consumers and smaller merchants over time. These findings fit well with empirical evidence. In another difference with existing models, cash users in our model benefit when a store has sales paid for both by cards and cash. In other words, cash users are "subsidized" by card users.[21]

Our analysis shows that the market-determined merchant fees tend to be too high

[21] Compared with cash and checks, payment cards are more cost effective for merchants in terms of lowering payment processing, fraud and monitoring costs. As a result, accepting cards allows merchants to lower retail prices, which can also benefit cash users who make purchases in the same store. In reality, merchants do complain about the interchange fees they pay to accept cards. Nevertheless, the fact that most merchants continue to accept payment cards suggests that cards are still more cost effective than alternative paper payments in spite of the high interchange fees.

from a consumer welfare point of view. Moreover, as the payment card market evolves, merchant fees tend to rise but consumer fees tend to fall, as observed in the data. We also show that imposing a cap on merchant (interchange) fees may improve consumer welfare. These findings provide some support for regulating down merchant (interchange) fees, though there are additional challenging issues that policymakers may need to address.

Our analysis can be extended in several directions. First, we may consider modeling consumer credit constraints and the provision of credit via payment cards. Adding the credit function may increase the card benefits to consumers and merchants. Taking that into consideration, Rochet and Wright (2010) find that a monopoly card network may always set an interchange fee that exceeds the level that maximizes consumer surplus. Second, we may consider extending our framework to introduce multiple card networks. It would be interesting to study how cooperation and competition among card networks affect the market outcome. Third, our model can be applied to emerging payment means, such as mobile payments. Compared with existing payment means, emerging payments like mobile require a substantial adoption cost but yield a lower marginal cost of transaction for consumers and merchants. Therefore, our analysis may also shed light on the development of emerging payments. Finally, and more broadly, our paper provides a case study of technology (product) adoption among heterogenous agents in a two-sided market setting. It would be interesting to explore pricing, adoption and usage issues in other industries featuring two-sided market effects, such as computer operating systems, video game consoles, online search engines, and communication networks.

Appendix:

1. *Heterogenous unit costs*

Consider a general setup where goods with the same value of α may be different in unit cost μ. We assume $\alpha \in (0, \overline{\alpha})$ and $\mu \in (\underline{\mu}, \overline{\mu})$ are jointly distributed across goods according to a cumulative distribution function $H(\alpha, \mu)$.

Merchants incur a transaction cost τ_m per dollar for accepting cash. The market is contestable so the cash price for a good α is determined as

$$p^h_{(\alpha,\mu)} = \frac{\mu}{1-\tau_m}.$$

A consumer, indexed by her income I, has Cobb-Douglas preferences. In a cash economy, the consumer maximizes her utility subject to the budget constraint:

$$U = Max \int_0^{\overline{\alpha}} \int_{\underline{\mu}}^{\overline{\mu}} \alpha \ln x_{(\alpha,\mu),I} dH(\alpha,\mu) \qquad s.t. \qquad \int_0^{\overline{\alpha}} \int_{\underline{\mu}}^{\overline{\mu}} (1+\tau_c) p^h_{(\alpha,\mu)} x_{(\alpha,\mu),I} dH(\alpha,\mu) \ = I,$$

where $x_{(\alpha,\mu),I}$ is her quantity of demand for a good (α, μ) and τ_c is the consumer's transaction cost of using cash. Solving the maximization problem yields consumer I's demand for good (α, μ)

$$x_{(\alpha,\mu),I} = \frac{\alpha I}{(1+\tau_c) p^h_{(\alpha,\mu)} E(\alpha)}.$$

Assume income $I \in (0, \overline{I})$ is distributed across consumers according to a cumulative distribution function $F(I)$. We denote $E(I)$ as the mean income and normalize the measure of consumers to be unity. At equilibrium, the supply of each good (α, μ) equals its total demand

$$x_{(\alpha,\mu)} = \int_0^{\overline{I}} x_{(\alpha,\mu),I} dF(I) = \frac{\alpha E(I)}{(1+\tau_c) p^h_{(\alpha,\mu)} E(\alpha)}.$$

The above results show that individual consumer spending $p^h_{(\alpha,\mu)} x_{(\alpha,\mu),I}$ and total consumer spending $p^h_{(\alpha,\mu)} x_{(\alpha,\mu)}$ on a good (α, μ) depends only on consumer preference α but not the unit cost μ. As a result, when the payment card is introduced, the efficiency gain to a merchant from adopting payment cards relative to accepting cash only depends on α but not μ. In fact, merchants with the same value of α would make the same card adoption decision regardless of their differences in unit cost μ. Therefore, our analysis in the paper holds for the more general setup.

2. Derivation of Eq (L2)

Given that α is uniformly distributed, Eq (18) suggests:

$$I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} k_c}{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} - \exp(2\int_{\alpha_0}^{\frac{Z_0}{Z_1}\alpha_0} \alpha \ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0})d\alpha)}.$$

Note that

$$\begin{aligned}
& \int \alpha \ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0})d\alpha \\
= \; & \frac{\alpha^2}{2}\ln(1-\tau_m) - \frac{1}{2}\alpha^2 \ln((1-f_m) - \frac{\alpha_0 Z_0(1+f_c)}{\alpha}) \\
& + \frac{\alpha_0 Z_0(1+f_c)}{2(1-f_m)}\alpha + \frac{\alpha_0^2 Z_0^2(1+f_c)^2}{2(1-f_m)^2}\ln(\alpha - \frac{\alpha_0 Z_0(1+f_c)}{(1-f_m)}).
\end{aligned}$$

Hence, we derive

$$\begin{aligned}
& 2\int_{\alpha_0}^{\frac{Z_0}{Z_1}\alpha_0} \alpha \ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0})d\alpha \\
= \; & \alpha_0 \left\{ \ln\frac{1+f_c}{1+\tau_c} + \frac{Z_0(1+f_c)}{(1-f_m)}(\frac{Z_0}{Z_1}-1) + \frac{Z_0^2(1+f_c)^2}{(1-f_m)^2}\ln(\frac{(\frac{Z_0}{Z_1}) - \frac{Z_0(1+f_c)}{(1-f_m)}}{1-\frac{Z_0(1+f_c)}{(1-f_m)}}) \right\}.
\end{aligned}$$

Equation (18) can then be rewritten into Eq (L2).

3. Monopoly network's problem

Assuming that α is uniformly distributed and I is exponentially distributed, the monopoly card network maximizes its profit as follows.

$$\underset{f_c, f_m}{Max} \quad e^{(-\lambda I_0)}(\frac{1}{\lambda} + I_0 - k_c)(\frac{1-\alpha_0^2}{1+f_c})(f_c + f_m - d_c - d_m)$$

$$s.t. \quad \alpha_0 = \frac{k_m}{2e^{(-\lambda I_0)}(\frac{1}{\lambda} + I_0 - k_c)Z_0},$$

$$\alpha_1 = \frac{Z_0}{Z_1}\alpha_0 \quad \text{if} \quad f_m \leq \tau_m,$$

$$I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} k_c}{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} - \exp(2\int_{\alpha_0}^{\min(\alpha_1,1)} \alpha \ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0})d\alpha)},$$

$$\tau_c \geq f_c, \quad \frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}.$$

4. *Proof of Proposition 3*

Proposition 3. *The increase of consumer mean income ($1/\lambda$) and the joint decline of adoption costs (k_c, k_m) have equivalent effects on card pricing and adoption. Formally, the network profit maximization yields the same card fees and adoption rates under the parameter values* ($1/\lambda$, k_c, k_m) *and* (θ/λ, θk_c, θk_m).

Proof. Given the monopoly network's problem stated above in Appendix 3, we first prove if (f_c^*, f_m^*, I_0^*, α_0^*) maximizes the network's profit under the parameter values ($1/\lambda$, k_c, k_m), then (f_c^*, $f_m^*, \theta I_0^*, \alpha_0^*$) maximizes the network's profit under the parameter values (θ/λ, θk_c, θk_m).

This can be shown by constructing a contradiction. Assume (f_c^*, f_m^*, I_0^*, α_0^*) generates the maximal network profit π^* under the parameter values ($1/\lambda$, k_c, k_m). Then, under parameter values ($\theta/\lambda, \theta k_c, \theta k_m$), ($f_c^*$, $f_m^*, \theta I_0^*, \alpha_0^*$) satisfies the constraints and offers profit $\theta\pi^*$. Assume (f_c^*, $f_m^*, \theta I_0^*, \alpha_0^*$) does not maximize the profit, then there is another choice ($f_c^{'}$, $f_m^{'}, I_0^{'}, \alpha_0^{'}$) that offers profit $\pi^{'} > \theta\pi^*$. However, ($f_c^{'}$, $f_m^{'}, I_0^{'}/\theta, \alpha_0^{'}$) also satisfies the constraint and offers a profit $\pi^{'}/\theta > \pi^*$ under the parameter values ($1/\lambda$, k_c, k_m). This contradicts the assumption that π^* is the maximal profit under ($1/\lambda$, k_c, k_m). Therefore, under the parameter values $(1/\lambda,\ k_c,\ k_m)$ and $(\theta/\lambda,\ \theta k_c,\ \theta k_m)$, solving the monopoly network problem yields the same card fees $(f_c^*,\ f_m^*)$ and adoption rates $(e^{(-\lambda I_0^*)}, 1-\alpha_0^*)$. Note that under the parameter values $(1/\lambda,\ k_c,\ k_m)$, consumer card adoption rate is $1-F_\lambda(I_0^*) = e^{(-\lambda I_0^*)}$; with $(\theta/\lambda,\ \theta k_c,\ \theta k_m)$, consumer card adoption rate is again the same, $1-F_{\lambda/\theta}(\theta I_0^*) = e^{(-\lambda I_0^*)}$. ■

5. *Ramsey regulator's problem*

Assuming that α is uniformly distributed and I is exponentially distributed, the Ramsey regulator maximizes consumer surplus as follows.

$$\underset{f_c,f_m}{Max} \int_0^\infty (U_{I,d} - U_{I,h})\lambda e^{-\lambda I} dI$$

$$s.t. \quad (U_{I,d} - U_{I,h})_{I \geq I_0} = \int_{\alpha_0}^{\min(\alpha_1,1)} \alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) d\alpha + \int_{\min(\alpha_1,1)}^{\overline{\alpha}} \alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) d\alpha + (\frac{1-\alpha_0^2}{2}) \ln(\frac{1+\tau_c}{1+f_c}) + \frac{1}{2}\ln(\frac{I-k_c}{I}),$$

$$(U_{I,d} - U_{I,h})_{I<I_0} = \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) d\alpha,$$

$$(\frac{p_{\alpha,h}}{p_{\alpha,d}})_{\overline{\alpha} \geq \alpha \geq \min(\alpha_1,\overline{\alpha})} = \frac{\frac{1-f_m}{1+f_c} e^{-\lambda I_0}(\frac{1}{\lambda} + I_0 - k_c) + \frac{1-\tau_m}{1+\tau_c}[\frac{1}{\lambda} - e^{-\lambda I_0}(\frac{1}{\lambda} + I_0)] - \frac{k_m}{2\alpha}}{\frac{1-\tau_m}{1+f_c} e^{-\lambda I_0}(\frac{1}{\lambda} + I_0 - k_c) + \frac{1-\tau_m}{(1+\tau_c)}[\frac{1}{\lambda} - e^{-\lambda I_0}(\frac{1}{\lambda} + I_0)]},$$

$$(\frac{p_{\alpha,h}}{p_{\alpha,d}})_{\min(\alpha_1,\overline{\alpha}) > \alpha \geqslant \alpha_0} = \frac{(\frac{1-f_m}{1+f_c}) e^{-\lambda I_0}(\frac{1}{\lambda} + I_0 - k_c) - \frac{k_m}{2\alpha}}{\frac{1-\tau_m}{1+f_c} e^{-\lambda I_0}(\frac{1}{\lambda} + I_0 - k_c)},$$

$$\alpha_0 = \frac{k_m}{2e^{(-\lambda I_0)}(\frac{1}{\lambda} + I_0 - k_c) Z_0},$$

$$\alpha_1 = \frac{Z_0}{Z_1}\alpha_0 \quad \text{if} \quad f_m \leq \tau_m,$$

$$I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} k_c}{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} - \exp(2\int_{\alpha_0}^{\min(\alpha_1,1)} \alpha \ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0}) d\alpha)},$$

$$\tau_c \;\; \geq \;\; f_c, \;\; \frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}, \qquad f_m + f_c \geq d_c + d_m.$$

6. *Simulations (Figs. A1-A4)*

The simulations compare three regimes: (1) a monopoly card network, (2) a card network run by a Ramsey regulator, and (3) a monopoly network under a merchant (interchange) fee cap regulation. We consider five cases of parameterization as follows.

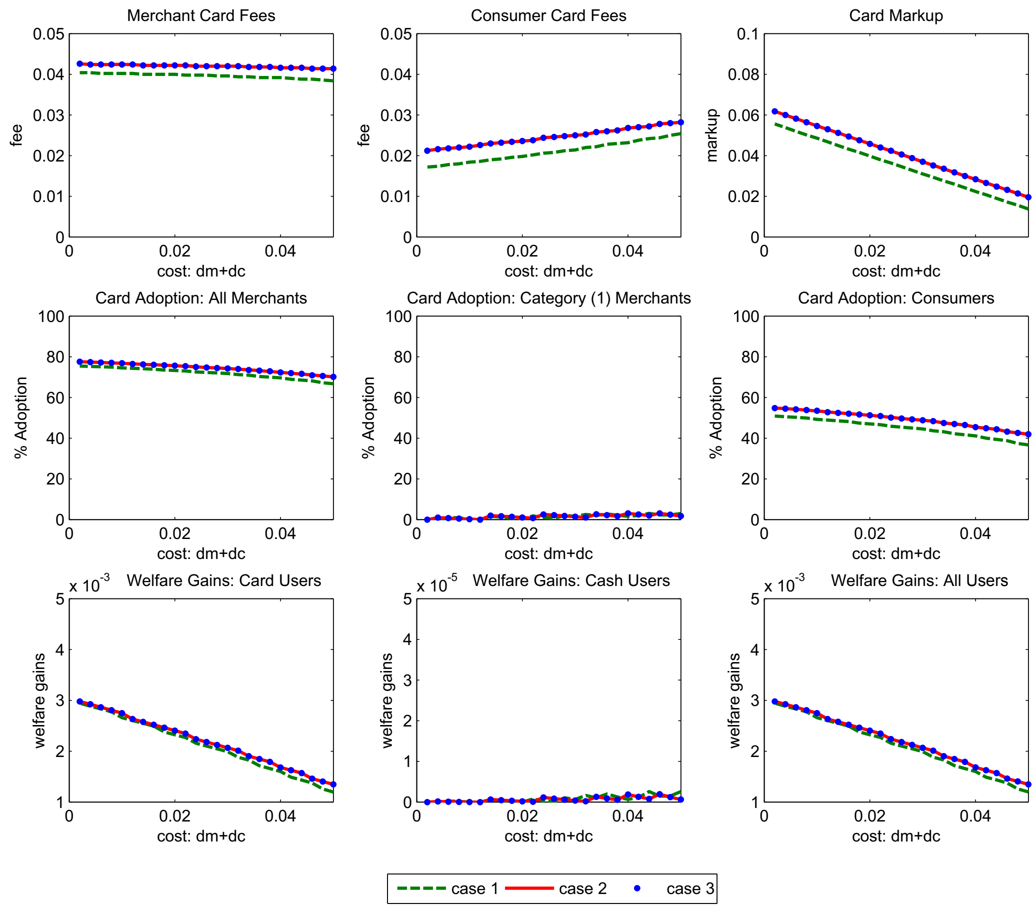
Simulation Cases (1-5)

	k_m	k_c	$1/\lambda$	τ_m	τ_c	d_m+d_c
Case 1	160	160	$10,000$	0.05	0.05	$(0, 0.05)$
Case 2	128	128	$10,000$	0.05	0.05	$(0, 0.05)$
Case 3	160	160	$12,500$	0.05	0.05	$(0, 0.05)$
Case 4	300	20	$10,000$	0.05	0.05	$(0, 0.05)$
Case 5	160	160	$10,000$	0.08	0.02	$(0, 0.05)$

The simulation results are presented in the attached Figs A1-A4 as follows.

- Figure A1 compares monopoly outcomes for Cases 1, 2 and 3.
- Figure A2 compares monopoly outcomes for Cases 1, 4 and 5.
- Figure A3 compares the monopoly outcome with the Ramsey regulator's outcome for Case 1.
- Figure A4 compares the monopoly outcomes for Case 1 with and without an merchant fee cap, where the cap is set as $\bar{f}_m = 0.03$.

Figure A1: Monopoly Network (Case 1, 2, 3)



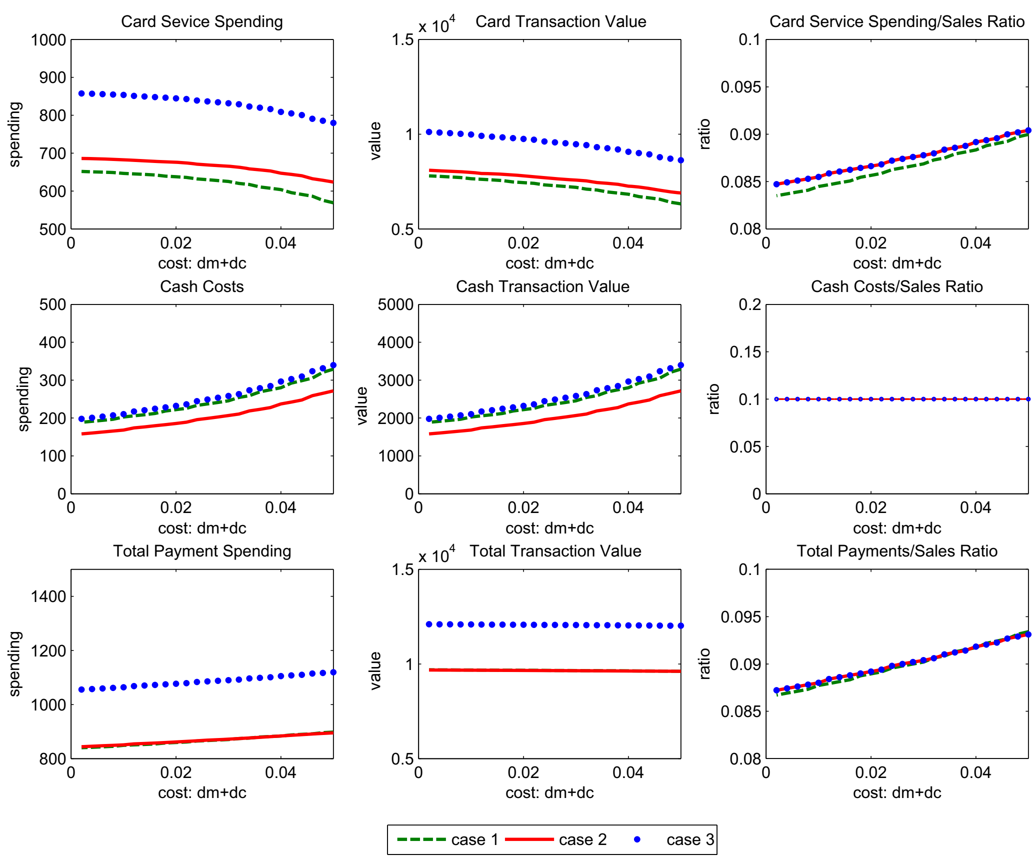
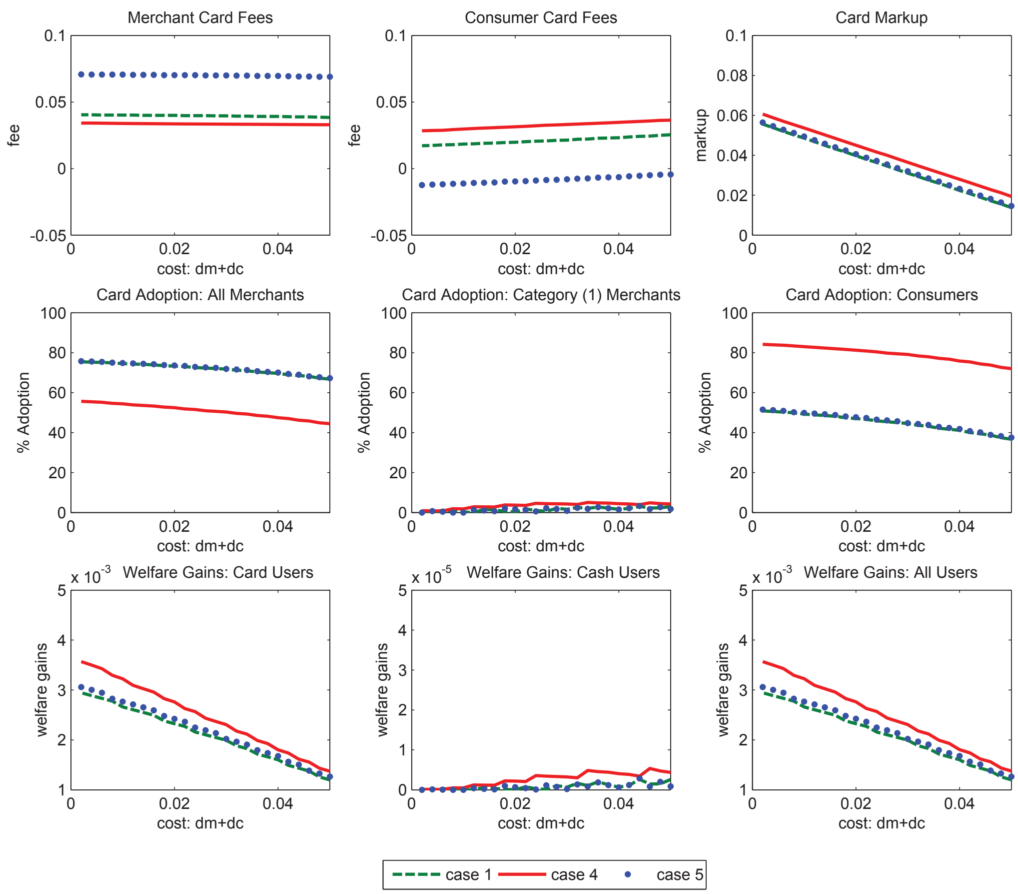


Figure A2: Monoply Network (Case 1, 4, 5)



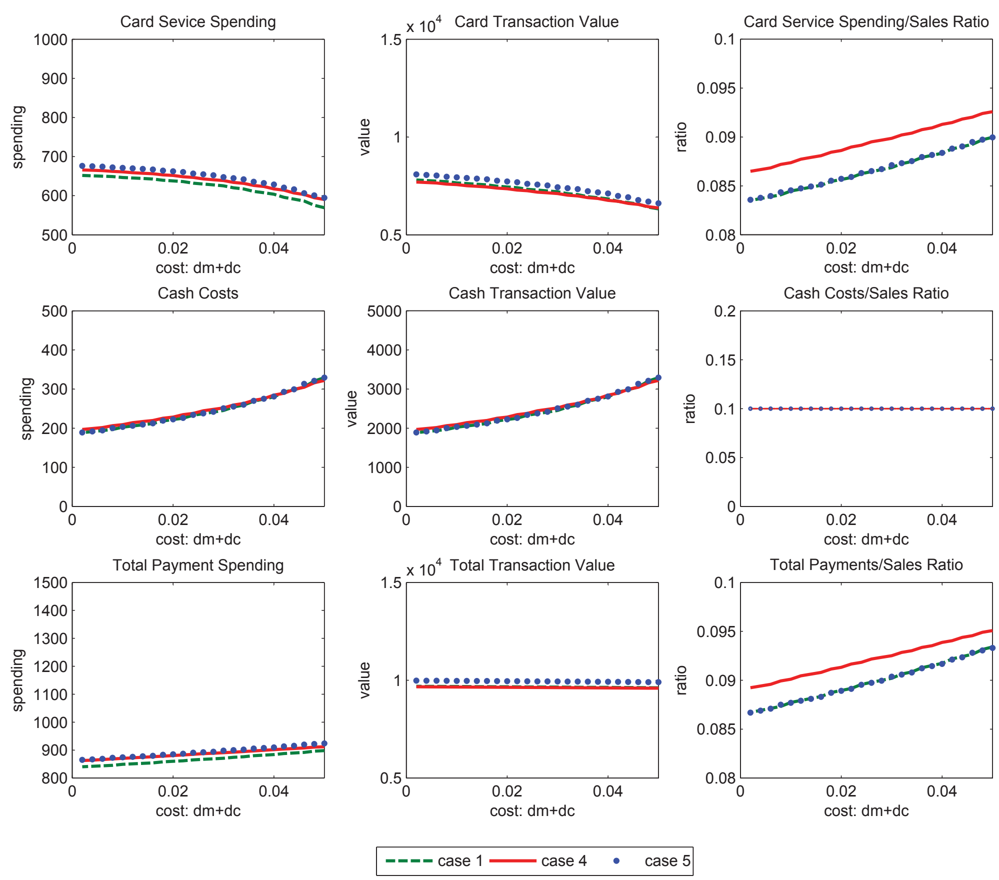
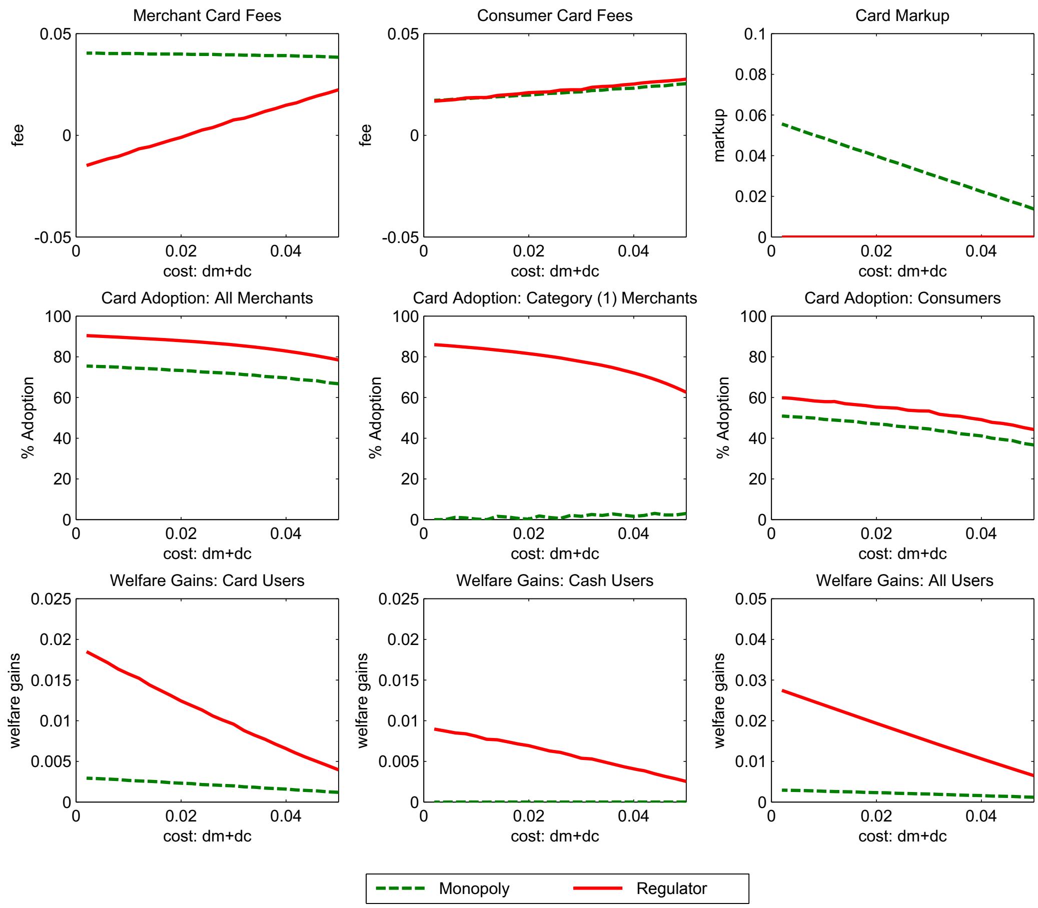


Figure A3: Monopoly Network vs. Ramsey Regulator (Case 1)



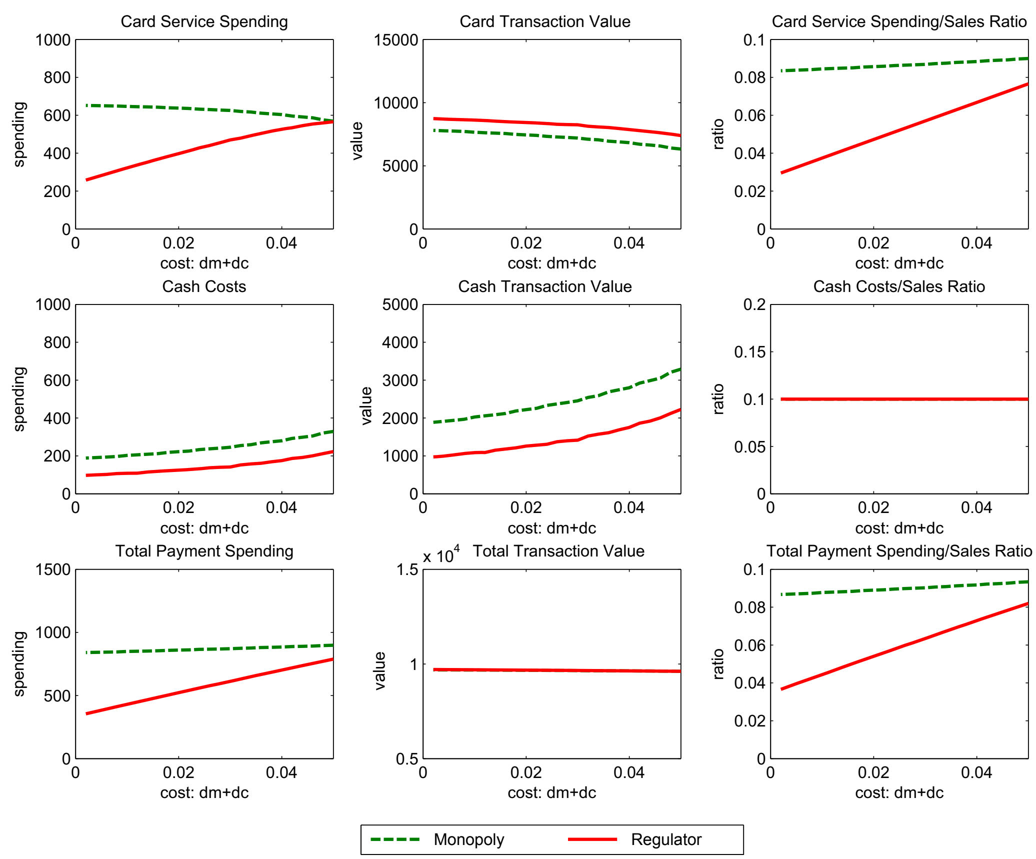
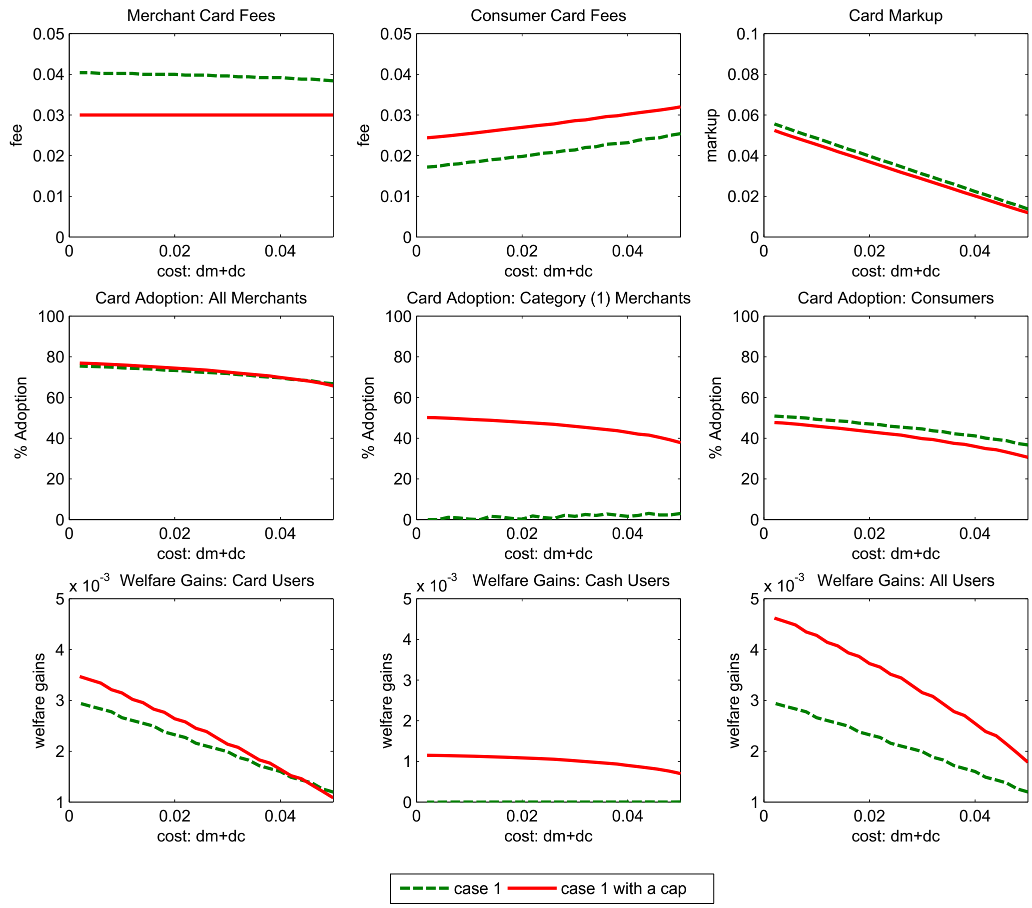
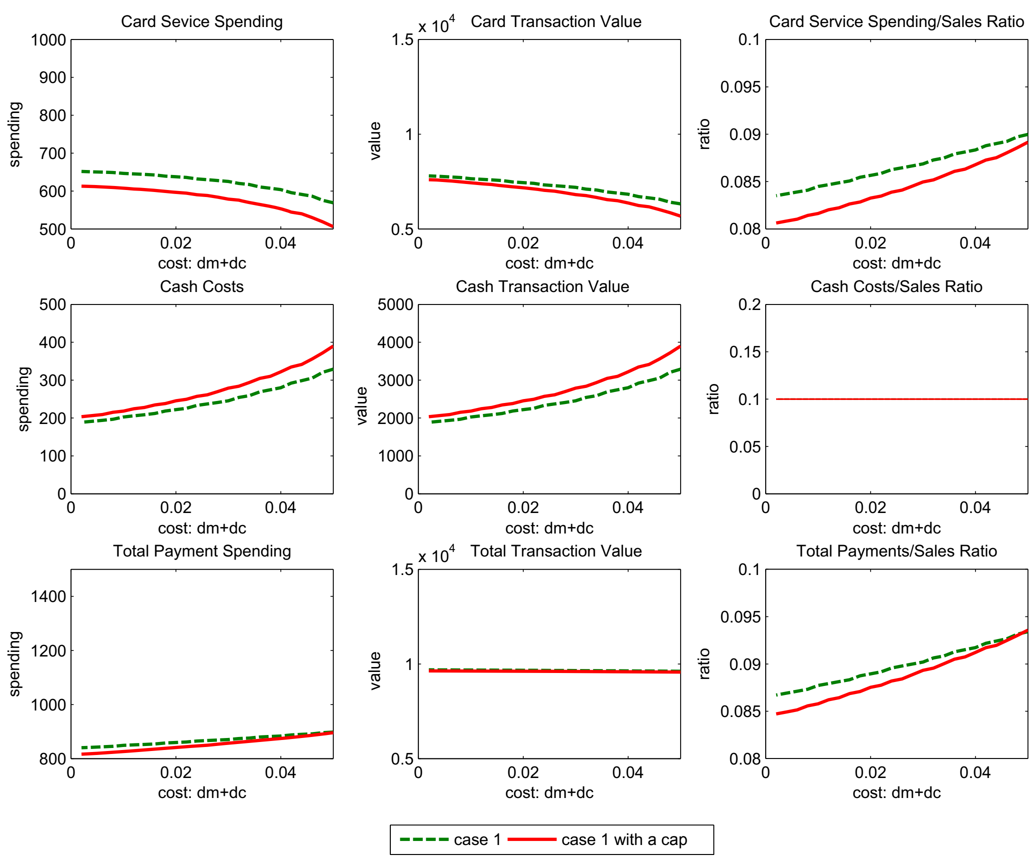


Figure A4: Monopoly Network under a Merchant Fee Cap (Case 1)





References

[1] Armstrong, Mark (2006), "Competition in Two-Sided Markets," *RAND Journal of Economics*, 37(3), 668–681.

[2] Baxter, William (1983), "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics* 26: 541-588.

[3] Bedre, Ozlem and Emilio Calvano (2009), "Pricing Payment Cards," European Central Bank Working Paper Series No. 1139.

[4] Bolt, Wilko and Sujit Chakravorti (2008), "Economics of Payment Cards: A Status Report," *Economic Perspectives*, Federal Reserve Bank of Chicago, issue Q IV, pages 15-27.

[5] Bradford, Terri and Fumiko Hayashi (2008), "Developments in Interchange Fees in the United States and Abroad," *Payments System Research Briefings,* May, Federal Reserve Bank of Kansas City.

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

[7] Fudenberg, Drew and Jean Tirole (1991), *Game Theory.* MIT Press, Cambridge, MA.

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

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

[10] Prager, Robin, Mark Manuszak, Elizabeth Kizer, and Ron Borzekowski, (2009), "Interchange Fees and Payment Card Networks: Economics, Industry Developments, and Policy Issues." *Finance and Economics Discussion Series* 2009-23. Washington: Board of Governors of the Federal Reserve System.

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

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

[13] Rochet, Jean-Charles and Jean Tirole (2011), "Must-Take Cards: Merchant Discounts and Avoided Costs," *Journal of the European Economic Association,* 9 (3): 462–495.

[14] Rochet, Jean-Charles and Julian Wright (2010), "Credit Card Interchange Fees," *Journal of Banking and Finance,* 34: 1788-1797.

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

[16] Rysman, Marc (2009), "The Economics of Two-Sided Markets." *Journal of Economic Perspectives*, 23: 125-144.

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

[18] Schuh, Scott, Oz Shy and Joanna Stavins (2010), "Who Gains and Who Loses from Credit Card Payments?" Federal Reserve Bank of Boston Public Policy Discussion Papers Series, No. 10-3.

[19] Shy, Oz and Zhu Wang (2011), "Why Do Payment Card Networks Charge Proportional Fees?" *American Economic Review*, 101 (4): 1575-1590.

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

[21] Weyl, Glen, (2010), "A Price Theory of Multi-Sided Platforms," *American Economic Review*, 100 (4): 1642-1672.

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

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