International Journal of Industrial Organization 28 (2010) 86–98

Contents lists available at ScienceDirect

International Journal of Industrial Organization

journal homepage: www.elsevier.com/locate/ijio

Market structure and payment card pricing: What drives the interchange?☆

Zhu Wang

Economic Research Department, Federal Reserve Bank of Kansas City, 1 Memorial Drive, Kansas City, MO 64198, United States

ARTICLE INFO

Article history:
Received 7 November 2008
Received in revised form 24 June 2009
Accepted 7 July 2009
Available online 16 July 2009

JEL classification:
D4
L1
G2

Keywords:
Payment cards
Market structure
Interchange fee

ABSTRACT

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

1. Introduction

1.1. Motivation

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

Interchange fees are set by card networks. Two major card networks, Visa and MasterCard, each set their interchange fees collectively for tens of thousand member financial institutions that issue and market their cards.[2] For a simple example of how interchange functions, imagine a consumer making a $100 purchase with a credit card. For that $100 item, the retailer would get approximately $98. The remaining $2, known as the merchant discount fees, gets divided up. About $1.75 would go to the card issuing bank as interchange fees, and $0.25 would go to the merchant acquiring bank (the retailer's account provider). Interchange fees serve as a key element of the card business model and generate significant revenues for card issuers.[3] In 2007, the US card issuers made $42 billion revenue in interchange fees.

In recent years, merchants have become increasingly critical on interchange fees, claiming the fees are excessively high. They pointed out that, despite of falling costs in the card industry, interchange rates in the US have been rising over the last ten years and are among the largest and fast-growing costs of doing business for many retailers (See Fig. 1).[4] However, card networks disagree, arguing interchange fees serve the needs of all parties in the card system, including funding better consumer reward programs that could also benefit merchants.

In the meantime, many competition authorities and central banks around the world have taken action (Weiner and Wright, 2006). In Australia, the Reserve Bank of Australia mandated a sizeable reduction in credit card interchange fees in 2003. EU, UK, Belgium, Israel, Poland, Portugal, Mexico, New Zealand, Netherlands, Spain and Switzerland have made similar decisions and moves. In the US, interchange fees have been

☆ The views expressed herein are solely those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

E-mail address: zhu.wang@kc.frb.org.

[1] There are four types of general purpose payment cards in the US: (1) credit cards; (2) charge cards; (3) signature debit cards; and (4) PIN debit cards. The analysis of this paper applies to the first three types of cards, which are routed over credit card networks and account for 90% of total card purchase volume. Since our analysis focuses on the payment function but not the credit function of cards, their differences are immaterial for our purpose.

[2] Visa and MasterCard provide card services through member financial institutions (card-issuing banks and merchant-acquiring banks). They are called "four-party" systems and account for approximately 80% of the US credit card market. Amex and Discover primarily handle all card issuing and acquiring by themselves. They are called "three-party" systems and account for the remaining 20% of the market. In a "three-party" system, interchange fees are internal transfers and hence not directly observable. This paper provides a model for four-party systems, but the analysis can also be applied to three-party systems.

[3] Note that credit cards may serve two functions: payment and credit. The payment function allows cardholders to make transaction with cards and generate interchange revenues to card issuers. The credit function allows cardholders to borrow funds and generate finance revenues. While this paper focuses on card payment function and interchange revenues, we need to note that interchange fees may help increase card transaction values, so they also contribute to finance revenues for card issuers.

[4] Data sources: *Nilson Report* and *American Banker*, various issues.

0167-7187/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijindorg.2009.07.001

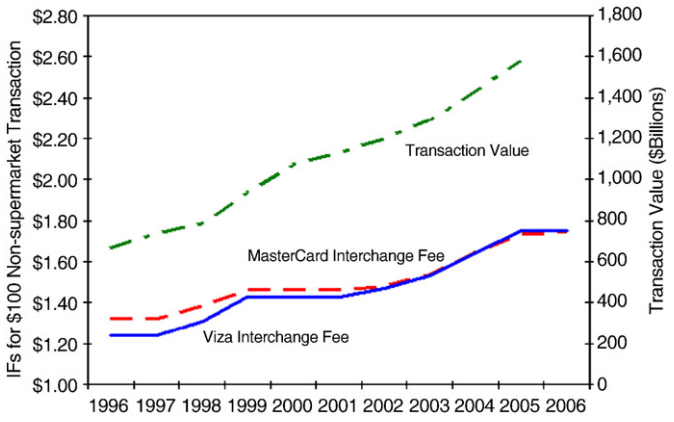

Fig. 1. Credit card interchange fees (IFs) and transaction values in the US.

mainly challenged by private litigation. Since 2005, more than 50 antitrust cases have been filed by merchants contesting interchange fees.

The performance of the card industry raises following challenging questions:

- Why have interchange fees been increasing given falling costs and increased competition in the card industry?[5]
- Given the rising interchange fees, why can't merchants refuse to accept cards? Why have total card transaction values been growing rapidly?
- What are the causes and consequences of increasing consumer card rewards?
- What are the choices and consequences of policy interventions?

In order to answer these questions, a growing literature on payment card markets has been developed recently.[6] These models, following the pioneering work of Baxter (1983), emphasize two-sided market externalities in card payment systems.[7] For example, Rochet and Tirole (2002) consider strategic interactions of consumers and merchants. In their model, two identical Hotelling merchants make card acceptance decisions to compete for consumers who have fixed demand for goods but heterogeneous benefits from using cards. Wright (2004) extends the framework by considering heterogenous merchants who receive different benefits from accepting cards. These models show that merchant card acceptance and consumer card usage depend on each other, and card networks need to set card fees to balance the demand on the two sides of the market. However, because these analyses rely on the distribution of merchant and consumer card benefits as well as the strategic competition among merchants, the results are less conclusive in terms of evaluating card market performance and explaining stylized facts (Katz, 2001; Hunt, 2003; Rochet, 2003; Rochet and Tirole, 2006).

1.2. A different approach

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

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

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

Our analysis and findings depart from the existing two-sided market theories in important ways. First, we assume free entry and exit of heterogeneous issuers, each incurring a convex cost that depends on the card transaction value. This allows us to pin down a unique equilibrium interchange fee under the Tiebout sorting of card users and cash users. This result is in contrast to previous studies, which found the neutrality of the interchange fee under the separation of card and cash payments.[9] Second, we found that the market equilibrium interchange fee is higher than or equal to the socially optimal level. Rochet and Tirole (2002) obtained similar results by considering the network externalities, which requires cash users subsidize card users under the no-surcharge rule.[10] They also show when the no-surcharge rule is lifted, the interchange fee then becomes undetermined and ceases to matter (neutrality). In contrast, our analysis is based on the entry and competition of heterogeneous issuers but not the card usage externalities, so our findings hold regardless of the no-surcharge rule. Third, unlike previous studies, we allow for an elastic consumer demand and show that the consumer demand elasticity is a key parameter that determines the equilibrium card fees. We found that the market equilibrium interchange fee tends to exceed the socially optimal level if the consumer demand is very elastic. Moreover, we found that the consumer demand elasticity also affects the impact of policy interventions. Particularly, under an interchange fee ceiling, the efficiency gains in the card industry could be distributed very differently depending on the consumer demand elasticity.[11]

[5] As shown in Fig. 2, card processing, borrowing and fraud costs have all declined, while the number of issuers and card solicitations have been rising over recent years. Data sources: Visa USA, Federal Reserve Board, Evans and Schmalensee (2005) and Frankle (2006).

[6] For example, Schmalensee (2002), Rochet and Tirole (2002), Wright (2003, 2004), Schwartz and Vincent (2006), Hayashi (2006), McAndrews and Wang (2008).

[7] Payment card systems are not the only case of such two-sided markets. Rochet and Tirole (2003) provide a detailed analysis of other examples, such as the software industry, video games, internet portals, medias, and shopping malls. In all these industries as well, the platforms may price differently to each side of the markets in order to balance the demand, while making a profit overall.

[8] Alternative payment instruments may include cash, check, PIN debit cards, stored value cards, automated clearing houses (ACH) and etc.

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

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

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