International Journal of Industrial Organization 45 (2016) 28–37

Contents lists available at ScienceDirect

International Journal of Industrial Organization

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

Price cap regulation in a two-sided market: Intended and unintended consequences☆

Zhu Wang

Research Department, Federal Reserve Bank of Richmond, USA

ARTICLE INFO

Article history:
Received 7 July 2014
Received in revised form 18 December 2015
Accepted 26 December 2015
Available online 2 January 2016

JEL classification:
D4
L5
G2

Keywords:
Price cap regulation
Two-sided market
Demand externalities

ABSTRACT

This paper studies intended and unintended consequences of price cap regulation in the two-sided payment card market. The recent U.S. debit card regulation was intended to lower merchants' card acceptance costs by capping interchange fees at the issuer cost, but for small-ticket transactions the interchange fee instead rose post-regulation. To address the puzzle, I construct a two-sided market model and show that card demand externalities between large-ticket and small-ticket transactions rationalize card networks' pricing response. Based on the model, I provide a welfare assessment of the issuer cost-based interchange regulation and discuss alternative regulatory approaches.

1. Introduction

Debit and credit cards have become an important part of the U.S. payments system and they affect a large number of consumers and merchants. Recent Federal Reserve studies show that 78% of U.S. consumers have debit cards and 70% have credit cards. In a typical month, 31% of consumer payments are made with debit cards, and 18% with credit cards (Foster et al., 2013).

However, pricing in the payment card markets has been controversial. As Rochet and Tirole (2006) pointed out, payment cards are so-called "two-sided markets," in which card networks serve two distinct end-user groups, namely, cardholders and merchants.[1] In practice, card networks and their issuers typically charge high interchange fees to merchants for card acceptance but provide rewards to consumers for card usage. Many industry observers and policymakers have become concerned that this highly skewed pricing structure may distort payments efficiency by inflating merchants' costs of accepting cards. Meanwhile, more than 20 countries have regulated or started investigating interchange fees.

In the United States, the Durbin Amendment of the Dodd–Frank Act has recently required the Federal Reserve to regulate debit card interchange fees. Under the regulation, the maximum permissible debit interchange fee for covered issuers is capped at half of its pre-regulation industry average level. As a result, covered issuers are losing billions of dollars in annual interchange revenues. However, the regulation has also generated unintended consequences. Particularly, prior to the regulation, merchants were charged differentiated interchange fees based on transaction sizes. Post-regulation, however, card networks set a uniform interchange fee at the maximum cap amount. As a result, small-ticket transactions that used to pay lower interchange fees now face an increased rate. In essence, the price cap has become a price floor.

The unintended consequence on small-ticket transactions made headlines and resulted in a lawsuit filed by several merchant groups against the Federal Reserve's debit interchange regulation.[2] This presents a puzzle: Why would card networks raise fees on small-ticket transactions in response to a fee cap? If networks used to maximize profits by charging lower fees to small-ticket transactions, it is not obvious why they would abandon that strategy in the face of a cap that is higher than the fees they used to charge.

This puzzle is not readily explained by the existing two-sided payment card market models (e.g., Rochet and Tirole, 2002, 2011; Wright, 2003,

☆ I thank the editor, Yongmin Chen, and two anonymous referees for very valuable comments and suggestions. I also thank Wilko Bolt, Huberto Ennis, Darren Filson, Boyan Jovanovic, Grace Bin Li, and participants at the Economics of Payments VI Conference hosted by the Bank of Canada, 2013 International Industrial Organization Conference, and various seminars for helpful comments and discussions. The views expressed herein are solely those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

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

[1] The research on two-sided markets recently has gained wide attention (Rysman, 2009). Other two-sided market examples include HMOs (patients and doctors), operating systems (computer users and software developers), video game consoles (gamers and game developers), and newspapers (advertisers and readers).

[2] E.g., see "Debit-Fee Cap Has Nasty Side Effect," *Wall Street Journal*, December 8, 2011.

http://dx.doi.org/10.1016/j.ijindorg.2015.12.004
0167-7187/

2012). Those theories point out that privately determined interchange fees tend to exceed the socially efficient level because of the wrong incentives at the point of sale, i.e., consumers pay the same retail price regardless of the payment instrument. However, those models typically assume card acceptance and usage are independent across transactions, so they do not predict or explain why some transactions would be adversely affected by an interchange cap that is not binding for them.

In this paper, I address this puzzle by introducing card demand externalities into a two-sided market framework. In the model, merchants engage in transactions of different sizes, and consumers' benefits from using cards at certain transaction sizes are positively affected by their card usage in others, which I call "ubiquity externalities."[3] I show that this type of demand externality drove card networks' response to the cap regulation: Prior to the regulation, card networks and issuers were willing to offer subsidized interchange fees to small-ticket transactions because their card acceptance boosted consumers' card usage for large-ticket purchases from which card issuers could collect higher interchange fees. Once a cap on interchange fees was imposed, however, card issuers profited less from this kind of externality, so they discontinued the subsidy.

Based on the model, I provide a welfare assessment of the regulation. The analysis shows that absent regulation, the market-determined interchange fees yield little total user surplus (i.e., the sum of consumer surplus and merchant profit). This explains why policymakers who care about end users (i.e., consumers and merchants) wanted the regulation in the first place. The analysis also shows that in spite of the negative impact on small-ticket transactions, the regulation may indeed improve the total user surplus by capping down interchange fees. However, an issuer cost-based regulation lacks theoretical foundation and could result in unintended consequences.[4] For one thing, such a regulation ignores the two-sided nature of the market and may run the risk of undershooting or overshooting. Especially in the latter case, the regulation could push the interchange fee too low so that a higher interchange fee may actually improve both the total user surplus and the issuer profit, and hence increase social welfare. For another, the regulation overlooks card demand externalities across different transactions, which may lead to the negative impact on small-ticket transactions that we have seen in the market. In light of the model findings, I discuss some alternative regulatory approaches.

In a nutshell, the contribution of the paper is threefold. First, the paper identifies an important puzzle of the debit interchange regulation and provides a plausible explanation motivated by empirical evidence. Second, the paper embeds the analysis in an extended two-sided market model with endogenous issuer markup, heterogenous transactions, and card demand externalities. Exploring these features yields a better understanding of the determinants of interchange fees. Finally, the paper evaluates the intended and unintended consequences of the issuer cost-based interchange cap regulation and discusses possible improvement.

The paper is organized as follows. Section 2 provides the background of the payment card industry and the debit interchange fee regulation. Section 3 lays out a two-sided payment card market model with heterogenous transactions and differentiated interchange fees. The model also allows for card demand externalities between large and small transactions. Section 4 characterizes the model equilibria with and without the interchange cap regulation. Section 5 provides a welfare assessment of the regulation and discusses alternative regulatory approaches. Section 6 concludes.

2. Industry background

Credit and debit cards have become an increasingly important part of the U.S. payments system. Recent data show that the share of their transactions in personal consumption expenditures rose to 48% in 2011. Among those, credit cards were used in 26 billion transactions for a total value of $2.1 trillion, and debit cards were used in 49 billion transactions for a total value of $1.8 trillion.[5]

Credit cards typically provide float or credit to cardholders, while debit cards directly draw from the cardholder's bank account right after each transaction. In practice, debit card payments are authorized either by the cardholder's signature or by a personal identification number (PIN). The former accounts for 60% of debit transactions, and the latter accounts for 40%.

Visa and MasterCard are the two major card networks in the United States. They provide card services through member financial institutions (issuers and acquirers) and account for 85% of the U.S. consumer credit card market.[6] Visa and MasterCard are also the primary providers of debit card services. The two networks split the signature debit market, with Visa holding 75% of the market share and MasterCard holding 25%. In contrast, PIN debit transactions are routed over a dozen PIN debit networks. Interlink, Star, Pulse, and NYCE are the top four networks, together holding 90% of the PIN debit market share. The largest PIN network, Interlink, is operated by Visa.

2.1. Interchange controversy

Along with the development of payment card markets, there has been a long-running controversy about interchange fees. Merchants are critical of the fees that they pay to accept cards. These fees are referred to as "merchant discounts," which are composed mainly of interchange fees paid to card issuers (i.e., banks issuing cards and making payments on behalf of cardholders) through merchant acquirers (i.e., banks collecting payments on behalf of merchants). Merchants believe that the card networks and issuers have wielded their market power to set excessively high interchange fees. The card networks and issuers counter that these interchange fees are necessary for covering issuers' costs as well as providing rewards to cardholders, which may also benefit merchants by making consumers more willing to use the cards.

In recent years, merchant groups launched a series of litigations against what they claim is anticompetitive behavior by the card networks and their issuers. Some of the lawsuits have been aimed directly at the interchange fees of credit and debit cards. For example, a group of class-action suits filed by merchants against Visa and MasterCard alleged that the networks violated antitrust laws by engaging in price-fixing. As a result, Visa, MasterCard, and their major issuers reached a $5.7 billion settlement agreement with U.S. retailers in December 2013, which is the largest antitrust settlement in U.S. history.

The heated debate on interchange fees has also attracted attention from researchers and regulatory authorities. On the research side, a sizeable body of literature, called "two-sided market theory," has been

[3] Ubiquity has always been a top selling point for brand cards. This is clearly shown in card networks' campaign slogans, such as Visa's "It is *everywhere* you want to be," and MasterCard's "There are some things money can't buy. For *everything* else, there's MasterCard." Ubiquity externalities may arise because of habit formation or some cost-saving motives for sticking to a single payment type: Getting buyers used to using cards for small transactions (where they may sometimes be very convenient) may also increase their demand for using them in other (including large) transactions.

[4] Two types of interchange fee regulations are currently in practice. One is based on issuers' costs, first adopted by the Reserve Bank of Australia in early 2000. The Durbin regulation in the United States is a recent example. The issuer cost-based regulation has been criticized for ignoring the two-sided nature of payment card markets. Instead, Rochet and Tirole (2011) proposed regulating the interchange fee based on merchant transaction benefit of card acceptance, which was adopted by the European Commission. The merchant benefit-based regulation addresses the two-sided market concerns, but it relies on a strong assumption that issuers set a constant markup. Moreover, neither type of regulation has considered card demand externalities across different transactions.

[5] Source: *Nilson Report*, December 2011. Prepaid cards are another type of general-purpose card but with much smaller volumes. They accounted for 2% of U.S. personal consumption expenditures in 2011.

[6] American Express and Discover are the other two credit card networks holding the remaining market share. They handle most card issuing and merchant acquiring by themselves and are called "three-party" systems. For a "three-party" system, interchange fees are internal transfers.