ISSN 1936-5330

THE FEDERAL RESERVE BANK *of* KANSAS CITY
ECONOMIC RESEARCH DEPARTMENT

Why Do Payment Card Networks Charge Proportional Fees?

Ozy Shy and Zhu Wang

December 2008; Revised January 2010
RWP 08-13

RESEARCH WORKING PAPERS

Why Do Payment Card Networks Charge Proportional Fees?*

Oz Shy[†] and Zhu Wang[‡]

January, 2010

Abstract

This paper explains why payment card networks charge fees that are proportional to the transaction values instead of charging fixed per-transaction fees. We show that, when card networks and merchants both have market power, card networks earn higher profits by charging proportional fees. It is also shown that competition among merchants reduces card networks' gains from using proportional fees relative to fixed per-transaction fees. Merchants are found to earn lower profits under proportional fees whereas consumer utility and social welfare are higher. Our welfare results are then evaluated with respect to the current regulatory policy debates.

JEL classification: D4, L1, G2

Keywords: Payment cards, Pricing structures, Double marginalization

*Forthcoming, *American Economic Review*. We thank Fumiko Hayashi, the editor and an anonymous referee for most valuable comments and suggestions. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Boston, the Federal Reserve Bank of Kansas City or the Federal Reserve System.

†Shy: Research Department, Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02210. Email: oz.shy@bos.frb.org.

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

1 Introduction

Credit and debit cards have become an important part of our payment system and they affect a large portion of the consumer and merchant populations. Recent Federal Reserve studies found that 80.2% of the U.S. consumers have debit cards and 78.3% have credit cards. In a typical month, 31.0% of consumer payments are paid for with debit cards, and 21.3% with credit cards (Foster et al., 2009). In 2006, debit cards were used in 25.3 billion transactions for a total value of $1.0 trillion, and credit cards were used in 21.7 billion transactions for a total value of $2.1 trillion (Gerdes, 2008).

Along with this development, some controversy has emerged because merchants are critical of the card fees, challenging both the fee structure and the level.[1] This has generated heated debates among researchers, practitioners and policymakers in recent years. On the research side, a large body of literature, called "two-sided market theory," has been developed to evaluate payment card market competition and pricing issues.[2] On the policy side, three bills are currently under consideration in Congress. If passed, the new legislation could open the possibility of regulating card fees in the United States. Similar trends are also taking place in many other countries. More than 20 countries and areas around the world have started regulating or investigating card fees.[3]

There are several main issues at the heart of the card fee controversy. For example, should card issuers be allowed to collectively set merchant (interchange) fees? How do the resulting fee levels correspond to social optimum? Why are these fees proportional to the transaction values rather than fixed per-transaction fees? While the academic literature

[1]Recently, there have been more than 50 lawsuits filed by merchants and merchant associations against card networks and card issuers regarding merchant card fees. Many of the recent lawsuits have been consolidated in an ongoing case before the U.S. District Court of the Eastern District of New York.

[2]For example, Baxter (1983), Katz (2001), Schmalensee (2002), Rochet and Tirole (2002, 2003), Gans and King (2003), Wright (2003, 2004), Schwartz and Vincent (2006), McAndrews and Wang (2008), Rysman (2009) and Wang (2010a, 2010b).

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

has so far focused on the collective setting and level of merchant (interchange) fees, this paper solves the puzzle why card networks charge proportional fees.

There are four types of general purpose payment cards in the United States: Credit cards, charge cards, signature debit cards and PIN debit cards. Charge cards are similar to credit cards, except that they require cardholders to pay off full charges every month. Signature debit cards and PIN debit cards do not extend credit to cardholders. Instead, they debit the cardholder's bank account right after each transaction.[4] The first three types of cards are routed over the credit card networks and charge fees that are proportional to the transaction values.[5] In contrast, PIN debit cards are routed over the PIN debit networks, which used to charge fixed per-transaction fees. However, in recent years, PIN debit cards have been shifting to the proportional fee model, though typically with a cap.[6]

Figure 1 shows the average total fees that merchants pay for accepting various payment cards. These fees are often referred to as the "merchant discount," and a large portion of the merchant discount comes from interchange fees, paid by merchants to card issuers through merchant acquirers. The fees for credit card transactions are the highest, followed by the fees for signature debit cards and PIN debit cards. On average, merchants pay about 1.75-2.41% of the transaction value for accepting signature debit cards and credit cards, and 0.62% for accepting PIN debit cards. These fees generate significant revenue for card service providers. In 2008, U.S. merchants paid $60.9 billion in merchant discount

[4] Debit card payments are typically authorized either with a PIN or by the cardholder's signature. In terms of transaction volume, signature debit cards account for 60 percent of debit transactions, and PIN debit cards account for 40 percent.

[5] Visa and MasterCard are the two major credit card networks in the United States. They provide card services through member financial institutions (card issuers and merchant acquirers), and account for 80 percent U.S. credit card market share. These two networks are also the sole providers of signature debit cards, with Visa holding 75 percent market share and MasterCard holding 25 percent.

[6] There are 14 major PIN debit networks in the United States. The proportional fees on PIN debit transactions started in 1996 by Interlink for non-supermarket transactions. Star and NYCE followed suit in 2000. Exchange started using proportional fees in 2002, Maestro in 2003, and Pulse and Jeanie in 2005. Major PIN debit cards currently charge between 0.45 to 0.75 percent of the transaction value plus 5 to 15 cents, capped around 40 to 65 cents for non-supermarket transactions. More recently, some leading networks removed "caps" for small retail merchant categories, and many networks raised "capped amount." (Sources: *EFT Data Book*, various issues, MasterCard International and Visa USA).

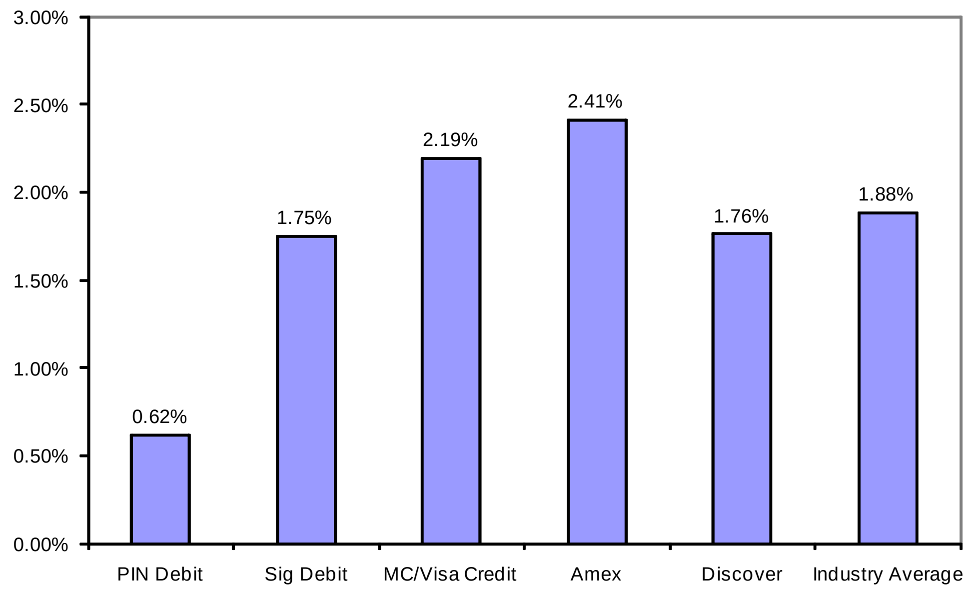

Figure 1: Average merchant discount rates. *Source*: Hayashi (2009).

fees (*Nilson Report*, 2009).

Card networks have often argued that the reason why they impose proportional fees stems from the costs they bear from processing card transactions and providing "payment guarantee" services. However, empirical evidence suggests that the cost components cannot explain the merchant fees commonly set at the 1-3% level for brand name cards.[7] Particularly, debit cards do not provide credit float and bear very small fraud risk, so there appears to be no cost basis for charging proportional fees. In fact, the PIN debit networks used to work well (some are still working) by charging fixed per-transaction fees. In many other countries, fixed per-transaction fees are the norm for debit cards.[8] Therefore, a natural question to ask is why U.S. card networks charge proportional fees.

In this paper we demonstrate that the "double marginalization" market structure pro-

[7] For example, industry studies show that the average net fraud losses to card issuers are 0.08% for credit cards, 0.05% for signature debit cards and 0.01% for PIN debit cards.

[8] Several countries have already debated whether to allow proportional fees on debit transactions. Most recently, the Canadian Senate held hearings on this issue and rejected the request of Interac, the Canadian debit system, for charging proportional fees (See "Transparency, Balance and Choice: Canada's Credit Card and Debit Card Systems," the Standing Senate Committee on Banking, Trade and Commerce, Canada, June 2009).

vides the major explanation for this puzzle. Viewing card markets and merchant markets as vertically related, we find that upstream card networks can better restrain the market power of downstream merchants by charging proportional fees. As a result, card networks earn higher profits by charging proportional fees compared with fixed per-transaction fees. Using a simple model, we also investigate how changing the degree of competition among merchants affects card networks' gains from proportional fees relative to fixed per-transaction fees. Finally, we investigate and compare merchant profits, consumer utility and social welfare under the two types of fees.

The paper is organized as follows. Section 2 sets up a simple model of a card network, merchants, and consumers. Section 3 solves for the proportional card fees which maximize the profit of the card network. Section 4 computes the fixed per-transaction card fees. Section 5 compares equilibrium market allocations under proportional fees and fixed per-transaction fees. Section 6 incorporates card annual fees into the analysis. Section 7 concludes with some policy discussions and other applications.

2 The Model

We study a simple model of a payment card system illustrated in Figure 2. Consumers use a payment card to pay merchants. Merchants submit charges to a card network which then bills consumers. For the purpose of this paper we deliberately abstract from the internal organization of card networks which consist of merchant acquirers who receive requests for payments from merchants, and card issuers who bill consumers and send the money to merchant acquirers who then pay the merchants.[9]

2.1 Consumers

Consumers buy two types of goods. Let Y denote the consumption of goods for which consumers pay only with cash (cash goods in what follows). Let Q denote the consumption

[9] There is no loss of generality with this simplification because considering additional players within card networks would further enhance card networks' incentive to charge proportional fees.

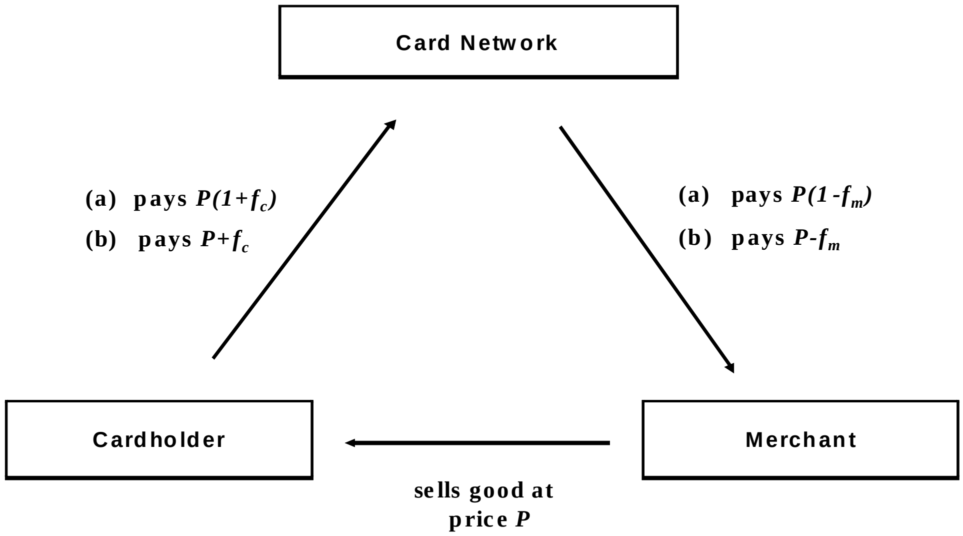

Figure 2: A card system with (a) proportional fees *or* (b) fixed per-transaction fees

of goods for which consumers pay only with cards (card goods).[10] Consumers' utility function is assumed to take the quasi-linear form given by

$$U = Y + \gamma Q^{1-\beta}, \quad \text{where} \;\; 0 < \beta < 1, \; \gamma > 0. \tag{1}$$

The price of card goods is denoted by P_c. The price of cash goods is normalized to unity. Consumers' budget constraint is given by

$$Y + P_c Q = I, \tag{2}$$

where I denotes consumers' exogenously-given income.

Substituting (2) for Y in (1), and maximizing with respect to Q obtains consumers'

[10] We interpret cash as non-card payments which also include checks and bank transfers. Here we implicitly assume that the cost of using cash is too high for card goods. Alternatively, we could explicitly model consumers' decisions regarding which payment instrument to use, but that would not affect our analysis. In fact, we may assume that each consumer has a unit demand for goods, but consumers are heterogenous with respect to the benefits they derive from using cards relative to cash. Assuming that consumers' benefits from using cards follow a Pareto distribution would then generate the same isoelastic demand function (3) for card goods (See Wang, 2010b).

isoelastic demand for card goods

$$P_c = \alpha Q^{-\beta}, \quad \text{where } \alpha = \gamma(1-\beta). \tag{3}$$

The price consumers pay for card goods P_c is composed of two parts: (a) the retail price charged by merchants, denoted by P, and (b) a consumer card fee charged by the card network denoted by f_c. The consumer card fee could be proportional to the price or a fixed per-transaction fee.[11] For consumers, the overall price they pay for card goods is

$$P_c = \begin{cases} (1+f_c)P & \text{under proportional fee,} \\ P + f_c & \text{under fixed per-transaction fee.} \end{cases} \tag{4}$$

2.2 Merchants

There are $\phi \geq 1$ merchants who sell card goods. Each merchant obtains the goods at a unit cost μ and sells them at a retail price P. The card network charges merchants a card fee f_m.

Let q_i denote the output sold by merchant i. Then, the total cost borne by each merchant i $(i = 1, 2, \ldots, \phi)$ is

$$c(f_m,\, q_i) = \begin{cases} (\mu + f_m P)q_i & \text{under proportional fee,} \\ (\mu + f_m)q_i & \text{under fixed per-transaction fee.} \end{cases} \tag{5}$$

Each merchant i sets the output level q_i taking the output by competing merchants $q_{-i} = Q - q_i$ as given, and maximizes profit given by

$$\pi_i(f_m,\, q_i) = Pq_i - c(f_m,\, q_i), \tag{6}$$

where P can be extracted from (3) and (4), and expressed as a function of (q_i, q_{-i}).

[11] Note that a fixed per-transaction fee is analytically equivalent to a fixed per-item fee as long as the number of transactions is proportional to the number of items sold.

2.3 The card network and timing

The card network is assumed to incur a fixed cost ν for processing each card transaction.[12] The profit earned by the card network is therefore

$$\Pi_n(f_c,\ f_m) = \begin{cases} (f_c + f_m)PQ - \nu Q & \text{under proportional fees,} \\ (f_c + f_m - \nu)Q & \text{under fixed per-transaction fees.} \end{cases} \tag{7}$$

The following two sections solve for a subgame perfect equilibrium of the following two-stage game:

Stage I. The card network chooses card fees f_c and f_m to maximize profit (7).

Stage II. Each merchant takes card fees f_c and f_m as given and chooses output level q_i to maximize profit (6), where in this stage output levels must satisfy the Nash-Cournot outcome.

3 Proportional Card Fees

We first consider proportional card fees. Notice that the equilibrium market allocation is unique up to a combination of f_c and f_m in the two-sided card market that we consider. Therefore, without loss of generality, we set $f_c = 0$ in the following analysis.[13] The consumer demand for card goods (3) becomes

$$P = \alpha Q^{-\beta}. \tag{8}$$

The cost borne by each merchant is already given in the top row of (5).

[12] This simplifying assumption suits well with debit cards, which are at the center of the controversy regarding proportional fees versus fixed per-transaction fees.

[13] We thank the referee for suggesting this. Note that the equalibrium market allocation is unique up to a combination of f_c and f_m is consistent with the neutrality result found in the two-sided market literature, e.g., Rochet and Tirole (2002), Wright (2003) and Gans and King (2003).

In Stage II, each merchant i takes q_{-i} as given and chooses output q_i to solve

$$\max_{q_i} \pi_i = Pq_i - (\mu + f_m P)q_i = \alpha(q_i + q_{-i})^{-\beta}(1 - f_m)q_i - \mu q_i, \tag{9}$$

where the price P was substituted from (8).

The first-order condition for the profit maximization problem (9) yields

$$(1 - f_m)\,\alpha(q_{-i} + q_i)^{-\beta}(1 - \frac{\beta q_i}{q_{-i} + q_i}) = \mu.$$

In a symmetric Nash-Cournot equilibrium all merchants sell the same amount, so that $q_i = q$ for each merchant $i = 1, \ldots, \phi$. Therefore, the total merchant output, $Q^{pr} = \phi q$, and market price P^{pr} are

$$Q^{pr} = \left[\frac{\phi\mu}{\alpha(\phi - \beta)(1 - f_m)}\right]^{-\frac{1}{\beta}} \quad \text{and} \quad P^{pr} = \frac{\phi\mu}{(\phi - \beta)(1 - f_m)}, \tag{10}$$

where the superscript "pr" stands for proportional card fees. As expected, the equilibrium price (output) increases (decreases) with the merchant card fee f_m.

In Stage I, the card network takes the merchant equilibrium (10) as given and chooses the merchant card fee f_m to solve

$$\max_{f_m} \Pi_n^{pr} = f_m P^{pr} Q^{pr} - \nu Q^{pr} = \left[\frac{f_m \phi\mu}{(\phi - \beta)(1 - f_m)} - \nu\right] \left[\frac{\phi\mu}{\alpha(\phi - \beta)(1 - f_m)}\right]^{-\frac{1}{\beta}}, \tag{11}$$

which yields the merchant card fee

$$f_m^{pr} = \frac{\beta\phi\mu + \nu(\phi - \beta)}{\phi\mu + \nu(\phi - \beta)}. \tag{12}$$

The resulting equilibrium profit of the card network is therefore

$$\Pi_n^{pr} = \beta\alpha^{\frac{1}{\beta}} \left[\frac{\phi\mu + \nu(\phi - \beta)}{(\phi - \beta)(1 - \beta)}\right]^{1 - \frac{1}{\beta}}. \tag{13}$$

The equilibrium price and total merchant output can be computed from (10) to be

$$P^{pr} = \frac{\phi\mu + \nu(\phi - \beta)}{(\phi - \beta)(1 - \beta)} \quad \text{and} \quad Q^{pr} = \left[\frac{\phi\mu + \nu(\phi - \beta)}{\alpha(\phi - \beta)(1 - \beta)}\right]^{-\frac{1}{\beta}}. \tag{14}$$

Substituting (14) into (9) for $q_i = Q^{pr}/\phi$ yields the total profit of all merchants

$$\Pi_m^{pr} = \phi\pi_i^{pr} = [(1 - f_m)P^{pr} - \mu]Q^{pr} = \frac{\beta\mu}{\phi - \beta}\left[\frac{\phi\mu + \nu(\phi - \beta)}{\alpha(\phi - \beta)(1 - \beta)}\right]^{-\frac{1}{\beta}}. \tag{15}$$

Adding up (13) and (15) yields the sum of card network and merchant profits

$$\Pi_{m+n}^{pr} = \Pi_m^{pr} + \Pi_n^{pr} = \alpha^{\frac{1}{\beta}}\beta\left[\frac{\phi\mu + \nu(\phi - \beta)}{(1 - \beta)(\phi - \beta)}\right]^{1-\frac{1}{\beta}}\left[1 + \frac{\mu(1 - \beta)}{\phi\mu + \nu(\phi - \beta)}\right]. \tag{16}$$

Substituting (14) into consumers' utility function (1) and (2) obtains the equilibrium consumer utility

$$U^{pr} = I - P^{pr}Q^{pr} + \frac{\alpha}{1 - \beta}(Q^{pr})^{1-\beta} = I + \alpha^{\frac{1}{\beta}}\left(\frac{\beta}{1 - \beta}\right)\left[\frac{\phi\mu + \nu(\phi - \beta)}{(1 - \beta)(\phi - \beta)}\right]^{1-\frac{1}{\beta}}. \tag{17}$$

Altogether, the total social welfare (defined as the sum of consumer utility, merchant and card network profits) is given by

$$W^{pr} = U^{pr} + \Pi_{m+n}^{pr} = I + \alpha^{\frac{1}{\beta}}\beta\left[\frac{\phi\mu + \nu(\phi - \beta)}{(1 - \beta)(\phi - \beta)}\right]^{1-\frac{1}{\beta}}\left[1 + \frac{1}{1 - \beta} + \frac{\mu(1 - \beta)}{\phi\mu + \nu(\phi - \beta)}\right]. \tag{18}$$

We now investigate the effects of changing the degree of competition among merchants.

Result 1. *Suppose that the card network imposes proportional card fees. Then, an increase in the number of competing merchants would*

(a) increase the profit of the card network Π_n^{pr}*,*

(b) decrease the equilibrium price P^{pr} *and increase total merchant output* Q^{pr}*,*

(c) decrease total merchant profits Π_m^{pr}*,*

(d) increase the sum of card network and merchant profits Π_{m+n}^{pr}*,*

(e) increase consumer utility U^{pr},

(f) increase social welfare W^{pr}.

Proof. Equations (13)-(18) imply that $\partial \Pi_n^{pr}/\partial\phi > 0$, $\partial P^{pr}/\partial\phi < 0$, $\partial Q^{pr}/\partial\phi > 0$, $\partial \Pi_m^{pr}/\partial\phi < 0$, $\partial \Pi_{m+n}^{pr}/\partial\phi > 0$, $\partial U^{pr}/\partial\phi > 0$ and $\partial W^{pr}/\partial\phi > 0$. ■

The findings in Result 1 can be explained as follows. Although this paper investigates a two-sided payment card market, much of the intuition can be derived from a one-sided market analogy. The card network can be viewed as an upstream firm that charges the downstream merchants a proportional fee f_m for card services. The downstream merchants then take f_m as given, and sell the final goods to consumers at price P. When $\phi = 1$, both the upstream and downstream markets are monopolized, in which case we have the classic "double marginalization" problem. As ϕ increases, the downstream market becomes more competitive, which mitigates the double marginalization effect. Consequently, this decreases the profit of downstream merchants but increases the profit of the upstream card network. Altogether, the sum of card network and merchant profits are higher, so are consumer utility and social welfare.

4 Fixed Per-transaction Card Fees

Suppose now the card network sets fixed per-transaction fees that are independent of the price of goods. Again, without loss of generality, we set $f_c = 0$ and the consumer demand for card goods becomes

$$P = \alpha Q^{-\beta}.$$

The cost borne by each merchant is given in the bottom row of (5).

In Stage II, each merchant i ($i = 1, 2, \ldots, \phi$) takes q_{-i} as given and chooses output level q_i to solve

$$\max_{q_i} \pi_i = P q_i - (\mu + f_m) q_i = \left[\alpha (q_i + q_{-i})^{-\beta} - \mu - f_m\right] q_i. \tag{19}$$

The first-order condition for a maximum yields

$$\alpha(q_{-i}+q_i)^{-\beta}\left(1-\frac{\beta q_i}{q_{-i}+q_i}\right)=\mu+f_m.$$

In a symmetric Nash-Cournot equilibrium $q_i = q$ for each merchant i. Therefore, the total merchant output, $Q^{fix} = \phi q$, and market price P^{fix} are

$$Q^{fix}=\left[\frac{\phi(\mu+f_m)}{\alpha(\phi-\beta)}\right]^{-\frac{1}{\beta}} \quad \text{and} \quad P^{fix}=\frac{\phi(\mu+f_m)}{\phi-\beta}, \tag{20}$$

where the superscript "fix" denotes for fixed per-transaction card fees.

In Stage I, the card network takes the merchant equilibrium (20) as given and chooses the merchant card fee f_m to solve

$$\max_{f_m} \Pi_n^{fix}=(f_m-\nu)Q=(f_m-\nu)\left[\frac{\phi(\mu+f_m)}{\alpha(\phi-\beta)}\right]^{-\frac{1}{\beta}}, \tag{21}$$

yielding the merchant card fee

$$f_m^{fix}=\frac{\beta\mu+\nu}{1-\beta}. \tag{22}$$

Therefore, the equilibrium profit of the card network is

$$\Pi_n^{fix}=\left[\frac{\beta(\mu+\nu)}{1-\beta}\right]\left[\frac{\phi(\mu+\nu)}{\alpha(\phi-\beta)(1-\beta)}\right]^{-\frac{1}{\beta}}. \tag{23}$$

The equilibrium price and total merchant output can be computed from (20) to be

$$P^{fix}=\frac{\phi(\mu+\nu)}{(\phi-\beta)(1-\beta)} \quad \text{and} \quad Q^{fix}=\left[\frac{\phi(\mu+\nu)}{\alpha(\phi-\beta)(1-\beta)}\right]^{-\frac{1}{\beta}}. \tag{24}$$

The total profit earned by all merchants is

$$\Pi_m^{fix}=\phi\pi_i^{fix}=(P^{fix}-\mu-f_m)Q^{fix}=\beta\left(\frac{\alpha}{\phi}\right)^{\frac{1}{\beta}}\left[\frac{\mu+\nu}{(1-\beta)(\phi-\beta)}\right]^{1-\frac{1}{\beta}}. \tag{25}$$

Adding up (23) and (25) yields the sum of card network and merchant profits

$$\Pi_{m+n}^{fix} = \Pi_m^{fix} + \Pi_n^{fix} = \beta(1-\beta+\phi)\left(\frac{\alpha}{\phi}\right)^{\frac{1}{\beta}}\left[\frac{\mu+\nu}{(1-\beta)(\phi-\beta)}\right]^{1-\frac{1}{\beta}}. \tag{26}$$

In view of consumers' utility function (1) and (2), equilibrium consumer utility is

$$U^{fix} = I - P^{fix}Q^{fix} + \frac{\alpha}{1-\beta}(Q^{fix})^{1-\beta} = I + \alpha^{\frac{1}{\beta}}\left(\frac{\beta}{1-\beta}\right)\left[\frac{\phi(\mu+\nu)}{(1-\beta)(\phi-\beta)}\right]^{1-\frac{1}{\beta}}. \tag{27}$$

Summing up, the total social welfare is given by

$$W^{fix} = U^{fix} + \Pi_{m+n}^{fix} = I + \alpha^{\frac{1}{\beta}}\beta\left[\frac{\phi(\mu+\nu)}{(1-\beta)(\phi-\beta)}\right]^{1-\frac{1}{\beta}}\left(1 + \frac{1}{1-\beta} + \frac{1-\beta}{\phi}\right). \tag{28}$$

Finally, note that Result 1 for the proportional card fee case applies also to the fixed per-transaction fee case, and will therefore not be repeated here. To see this note that equations (23)–(28) imply that $\partial\Pi_n^{fix}/\partial\phi > 0$, $\partial P^{fix}/\partial\phi < 0$, $\partial Q^{fix}/\partial\phi > 0$, $\partial\Pi_m^{fix}/\partial\phi < 0$, $\partial\Pi_{m+n}^{fix}/\partial\phi > 0$, $\partial U^{fix}/\partial\phi > 0$ and $\partial W^{fix}/\partial\phi > 0$.

5 Comparing Proportional and Fixed Card Fees

We now approach our main investigation which is to explain why card networks impose proportional fees rather than fixed per-transaction fees.

Result 2.

(a) The card network earns higher profits by charging proportional fees compared with fixed per-transaction fees. Formally, $\Pi_n^{pr} > \Pi_n^{fix}$.

(b) An increase in the number of competing merchants decreases the card network's gains from using proportional fees relative to fixed per-transaction fees. Formally, $\Pi_n^{pr} - \Pi_n^{fix}$ *decreases with the number of merchants* ϕ.

Proof. (a) Equations (13) and (23) imply $\Pi_n^{pr} - \Pi_n^{fix} > 0$; (b) Follows from (a) by computing $\partial(\Pi_n^{pr} - \Pi_n^{fix})/\partial\phi < 0$. ■

Result 2(a) explains why card networks prefer proportional fees to fixed per-transaction fees, and 2(b) suggests that card networks' incentive to use proportional fees increases with the concentration of merchant markets. These results constitute the main motivation for this research. We now investigate how the two types of card fees affect merchant profits, consumer utility and social welfare.

Result 3.

(a) Merchants earn lower profits under proportional card fees compared with fixed per-transaction fees. Formally $\Pi_m^{pr} < \Pi_m^{fix}$.

(b) For a sufficiently large number of competing merchants, a further increase of merchant competition decreases the difference of merchant profits under proportional fees and fixed per-transaction fees. Formally, there exists $\phi^* > 1$ such that for all $\phi > \phi^*$, $\Pi_m^{fix} - \Pi_m^{pr}$ *decreases with the number of merchants* ϕ.

Proof. (a) Equations (15) and (25) imply $\Pi_m^{pr} - \Pi_m^{fix} < 0$; (b) Follows from (a) by computing $\partial(\Pi_m^{fix} - \Pi_m^{pr})/\partial\phi < 0$ for $\phi > \phi^*$, where ϕ^* solves $(\mu+\nu)^{1-\frac{1}{\beta}}(\phi^* - 1)(\mu + \nu - \frac{\beta}{\phi^*}\nu)^{1+\frac{1}{\beta}} = \mu(1-\beta)(\mu\phi^* + \nu\phi^* - \mu - \nu\beta)$. ■

Result 3(a) explains why merchant associations complain about proportional fees and lobby for fixed per-transaction fees. Result 3(b) suggests that merchants' incentives to lobby for fixed per-transaction fees tend to increase with market concentration.

Results 2 and 3 are illustrated in Figure 3. The figure shows that the card network earns a higher profit under proportional fees compared with fixed per-transaction fees, but merchants earn lower profits. As the number of competing merchants increases, card network profit increases and merchant profits decrease under both types of fees. In addition, for both the card network and merchants, the difference in profits between the two types of fees decreases with the number of merchants.

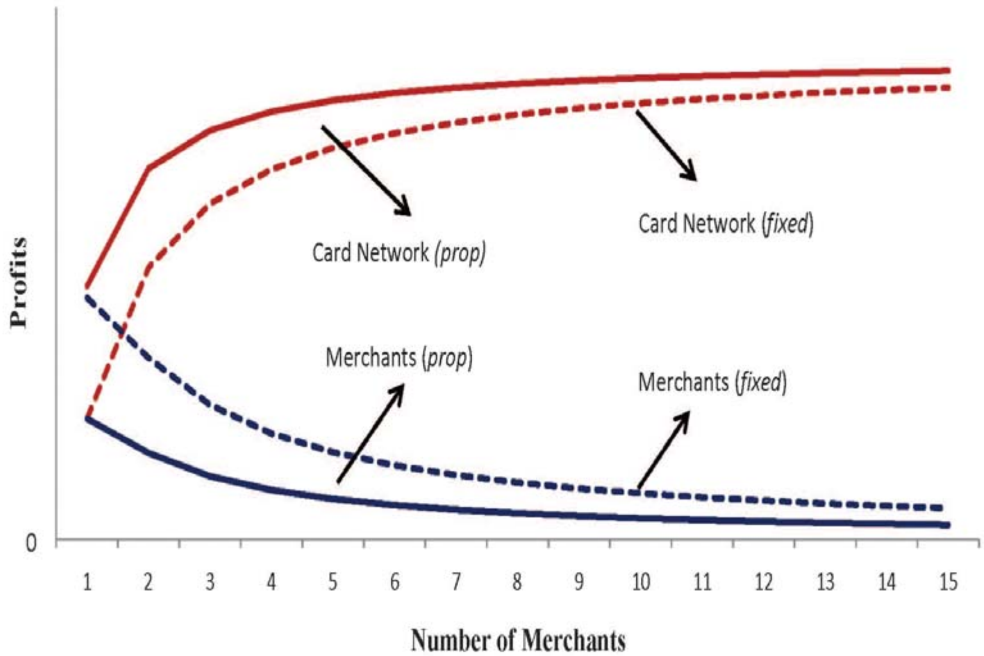

Figure 3: Card network and merchant profits under proportional and fixed fees

Next, Result 4 provides welfare comparisons between proportional fees and fixed per-transaction fees.

Result 4.

(a) The sum of card network and merchant profits is higher under proportional card fees compared with fixed per-transaction fees. Formally, $\Pi^{pr}_{m+n} > \Pi^{fix}_{m+n}$.

(b) The equilibrium price is lower and total merchant output is higher under proportional fees. Formally, $P^{pr} < P^{fix}$ *and* $Q^{pr} > Q^{fix}$.

(c) Consumer utility is higher under proportional fees. Formally, $U^{pr} > U^{fix}$.

(d) Social welfare is higher under proportional fees. Formally, $W^{pr} > W^{fix}$.

(e) The above differences between proportional fees and fixed per-transaction fees diminish with the number of competing merchants. Formally, $\Pi^{pr}_{m+n} - \Pi^{fix}_{m+n}$, $P^{fix} - P^{pr}$, $Q^{pr} - Q^{fix}$, $U^{pr} - U^{fix}$ *and* $W^{pr} - W^{fix}$ *all decrease with the number of merchants* ϕ.

Proof. (a) Follows (16) and (26). (b) Follows (14) and (24). (c) Follows (17) and (27). (d) Follows (18) and (28). (e) Follows from (a)-(d) by computing $\partial(\Pi^{pr}_{m+n} - \Pi^{fix}_{m+n})/\partial\phi < 0$,

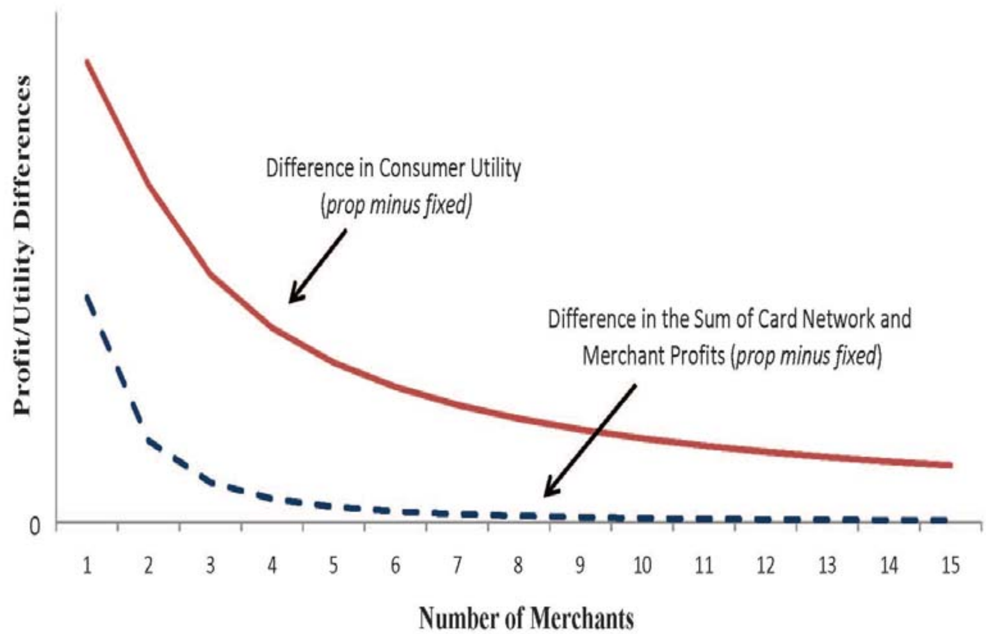

Figure 4: Profit and utility differences between proportional and fixed fees

$\partial(P^{fix} - P^{pr})/\partial\phi < 0$, $\partial(Q^{pr} - Q^{fix})/\partial\phi < 0$, $\partial(U^{pr} - U^{fix})/\partial\phi < 0$ *and* $\partial(W^{pr} - W^{fix})/\partial\phi < 0$. ■

Result 4 shows that the equilibrium market allocation under proportional fees generates a higher sum of card network and merchant profits and also generates higher consumer utility and social welfare compared with fixed per-transaction fees. Also, these differences diminish with the degree of competition among merchants, as illustrated in Figure 4. Finally, the following Result 5 shows that there is no difference between proportional fees and fixed per-transaction fees when the merchant market is perfectly competitive.

Result 5.

When the merchant market is perfectly competitive, proportional fees and fixed per-transaction fees yield the same market outcome. Formally, $\Pi_n^{pr} - \Pi_n^{fix}$, $\Pi_m^{pr} - \Pi_m^{fix}$, $\Pi_{m+n}^{pr} - \Pi_{m+n}^{fix}$, $P^{pr} - P^{fix}$, $Q^{pr} - Q^{fix}$, $U^{pr} - U^{fix}$ *and* $W^{pr} - W^{fix}$ *all converge to zero as the number of merchants* ϕ *goes to infinity.*

Proof. Follows equations (13)–(18) and (23)–(28) by computing $\lim_{\phi\to\infty}(\Pi_n^{pr} - \Pi_n^{fix}) = 0$, $\lim_{\phi\to\infty}(\Pi_m^{pr} - \Pi_m^{fix}) = 0$, $\lim_{\phi\to\infty}(\Pi_{m+n}^{pr} - \Pi_{m+n}^{fix}) = 0$, $\lim_{\phi\to\infty}(P^{pr} - P^{fix}) = 0$, $\lim_{\phi\to\infty}(Q^{pr} - Q^{fix}) = 0$, $\lim_{\phi\to\infty}(U^{pr} - U^{fix}) = 0$ *and* $\lim_{\phi\to\infty}(W^{pr} - W^{fix}) = 0$. ■

Intuitively, our findings can be explained as follows. When the merchant market is monopolized ($\phi = 1$), there are two monopolies in the vertical card payment system, the card network being the upstream monopoly and the merchant being the downstream monopoly. They both charge a markup over marginal cost, which causes the "double marginalization" problem. As it turns out, in the presence of double marginalization, the upstream monopoly can better restrain the market power of the downstream monopoly by charging proportional fees. As a result, the card network earns a higher profit at the expense of the merchant under proportional fees. Also, the sum of card network and merchant profits, consumer utility and social welfare are all higher than those under fixed per-transaction fees.

As the number of merchants ϕ increases, the downstream merchant market becomes more competitive, which mitigates the double marginalization problem. Consequently, under both types of fees, the sum of profits of upstream and downstream firms as well as consumer utility and social welfare all increase. However, the differences between proportional fees and fixed per-transaction fees diminish. Ultimately, the differences vanish as the number of merchants goes to infinity because the double marginalization problem gets eliminated.

6 Card Annual Fees

A natural extension of our analysis would be to incorporate card annual fees into the model. Some payment cards require cardholders to pay a lump-sum annual fee. In our framework, the annual fee could be viewed as a "two-part tariff" that card service providers use in order to extract a higher surplus from cardholders.

First, note that in the absence of card services, consumers' utility function (1) and (2) imply that consumers spend all their income on cash goods and gain utility $U = I$. In contrast, the availability of card services allows consumers to achieve a higher utility level. In theory, a monopoly card service provider should be able to extract the entire consumer surplus from card usage by charging cardholders a lump-sum annual fee. Incorporating card annual fees into the model implies that the card network would then maximize its overall profits, which is the sum of profits generated from merchant fees and cardholder annual fees. As before, the merchant fee could be either proportional to the transaction value or a fixed per-transaction fee, whereas the cardholder annual fee is intended to capture the remaining consumer surplus from card usage.

Extending our model to allowing the card network to charge cardholders annual fees yields the following results regarding the comparisons of proportional fees and fixed per-transaction fees: (a) Merchants earn lower profits under proportional fees than under fixed per-transaction fees. (b) For a sufficiently large card service cost ν, the card network's overall profits as well as social welfare are higher under proportional fees. (c) The market outcomes under proportional fees and under fixed per-transaction fees both converge to the socially efficient allocation as the number of merchants becomes large.

Finally, it should be noted that the card annual fee would be reduced from the monopoly level if card service providers were subjected to competition. For example, incumbent card service providers may want to keep their annual fees low in order to maintain their cardholder base and deter entry from potential competitors.

7 Additional Discussions

This paper explains why payment card networks charge fees that are proportional to the transaction values instead of charging fixed per-transaction fees. The theory shows that, when card networks and merchants both have market power, card networks earn higher profits by charging proportional fees. Merchants earn lower profits under proportional fees whereas consumer utility and social welfare are higher. These findings explain the existing payment card fee structure and shed light on related policy debates.

Our analysis shows that card networks with market power charge a markup over marginal cost, so the resulting market allocation may deviate from the social optimum. This concerns policymakers in many countries, who try to align payment card fees with the cost basis. While it may be difficult to directly regulate card fee levels, it seems natural and easy to regulate card fee structures, such as requiring fixed per-transaction fees for payment cards that incur only a fixed cost per transaction (e.g., debit cards and prepaid cards). However, our analysis suggests that if card networks have market power and merchants are perfectly competitive, then imposing a fixed per-transaction fee only would not affect the market outcome. Moreover, when merchants also have market power, a regulation like this may indeed increase merchant profits but at the expense of card networks and consumers, and is also likely to reduce total social welfare. Therefore, caution should be taken by policymakers who consider intervening in payment card markets.[14]

It should be noted that our analysis has abstracted from the argument whether card fees fulfill a special balancing purpose in two-sided markets because we believe that our findings should be useful to all researchers, those who believe that card fees play such a role, and those who argue that they don't and that all fees should be cost based.

While the paper studies the payment card industry, our analysis makes more general contribution to the broader industrial organization literature by studying the effects of pricing structures in vertically related industries, and by considering two-sided markets. Some of our results may also shed light on general public finance taxation issues, such as the efficiency of proportional tax versus flat tax.

[14] See Wang (2010b) for more discussions on regulating card fee structures versus fee levels.

References

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

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

[3] Foster, Kevin, Erik Meijer, Scott Schuh and Michael Zabek (2009), "The 2008 Survey of Consumer Payment Choice," Public Policy Discussion Paper, No. 09-10, Federal Reserve Bank of Boston.

[4] Gans, Joshua and Stephen King (2003), "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis & Policy*: 3(1), Article 1.

[5] Gerdes, Geoffrey (2008), "Recent Payment Trends in the United States," *Federal Reserve Bulletin*, 94, A75-A106.

[6] Hayashi, Fumiko (2009), "Do U.S. Consumers Really Benefit from Payment Card Rewards?" *Economic Review,* First Quarter, Federal Reserve Bank of Kansas City.

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

[8] McAndrews, James and Zhu Wang (2008), "The Economics of Two-Sided Payment Card Markets: Pricing, Adoption and Usage," Working Paper, Federal Reserve Bank of Kansas City.

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

[10] Rochet, Jean-Charles and Jean Tirole (2003), "An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems," *Review of Network*

Economics 2: 69-79.

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

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

[13] Schwartz, Marius and Daniel Vincent (2006), "The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network," *Review of Network Economics* 5: 72-102.

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

[15] Wang, Zhu (2010b), "Regulating Debit Cards: The Case of *Ad Valorem* Fees," *Economic Review,* First Quarter, Federal Reserve Bank of Kansas City.

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

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