Working Paper Series

Demand Externalities and Price Cap Regulation: Learning from a Two-Sided Market

WP 13-06

Zhu Wang
Federal Reserve Bank of Richmond

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Demand Externalities and Price Cap Regulation: Learning from a Two-Sided Market

Zhu Wang*

May 2013
Working Paper No. 13-06

Abstract

This paper studies unintended consequences of price cap regulation in the presence of demand externalities in the context of payment cards. The recent U.S. debit card regulation was intended to lower merchant card acceptance costs by capping the maximum interchange fee. However, small-ticket merchants found their fees instead higher after the regulation. To address this puzzle, I construct a two-sided market model and show that card demand externalities across merchant sectors rationalize card networks' pricing response. Based on the model, I study socially optimal card fees and an alternative cap regulation that may avoid the unintended consequence on small-ticket merchants.

Keywords: Price cap regulation; Demand externalities; Two-sided market
JEL Classification: D4; L5; G2

---

*Research Department, Federal Reserve Bank of Richmond. Email: zhu.wang@rich.frb.org. I thank Wilko Bolt, Huberto Ennis, Darren Filson, 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. 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.

1 Introduction

The research on two-sided markets recently has gained wide attention (Rysman 2009). Two-sided markets refer to economic platforms that serve two distinct end-user groups, examples including payment cards (composed of cardholders and merchants), HMOs (patients and doctors), operating systems (computer users and software developers), and video game consoles (gamers and game developers).

For a two-sided market to prosper, the pricing structure is pivotal. As Rochet and Tirole (2006) pointed out, the transaction volume of a two-sided market crucially depends on the fees that the platform charges each side of the market, not only on the overall level of the fees. Given the complexity of two-sided markets, regulatory authorities often need to evaluate whether a privately determined pricing structure is socially beneficial, and whether a regulatory intervention is warranted to improve the market outcome.

The recent U.S. debit card regulation provides an intriguing example. In a payment card system such as Visa or MasterCard, the network sets the interchange fee, paid by the merchant's bank (i.e. the acquirer) to the cardholder's bank (i.e. the issuer) allocating the total cost of the payment service between the cardholder and the merchant. While serving an important balancing function in the two-sided market, privately determined interchange fees have been deemed too high by regulatory authorities in many countries. In the U.S., the Durbin Amendment to the Dodd-Frank Act has recently required the Federal Reserve Board of Governors to regulate debit card interchange fees. The resulting regulation limits the maximum interchange fee that a covered issuer can collect from merchants for a debit card transaction.

The regulation was intended to lower merchant card acceptance costs. Under the regulation, the maximum permissible debit interchange fee is capped at nearly half of its pre-regulation industry average level. As a result, card issuers were expected to lose multibillion-dollar annual interchange revenues to merchants. However, the regulation has also generated unintended consequences on certain merchant groups. In fact, prior to the regulation, merchants were charged differentiated interchange fees based on their sectors. Post regulation, however, card networks adjusted their pricing by setting a uniform interchange fee at the maximum cap amount. As a result, small-ticket merchants who used to pay lower interchange fees found their rates instead increased. In essence, the price cap has become a price floor.

The unintended consequence on small-ticket merchants made headlines and resulted in a pending lawsuit filed by merchant groups against the Federal Reserve's debit interchange regulation.[1] To understand the issue, an important puzzle needs to be addressed: Why would card networks abandon price differentiation in response to a cap regulation? In other words, if it was profitable to charge a lower interchange fee to small-ticket merchants in the absence of regulation, why would card networks change the practice because of a cap that is not binding for that merchant group?

This puzzle is not readily explained by the existing two-sided payment card market models (e.g. Rochet and Tirole 2002, 2011, Wright 2003, 2011). Those studies find 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 they use. However, those models typically treat merchant sectors independent from one another in terms of card acceptance and usage, so they do not predict or explain why some merchants would be adversely affected by an interchange cap that is not binding for them.

In this paper, I extend the standard two-sided market model to allow for card demand externalities across merchant sectors. In the model, merchant sectors are charged differentiated interchange fees due to their (observable) heterogenous benefits of card acceptance and usage. In addition, consumers' benefits of using cards in a merchant sector are positively affected by their card usage in other sectors, which I call "ubiquity externalities."[2] This type of demand externalities is shown to drive card networks' response to the cap regulation: Before the regulation, card networks were willing to offer subsidized interchange

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

[2] 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 from various sources. First, in the presence of a fixed adoption cost, consumers are more likely to adopt payment cards if the card is accepted by more merchants. Second, for consumers who have adopted cards, universal card acceptance may allow them to carry less cash and as a result rely more on cards for making payments. Third, universal card usage may allow card networks and issuers to collect more complete information on consumer shopping patterns, so that they can design better services to encourage further card usage (e.g. by offering more targeted card reward programs). All these ubiquity externalities, regardless of their sources, are consistent with the following analysis.

fees to small-ticket merchants because their card acceptance boosts consumers' card usage for large-ticket purchases from which card issuers can collect higher interchange fees. After the regulation, however, card issuers profit less from this kind of externalities so they discontinued the subsidy. Based on the model, I then study socially optimal card fees and alternative regulations. Consistent with previous studies, I find that privately determined interchange fees tend to exceed the socially optimal level. Moreover, I show that in the presence of card demand externalities, capping the weighted average interchange fee, instead of the maximum interchange fee, may help restore the social efficiency and avoid the unintended consequence on small-ticket merchants.

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 merchant sectors and differentiated interchange fees. The model allows for card demand externalities across merchant sectors. Section 4 characterizes the model equilibria with and without the interchange cap regulation. Section 5 discusses socially optimal interchange fees and an alternative cap regulation. Section 6 provides concluding remarks.

2 Industry background

Credit and debit cards are two of the most popular general-purpose payment cards, which accounted for 48 percent of U.S. personal consumption expenditures 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.[3]

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

Visa and MasterCard are the two major card networks in the United States. They

[3]Source: *Nilson Report*, December 2011. Note that prepaid cards are another type of general-purpose card but with much smaller volumes. In 2011, they accounted for 2 percent of U.S. personal consumption expenditures.

provide card services through member financial institutions (issuers and acquirers) and account for 85 percent of the U.S. consumer credit card market.[4] Visa and MasterCard are also the primary providers of debit card services. The two networks split the signature debit market, with Visa holding 75 percent of the market share and MasterCard holding 25 percent.[5] In contrast, PIN debit transactions are routed over PIN debit networks. Interlink, Star, Pulse and NYCE are the top four networks, together holding 90 percent 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 the "merchant discounts," which are composed mainly of interchange fees paid to card issuers through merchant acquirers. 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 litigation against what they claim is anticompetitive behavior by the card networks and their issuers. Some of the lawsuits have been aimed directly at interchange fees of credit and debit cards. For example, a group of class-action suits filed by merchants against Visa and MasterCard in 2005 alleged that the networks violated antitrust laws by engaging in price-fixing. As a result, Visa and MasterCard recently agreed to a $7.25 billion settlement with U.S. retailers, which could be the largest antitrust settlement in U.S. history.[6]

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-

[4] 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.

[5] Discover has recently entered the signature debit market, but its market share is small.

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

sided market theory," has been developed to evaluate payment card market competition and pricing issues.[7] On the regulatory side, three bills restricting interchange fees were introduced in Congress shortly before the Durbin Amendment was passed.[8] Similar trends are also taking place in many other countries. More than 20 countries and areas around the world have regulated or started investigating interchange fees.[9]

2.2 Durbin regulation

In 2010, an amendment sponsored by Sen. Dick Durbin was added to the Dodd-Frank bill, which was passed and signed into law in July 2010. The Durbin Amendment to the Dodd-Frank Act directs the Federal Reserve Board to ensure that debit card interchange fees are "reasonable and proportional to the cost incurred by the issuer with respect to the transaction." The Federal Reserve Board thereafter issued Regulation II (Debit Card Interchange Fees and Routing), which went into effect on October 1, 2011.

The new regulation establishes a cap on the debit interchange fees that banks with more than $10 billion in assets can collect from merchants through merchant acquirers. The permissible fees were set based on the Fed's evaluation of issuers' costs associated with debit card processing, clearance and settlement. The resulting interchange cap is composed of the following: A base fee of 21 cents per transaction to cover the issuer's processing costs, a five basis point adjustment to cover potential fraud losses, and an additional 1 cent per transaction to cover fraud prevention costs if the issuer is eligible. This cap applies to both Signature and PIN debit transactions.

The regulation has a major impact on card issuers' interchange revenues. According to a recent Federal Reserve study, the average debit card transaction in 2009 was approximately $40. Based on the regulation, the interchange fee applicable to a typical debit

[7] For example: Baxter (1983), Carlton and Frankel (1995), Katz (2001), Schmalensee (2002), Rochet and Tirole (2002, 2006, 2011), Gans and King (2003), Wright (2003, 2004, 2010, 2012), Cabral (2005), Armstrong (2006), Schwartz and Vincent (2006), Rysman (2007, 2009), Bolt and Chakravorti (2008), Robin Prager et al. (2009), Rochet and Wright (2010), Wang (2010), Weyl (2010), Shy and Wang (2011), and McAndrews and Wang (2012).

[8] The three bills are a House version of the Credit Card Fair Fee Act of 2009, a Senate version of the same act, and the Credit Card Interchange Fees Act of 2009.

[9] Examples include Argentina, Australia, Austria, Brazil, Canada, Chile, China, Colombia, Denmark, European Union, France, Hungary, Israel, Mexico, New Zealand, Norway, Panama, Poland, Portugal, South Africa, South Korea, Spain, Switzerland, Turkey, and United Kingdom.

card transaction would be capped at 24 cents (21 cents + (\$40 $\times$ .05%) + 1 cent), which is about half of its pre-regulation industry average level. As a result, card issuers were expected to lose an estimated \$8.5 billion annual interchange revenues.[10]

In response to the reduced interchange revenues, many card issuing banks have cut back their debit reward programs and free checking services. A recent Pulse debit issuer study shows that 50 percent of regulated debit card issuers with a reward program ended their programs in 2011, and another 18 percent planned to do so in 2012.[11] Meanwhile, the Bankrate's 2012 Checking Survey shows that the average monthly fee of noninterest checking accounts rose by 25 percent compared with last year, and the minimum balance for free-checking services rose by 23 percent.[12] Several major banks including Bank of America, Wells Fargo, and Chase attempted to charge a monthly debit card fee to their customers, though they eventually backed out due to customer outrage.[13]

2.3 Small-ticket effect

Merchants as a whole have greatly benefited from the reduced debit interchange fees under the regulation.[14] However, the distribution of the benefits is quite uneven. Particularly, an unintended consequence quickly surfaced: Small-ticket merchants find their interchange fees higher after the regulation.

Prior to the regulation, Visa, MasterCard, and most PIN networks offered discounted debit interchange fees for small-ticket transactions as a way to encourage card acceptance by merchants specializing in those transactions.[15] For instance, Visa and MasterCard set the small-ticket signature debit interchange rates at 1.55 percent of the transaction value

[10] Wang (2012) provides some estimates of issuers' lost interchange revenues using Call Report data.

[11] The 2012 Debit Issuer Study, commissioned by Pulse, is based on research with 57 banks and credit unions that collectively represent approximately 87 million debit cards and 47,000 ATMs.

[12] Bankrate surveyed banks in the top 25 U.S. cities to compare the average fees associated with checking accounts in their annual Checking Account Survey.

[13] See "Banks Adding Debit Card Fees," *The New York Times*, September 29, 2011.

[14] Depending on merchant comptition, some of the benefits may be passed along to consumers through lower retail prices.

[15] Visa and MasterCard introduced small-ticket discounted interchange fees in the early 2000s. The rates were applied to merchant sectors specializing in small-ticket transactions, including Local Commuter Transport, Taxicabs and Limousine, Fast Food Restaurants, Coffee Shops, Parking Lots and Garages, Motion Picture Theaters, Video Rental Stores, Cashless Vending Machines and Kiosks, Bus Lines, Tolls and Bridge Fees, News Dealers, Laundries, Dry Cleaners, Quick Copy, Car Wash and Service Stations, etc. In October 2010, Visa expanded the program to include more merchant sectors.

plus 4 cents for sales of $15 and below. As a result, a debit card would only charge a 7 cents interchange fee for a $2 sale or 11 cents for a $5 sale. However, in response to the regulation, most card networks eliminated the small-ticket discounts, and all transactions (except those on cards issued by exempt issuers) have to pay the maximum cap rate set by the Durbin regulation.[16] For merchants selling small-ticket items, this means that the cost of accepting the same debit card doubled or even tripled after the regulation.

The increase of small-ticket interchange fees could affect a large number of transactions. According to the 2010 Federal Reserve Payments Study, debit cards were used for 4.9 billion transactions below $5, and 10.8 billion transactions between $5-$15 in 2009. The former accounts for 8.3 percent of all payment card transactions (including credit, debit, and prepaid cards), and the latter accounts for 18.3 percent. Depending on their compositions of transaction sizes, merchants in different sectors could be affected differently by the post-regulation debit interchange fees.[17] However, merchants who specialize in small-ticket transactions would be most adversely affected.

In response, many small-ticket merchants have tried to find ways to offset their higher interchange rates. Some raised prices, or chose to restrict or reject the use of debit cards.[18] Some others offer customers incentives to consolidate transactions using prepaid cards or online wallets.[19] In the meantime, a lawsuit was filed in November 2011 in federal court by three of the retail industry's largest trade associations and two retail companies against the Federal Reserve's debit interchange regulation. The lawsuit alleges that the Fed has set the interchange cap too high by including costs that were barred by the law, and "forcing small businesses to pay three times as much to the big banks on small purchases

[16] Hayashi (2013) compares the increases of interchange fees for small-ticket transactions for Visa, MasterCard, and most PIN debit networks.

[17] E.g. Shy (2012) used the data from a diary study of consumer payment choices to identify the types of merchants who are likely to pay higher or lower interchange fees under the debit regulation.

[18] Notable examples in the press include: the DVD-rental company Redbox raised rental prices from $1 to $1.20 to cover increased debit fees; USA Technologies and Apriva, two large payment facilitators in the vending industry, stopped accepting MasterCard debit cards; the fast food restaurant chain Dairy Queen asked customers to pay with cash for purchases under $10. See "Debit-Fee Cap Has Nasty Side Effect," *Wall Street Journal*, December 8, 2011.

[19] Merchants are charged one transaction fee when a customer loads the prepaid card or online wallet rather than multiple times each instance a user pays with a debit card. Notable examples in the press include coffeehouse chain Starbucks promoting in-store prepaid cards and Washington, D.C. parking operator Parkmobile offering discounts for customers who pay with an online wallet. E.g. see "Small-Ticket Retailers Squeezed By High Transaction Fees," *U.S.News & World Report*, October 26, 2012.

was clearly not the intent of the law and is further evidence that the Fed got it wrong."[20]

The unintended regulatory impact on small-ticket merchants calls for a further examination of the payment card market. According to the Federal Reserve's evaluation, debit card issuers incur a per-transaction cost around 21 cents, which exceeds the interchange fees that they charged for small-ticket transactions prior to the regulation. Considering that issuers typically do not recover costs from the cardholder side (cardholders often receive a reward rather than pay a fee for each card transaction), card issuers appeared to have subsidized small-ticket transactions. This is also the reason that card networks claim why they eliminate the small-ticket discounts under the regulation.[21] While the existing two-sided market theories have shed great light on the functioning of interchange fees, they do not explain the subsidies on small-ticket transactions prior to the Durbin regulation, nor do they explain why these subsidies were discontinued afterwards. In this paper, I address this puzzle and draw some new policy implications.

3 Model environment

I consider a payment card system composed of five types of players: consumers, merchants, acquirers, issuers, and the card network, as illustrated in Figure 1. The setup extends the standard two-sided market model, such as Rochet and Tirole (2002, 2011), to allow for card demand externalities across merchant sectors.

■ **Consumers:** There is a continuum of measure one of consumers, who purchase goods from two distinct merchant sectors h and l. In this setting, h (respectively, l) refers to the *large-ticket* (respectively, *small-ticket*) sector where merchants and consumers enjoy *high* (respectively, *low*) transaction benefits of card acceptance and usage.[22]

Consumers have inelastic demand and buy one good per sector. Within each sector,

[20] See "Merchants' Lawsuit Says Fed Failed to Follow Law on Swipe Fee Reform," *Business Wire*, November 22, 2011.

[21] According to MasterCard, "the company decided that it couldn't sustain the [small-ticket] discounts under the new rate model because the old rates had essentially subsidized the small-ticket discounts." See "Debit-Fee Cap Has Nasty Side Effect," *Wall Street Journal*, December 8, 2011.

[22] For both merchants and consumers, replacing cash with cards may reduce their handling, safekeeping and fraud expenses on payments, and the benefits typically increase with transaction values. Therefore, it is natural to assume that merchants and consumers benefit more from card usage in large-ticket transactions than in small-ticket transactions.

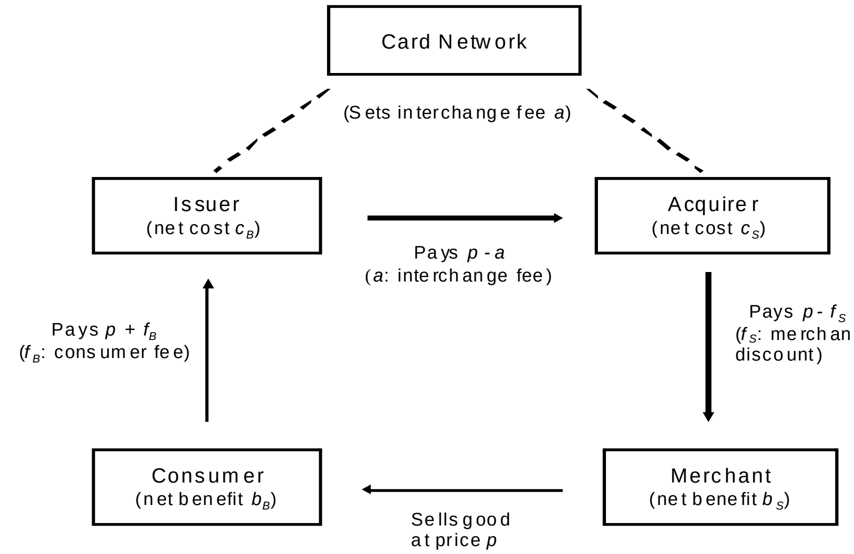

Figure 1: A Payment Card System

consumers need to decide which store to patronize. They know the stores' price and card acceptance policy before making the choice. Once in the store they then select a payment method (a card or an alternative payment method such as cash), provided that the retailer indeed offers a choice among payment means. I assume price coherence such that retailers charge the same price for purchases made by card and by cash.[23] Whenever a transaction between a consumer (buyer) and a retailer (seller) is settled by card, the buyer pays a fee f_B^i to her card issuing bank (issuer) and the seller pays a merchant discount f_S^i to her merchant acquiring bank (acquirer). These fees, f_B^i and f_S^i, depend on the merchant sector $i \in \{h, l\}$, and f_B^i is allowed to be negative, in which case the cardholder receives a reward. There are no annual fees and all consumers have a card.[24]

[23]Price coherence is the key feature that defines a two-sided market. Rochet and Tirole (2006) show that the two-sided market pricing structure (e.g. interchange fees) would become irrelevant without the price coherence condition. In reality, price coherence may result either from network rules or state regulation, or from high transaction costs for merchants to price discriminate based on payment means. In the U.S., while merchants are allowed to offer their customers discounts for paying with cash or checks, few merchants choose to do so. On the other hand, card network rules and some state laws explicitly prohibit surcharging on payment cards.

[24]For modeling purposes, I assume a representative consumer framework developed by Wright (2004) and used in the subsequent literature. Consumers are *ex ante* identical, and receive a random transaction benefit when they decide whether to pay with a card. As a result, the equilibrium number of card transactions depends on the intensive margin of card usage. Alternatively, the model could use the framework developed by Rochet and Tirole (2002) and assume heterogenous consumers who differ systematically

A consumer's transaction benefit of paying by card relative to using cash is a random variable b_B^i drawn from a cumulative distribution function H_i on the support $[\underline{b}_B^i, \overline{b}_B^i]$. I allow H_i to be heterogenous across merchant sectors. Denote μ^i as the mean of H_i and naturally $\mu^h > \mu^l$. In addition, I assume that a consumer's average transaction benefit of using a card for a large-ticket purchase μ^h is positively affected by her card usage in the small-ticket sector y^l, i.e. $d\mu^h/dy^l > 0$. (I define $y^l = q^l\chi_l$, where χ_l indicates whether l-sector merchants accept cards and q^l is a consumer's frequency of card usage in the l sector conditional on cards being accepted). This assumption captures the idea that ubiquity externalities shift up consumers' valuation of paying with cards in the h sector.[25] More specifically, I assume H_h to be a uniform distribution on the support $[\mu^h(y^l) - \gamma, \mu^h(y^l) + \gamma]$, while H_l is a degenerate distribution taking a single value μ^l.[26]

Cardholders are assumed to observe the realization of b_B^i once in the store. This is a standard assumption introduced by Wright (2004) and used in the subsequent literature. Because the net benefit of paying by card is equal to the difference $b_B^i - f_B^i$, a card payment is optimal for the consumer whenever $f_B^i \leq b_B^i$. Hence, whenever $f_B^h \leq \mu^h(y^l) + \gamma$, the proportion of card payments at an h-sector (i.e. large-ticket) store that accepts cards is

$$q^h(f_B^h) = \Pr(b_B^h \geq f_B^h) = \frac{\mu^h(y^l) + \gamma - f_B^h}{2\gamma}, \tag{1}$$

and the average net consumer benefit of paying with a card is

$$v^h(f_B^h) = E[b_B^h - f_B^h | b_B^h \geq f_B^h] = \frac{\mu^h(y^l) + \gamma - f_B^h}{2}. \tag{2}$$

Note that $q^h(f_B^h) = v^h(f_B^h) = 0$ if $f_B^h > \mu^h(y^l) + \gamma$.

---

in their transaction benefits of using cards. In that setup, the equilibrium number of card transactions depends on the extensive margin of card adoption and usage. As Rochet and Tirole (2011) show, these two alternative frameworks deliver convergent results, so the analysis and findings in this paper can be interpreted using either framework.

[25] For ease of exposition, I assume that consumers' transaction benefit of using cards in the l sector is fixed, unaffected by card usage in the h sector. However, relaxing this assumption would not change the qualitative findings.

[26] Under the assumption of H_l, consumer card usage in the l sector is a simple binary outcome, i.e. $q^l \in \{0, 1\}$. This makes it easier to model card usage externalities between the l and h sectors as a two-point function. Note that if H_l is a non-degenerate distribution, we then need to specify how card demand externalities vary by each of the multiple levels of card usage in the l sector, which significantly complicates the problem but does not provide greater qualitative intuition.

Similarly, whenever $f_B^l \le \mu^l$, the proportion of card payments at an l-sector (i.e. small-ticket) store that accepts cards is

$$q^l(f_B^l) = \Pr(\mu^l \ge f_B^l) = 1, \tag{3}$$

and the average net consumer benefit of paying with a card is

$$v^l(f_B^l) = \mu^l - f_B^l. \tag{4}$$

Note that $q^l(f_B^l) = v^l(f_B^l) = 0$ if $f_B^l > \mu^l$.

■ **Merchants:** Merchants belong to one of the two sectors, h and l. A merchant in a given sector $i \in \{h, l\}$ derives the transaction benefit b_S^i of accepting payment cards (relative to handling cash), and naturally $b_S^h > b_S^l$. Moreover, the heterogeneity between sectors is observable to the card network so that the card network can perfectly price discriminate by charging differentiated interchange fee a^i to the merchant sector i.

By accepting cards, under the price coherence assumption, a merchant is able to offer each of its card-holding customers an additional expected surplus of $q^i(f_B^i)v^i(f_B^i)$, but faces an additional expected net cost of $q^i(f_B^i)(f_S^i - b_S^i)$ per cardholder from doing so. Here, f_S^i is the sector-specific merchant discount paid to the acquirer. Therefore, a merchant accepts cards if and only if $f_S^i \le b_S^i + v^i(f_B^i)$. Rochet and Tirole (2011) show this condition holds for a variety of merchant competition setups, including monopoly, perfect competition and Hotelling-Lerner-Salop differentiated products competition with any number of retailers. Wright (2010) shows the same condition also holds for Cournot competition.

I denote χ_i as an indicator function whether merchants in sector i accept cards or not. Accordingly,

$$\chi_i = \begin{cases} 1 & \text{if } f_S^i \le b_S^i + v^i(f_B^i) \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

■ **Acquirers:** I assume acquirers incur a per-transaction cost c_S and are perfectly competitive. Thus, given the interchange fee a^i, they charge a sector-specific merchant discount f_S^i such that

$$f_S^i = a^i + c_S. \tag{6}$$

Because acquirers are competitive, they play no role in the analysis except passing through the interchange charge to merchants.

■ **Issuers:** There are $n \geq 1$ issuers who have market power.[27] Issuers incur a per-transaction cost c_B and receive an interchange payment of a^i in a card transaction. I consider a symmetric oligopolistic equilibrium at which all issuers charge the same consumer fee f_B^i, which can be negative if cardholders receive a reward.

As pointed out in Rochet and Tirole (2002, 2011), there are various ways to model issuer competition. To be concrete, I assume an explicit setting: Issuers coordinate on their pricing for small-ticket transactions (i.e. the l sector) where they make a loss, but engage in a Cournot competition for large-ticket transactions (i.e. the h sector) where they make a profit. The former assumption simplifies the setting of small-ticket card fees in order to focus on the card demand externalities, while the latter assumption allows endogenizing issuers' markup for large-ticket transactions.[28] Note that when $n = 1$, the model becomes a special case where there is a monopoly issuer.

For small-ticket transactions, issuers take the interchange fee a^l as given and set the consumer fee f_B^l to maximize their total profit conditional on merchants accepting cards (i.e. $\chi_l = 1$):

$$\hat{\Pi}^l = \max_{f_B^l} (f_B^l + a^l - c_B) q^l, \tag{7}$$

$$s.t.\ q^l = \begin{cases} 1 & \text{if } f_B^l \leq \mu^l \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where (8) follows (3). Therefore, whenever $q^l = 1$, issuers choose the highest possible

[27] This is a standard assumption in the literature. As pointed out in Rochet and Tirole (2002), the issuer market power may be due to marketing strategies, search costs, issuer reputation or the nature of the card. Note that were the issuing side perfectly competitive, issuers and card networks would have no preference over the interchange fee, and so the latter would be indeterminate.

[28] This is an extension to the existing literature which typically assumes a constant issuer markup in the spirit of Hotelling competition (e.g. Rochet and Tirole 2011, Rochet and Wright 2010). It is straightforward to show that the findings in this paper hold for the alternative and simpler setup where issuers set a constant markup for large-ticket transactions.

consumer fee such that

$$f_B^l = \mu^l, \tag{9}$$

and the total issuers' profit in the l sector is

$$\hat{\Pi}^l = \mu^l + a^l - c_B. \tag{10}$$

For large-ticket transactions, issuers engage in a Cournot competition if merchants accept cards (i.e. $\chi_h = 1$). Each issuer j sets the output level q_j^h taking the output by competing issuers, $q_{-j}^h = q^h - q_j^h$, as given and maximizes profit:

$$\hat{\pi}_j^h = \max_{q_j^h} (f_B^h + a^h - c_B) q_j^h, \tag{11}$$

$$s.t.\ f_B^h = \mu^h(y^l) + \gamma - 2\gamma(q_j^h + q_{-j}^h) \tag{12}$$

where (12) follows Eq (1). In a symmetric equilibrium, the total card usage q^h and the consumer fee f_B^h are pinned down as follows:

$$q^h = nq_j^h = \frac{n}{2\gamma(n+1)}[\mu^h(y^l) + \gamma + a^h - c_B], \tag{13}$$

$$f_B^h = \frac{1}{n+1}[\mu^h(y^l) + \gamma + n(c_B - a^h)], \tag{14}$$

and the total issuers' profit in the h sector is

$$\hat{\Pi}^h = \frac{n[\mu^h(y^l) + \gamma + a^h - c_B]^2}{2\gamma(n+1)^2}. \tag{15}$$

■ **Network:** I consider a monopoly network that sets sector-specific interchange fees a^h and a^l to maximize the total issuers' profit, namely

$$\Pi = \max_{a^h,\, a^l} (\hat{\Pi}^l \chi_l + \hat{\Pi}^h \chi_h), \tag{16}$$

where $\hat{\Pi}^l$ and $\hat{\Pi}^h$ are given by Eqs (10) and (15) above.

Because the network maximizes the issuers' profit, it makes a decision consistent with issuers on whether to provide card services to the l sector. Therefore, $y^l = \chi_l$ always holds at equilibrium, so we can simply replace $\mu^h(y^l)$ with $\mu^h(\chi_l)$ in the following analysis.

In the welfare and policy analysis (Section 5), I will also consider an alternative regime where the network is run by a social planner who maximizes social welfare or total user surplus.

■ **Timing**: I solve for a subgame perfect Nash equilibrium of the model. The timing of the game can be summarized in the following four stages.

1. The card network sets sector-specific interchange fees a^i.
2. Issuers and acquirers set fees f_B^i and f_S^i.
3. Depending on their value of b_S^i, merchants decide whether to accept cards and set retail prices.
4. Observing which merchants accept cards and their prices, consumers decide which merchants to purchase from. Once in the store, consumers receive their draw of b_B^i and decide how to pay.

4 Model characterization

I first consider a monopoly network, which sets sector-specific interchange fees a^i to maximize the total issuers' profit. In the absence of regulation, the network solves the following problem:

$$\Pi = \max_{a^h,\, a^l} \frac{n[\mu^h(\chi_l) + \gamma + a^h - c_B]^2}{2\gamma(n+1)^2}\chi_h + (\mu^l + a^l - c_B)\chi_l \tag{17}$$

$$s.t. \quad \chi_h = \begin{cases} 1 & \text{if } a^h \leq \frac{n[\mu^h(\chi_l)+\gamma+b_S^h-c_S-c_B]}{n+2} + b_S^h - c_S \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

$$\chi_l = \begin{cases} 1 & \text{if } a^l \leq b_S^l - c_S \\ 0 & \text{otherwise} \end{cases}. \tag{19}$$

The condition (18) is derived from (2), (5), (6) and (14), while the condition (19) is derived from (4), (5), (6) and (9).

Once an interchange fee cap $\overline{a}$ is introduced by regulation, the network then solves a similar problem as above but with an additional constraint:

$$a^i \leq \overline{a} \text{ for } i \in \{h, l\}. \tag{20}$$

To help characterize the model equilibrium, I make three basic assumptions on parameter values.

Assumption A1.

$$Z^h(\chi_l) = b_S^h + \mu^h(\chi_l) + \gamma - c_B - c_S > 0 \text{ for } \chi_l \in \{0, 1\}.$$

The first assumption states that the maximum merchant-and-consumer joint transaction benefit of using cards in the h sector net of costs is always positive. As will be shown, this ensures that issuers earn a positive profit for serving card transactions in the h sector.

Assumption A2.

$$Z^l = b_S^l + \mu^l - c_B - c_S < 0.$$

The second assumption states that the merchant-and-consumer joint transaction benefit of using cards in the l sector net of costs is negative. As will be shown, this implies that card issuers make a loss for serving card transactions in the l sector *per se*.

Assumption A3.

$$\mu^h(1) - \mu^h(0) = Z^h(1) - Z^h(0) > \frac{\gamma(n+2)^2(-Z^l)}{2n[Z^h(0) + Z^h(1)]}.$$

The third assumption states that card demand externalities are sufficiently large between the l and h sectors. As will be shown, this ensures that in the absence of regulation, the card network would charge differentiated interchange fees to serve card transactions in both the h and l sectors.

Under the above assumptions, I first characterize the model equilibrium in the absence of regulation. The findings are shown by the following proposition.

Proposition 1 *Under Assumptions A1-A3, an unregulated card network which maximizes total issuers' profit sets differentiated interchange fees such that cards are used in both the h and l sectors.*

Proof. Consider three options for the card network. First, when only the h sector is served with card services (i.e. $\chi_h = 1, \chi_l = 0$), the card network maximizes the total issuers' profit (17) by setting the h-sector interchange fee such that the constraint (18) is binding

$$a^h(\chi_l = 0) = b_S^h - c_S + \frac{n}{n+2} Z^h(0). \tag{21}$$

As a result, the total number of card transactions is

$$q^h = \frac{n}{\gamma(n+2)} Z^h(0), \tag{22}$$

and the total issuers' profit is

$$\Pi^h = \frac{2n}{\gamma(n+2)^2} [Z^h(0)]^2. \tag{23}$$

Under Assumption A1, this implies that $q^h > 0$ and $\Pi^h > 0$.

Second, when only the l sector is served with the card services (i.e. $\chi_h = 0, \chi_l = 1$), the card network maximizes the issuers' profit (17) by setting the l-sector interchange fee

$$a^l = b_S^l - c_S. \tag{24}$$

Under Assumption A2, the total issuers' profit is

$$\Pi^l = Z^l = b_S^l + \mu^l - c_S - c_B < 0. \tag{25}$$

Finally, when both the h and l sectors are served with card services (i.e. $\chi_h = \chi_l = 1$), the card network maximizes the issuers' profit (17) by charging differentiated interchange fees to the two sectors:

$$a^h(\chi_l = 1) = b_S^h - c_S + \frac{n}{n+2} Z^h(1), \tag{26}$$

$$a^l = b_S^l - c_S. \tag{27}$$

The resulting total issuers' profit is

$$\Pi^{h+l} = \frac{2n}{\gamma(n+2)^2}[Z^h(1)]^2 + Z^l. \tag{28}$$

Comparing Eqs (21), (26) and (27), it is found that the interchange fee is always higher in the h sector than the l sector, i.e.

$$a^h(\chi_l = 1) > a^h(\chi_l = 0) > a^l \tag{29}$$

given that $b_S^h > b_S^l$ and $Z^h(1) > Z^h(0) > 0$. Comparing (23) and (28), it is also verified that $\Pi^{h+l} > \Pi^h$ iff Assumption A3 holds. Therefore, under Assumptions A1-A3, the card network charges differentiated interchange fees given by (26) and (27) and serves card transactions in both the h and l sectors. ■

In comparison, I now characterize the model equilibrium under the interchange cap regulation. Under the regulation, the card network needs to solve the problem (17) subject to the cap constraint (20) in addition to (18)-(19). The goal here is to derive conditions that rationalize the card network's pricing response to the cap regulation as seen in the market. Namely, under the regulation, the card network charges a single interchange fee exactly at the cap level $\overline{a}$. As a result, merchants in the h sector continue to accept card, but merchants in the l sector do not.

Recall Eq (29) that $a^h(\chi_l = 1) > a^h(\chi_l = 0) > a^l$. For the purpose stated, I consider a cap level $\overline{a}$ that satisfies $a^h(\chi_l = 0) \geq \overline{a} > a^l$. This ensures that the cap is binding for the h sector regardless of whether or not the l sector is served with card services.[29] I now

[29] Note that if the cap value $\overline{a}$ is set at a level such that $a^h(\chi_l = 1) > \overline{a} > a^h(\chi_l = 0)$, the cap would not be binding for the h sector if the l sector is dropped out of the card services. The case could be a theoretical possibility, but is less relevant for explaining the market reality.

establish the following proposition.

Proposition 2 *Given any interchange cap $\overline{a}$ that satisfies $a^h(\chi_l = 0) \geq \overline{a} > a^l$, the card network sets a single interchange fee at $\overline{a}$ such that cards are used only in the h sector if the following condition holds*

$$\mu^h(1) - \mu^h(0) < \frac{2\gamma(n+1)^2(-Z^l)}{n[Z^h(1) + \frac{3n+2}{n+2}Z^h(0)]}. \tag{A4}$$

Proof. Given that $a^h(\chi_l = 0) \geq \overline{a} > a^l$, the cap $\overline{a}$ is binding for the h sector regardless of whether or not the l sector is served with card services. Therefore, Eqs (13) and (14) imply that

$$q^h = \frac{n}{2\gamma(n+1)}[\mu^h(\chi_l) + \gamma + \overline{a} - c_B],$$

$$f_B^h = \frac{1}{n+1}[\mu^h(\chi_l) + \gamma + n(c_B - \overline{a})].$$

If both the h and l sectors are served (i.e. $\chi_h = \chi_l = 1$), the total issuers' profit is

$$\Pi^{h+l} = (f_B^h + \overline{a} - c_B)q^h + Z^l = \frac{n}{2\gamma(n+1)^2}[\mu^h(1) + \gamma + \overline{a} - c_B]^2 + Z^l.$$

In contrast, if only the h sector is served (i.e. $\chi_h = 1$, $\chi_l = 0$), the total issuers' profit is

$$\Pi^h = (f_B^h + \overline{a} - c_B)q^h = \frac{n}{2\gamma(n+1)^2}[\mu^h(0) + \gamma + \overline{a} - c_B]^2.$$

Therefore, $\Pi^{h+l} < \Pi^h$ iff

$$\mu^h(1) - \mu^h(0) < \frac{2\gamma(n+1)^2(-Z^l)}{n[\mu^h(1) + \mu^h(0) + 2\gamma + 2\overline{a} - 2c_B]}. \tag{30}$$

Because $\overline{a} \leq a^h(\chi_l = 0)$, a sufficient condition for (30) to hold is that

$$\mu^h(1) - \mu^h(0) < \frac{2\gamma(n+1)^2(-Z^l)}{n[\mu^h(1) + \mu^h(0) + 2\gamma + 2a^h(\chi_l = 0) - 2c_B]}. \tag{31}$$

Inserting the expression of $a^h(\chi_l = 0)$ from Eq (21), the condition (31) can then be rewritten as (A4).

Under Assumptions A1-A2, it is straightforward to verify that

$$\frac{2\gamma(n+1)^2(-Z^l)}{n[Z^h(1)+\frac{3n+2}{n+2}Z^h(0)]} > \frac{\gamma(n+2)^2(-Z^l)}{2n[Z^h(0)+Z^h(1)]}.$$

Therefore, there exists a non-empty set of values that satisfy Assumption A3 and Condition A4. Hence, for any value of $\mu(1)-\mu(0)$ within that set, the card network sets differentiated interchange fees to serve both the h and l sectors in the absence of regulation, and sets a single interchange fee at $\overline{a}$ such that only the h sector is served with the card services under the cap regulation. ■

5 Welfare and policy analysis

I have provided a model that rationalizes card networks' interchange pricing before and after the cap regulation introduced by the Durbin Amendment. The analysis suggests that card demand externalities between the small-ticket and large-ticket sectors could play an important role in explaining card networks' response to the regulation. Based on the model framework, I now take a step further to conduct welfare and policy analysis.

5.1 Welfare maximization

I first consider an alternative regime where the network is run by a social planner who maximizes social welfare.[30] Social welfare is generated whenever consumers use cards for payment at retailers provided consumer-and-merchant joint transaction benefits exceed the joint costs (i.e., $b_S^i + b_B^i > c_B + c_S$), which is shown as

$$\sum_{i\in\{h,l\}} \left(\chi_i \int_{f_B^i}^{\overline{b}_B^i} [b_S^i + b_B^i - c_B - c_S] dH_i(b_B^i) \right). \tag{32}$$

To be comparable with the analysis in the previous section, I assume that the social planner can collectively set card fees (a^l, f_B^l) for small-ticket transactions to internalize

[30] In the welfare analysis, I abstract from the concern that social costs of alternative payment means may deviate from private costs (e.g. the cash and check services are partially sponsored by the government, so social costs of providing those services may diverge from private costs). Those are interesting but separate issues, which are beyond the scope of this paper.

card demand externalities (I will show later that this outcome can indeed be implemented by an alternative interchange cap regulation). Therefore, under the model's distributional assumptions of b_B^h and b_B^l, the social planner sets card fees a^h, a^l, f_B^l to maximize social welfare as follows,

$$W = \max_{a^h, a^l, f_B^l} \frac{\chi_h}{2\gamma}\left([b_S^h - c_B - c_S][\mu^h(\chi_l) + \gamma - f_B^h] + \frac{[(\mu^h(\chi_l)+\gamma]^2 - f_B^{h2}}{2}\right) \tag{33}$$

$$+ [b_S^l + \mu^l - c_B - c_S]\chi_l$$

$$s.t.\ (9),\ (14),\ (18),\ (19).$$

The following proposition characterizes the solution to the welfare maximization problem (33). The results show that under Assumptions A1-A3, the social planner would also set differentiated interchange fees to serve both the h and l sectors, but the fee level in the h sector tends to be lower than that set by the private network.

Proposition 3 *The social planner who maximizes social welfare sets differentiated interchange fees to serve card transactions in both the h and l sectors. In addition, (i) when issuer competition is high (i.e. $n > 2$), the h-sector interchange fee set by the social planner is lower than that set by the private network; (ii) when issuer competition is low (i.e. $n \leq 2$), the h-sector interchange fee set by the social planner coincides with that set by the private network.*

Proof. Consider that the card network is run by a social planner who maximizes social welfare. In the case where issuer competition is high (i.e. $n > 2$), the constraint (18) does not bind. The first order condition with regard to f_B^h yields that

$$\tilde{f}_B^h = c_B + c_S - b_S^h. \tag{34}$$

Eqs (14) and (34) then determine the interchange fee in the h sector

$$\tilde{a}^h = b_S^h - c_S + \frac{Z^h(\chi_l)}{n}. \tag{35}$$

The social planner can also set

$$\tilde{a}^l = b_S^l - c_S \quad \text{and} \quad \tilde{f}_B^l = \mu^l \tag{36}$$

to serve the l sector. Therefore, if the social planner sets a single interchange fee and only serves the h sector, the maximum welfare is determined by (33) as

$$W^h = \frac{[Z^h(0)]^2}{4\gamma}. \tag{37}$$

In contrast, if the social planner sets differentiated interchange fees and serves both the h and l sectors, the maximum welfare is

$$W^{h+l} = \frac{[Z^h(1)]^2}{4\gamma} + Z^l. \tag{38}$$

Under Assumption A3, $W^h < W^{h+l}$, so the social planner prefers the latter.[31]

In the case where issuer competition is low (i.e. $n \leq 2$), the constraint (18) is binding. Hence,

$$\tilde{a}^h = b_S^h - c_S + \frac{n}{n+2} Z^h(\chi_l). \tag{39}$$

Eqs (39) and (14) then determine the consumer fee in the h sector

$$\tilde{f}_B^h = \frac{2n(c_S + c_B - b_S^h) - (n-2)[\mu^h(\chi_l) + \gamma]}{n+2}. \tag{40}$$

Again, the social planner can also set card fees

$$\tilde{a}^l = b_S^l - c_S \quad \text{and} \quad \tilde{f}_B^l = \mu^l \tag{41}$$

to serve the l sector. Therefore, if the social planner sets a single interchange fee and only serves the h sector, the maximum welfare is determined by (33) as

[31] For issuers to participate in the card network, they need to make a non-negative profit. This can be satisfied under plausible parameter values, i.e. $[Z^h(1)]^2 > 2\gamma n(-Z^l)$, or the social planner is allowed to conduct lump-sum transfers.

$$W^h = \frac{2n[Z^h(0)]^2}{\gamma(n+2)^2}. \tag{42}$$

In contrast, if the social planner sets differentiated interchange fees and serves both the h and l sectors, the maximum welfare is

$$W^{h+l} = \frac{2n[Z^h(1)]^2}{\gamma(n+2)^2} + Z^l. \tag{43}$$

In this case, the social planner's decision is indeed equivalent to the private network's decision given by (23) and (28). Therefore, $W^h < W^{h+l}$, so the social planner prefers the latter as well. ■

The welfare findings can be intuitively explained as follows. There are two counteracting distortions in the card payment system that we consider (particularly in the h sector where consumer transaction benefit of card usage is assumed to follow a non-degenerate distribution).[32] On the one hand, price coherence allows consumers to pay the same retail price regardless of the payment method they use. As a result, merchants internalize consumers' inframarginal card usage benefits when they decide whether to accept cards. This raises the interchange fee that merchants are willing to accept.[33] On the other hand, issuers impose a markup when setting consumer fees, which drives down the inframarginal card usage benefits and lowers the interchange fee that merchants are willing to accept. In the case where the issuers' market power is small (i.e. $n > 2$), the distortion due to price coherence dominates, so the privately determined interchange fee in the h sector exceeds the socially optimal level. In the case where the issuers' market power is large (i.e. $n \leq 2$), the distortion due to issuer markup dominates. However, because the socially optimal interchange fee is limited by the merchant card acceptance constraint, the privately determined interchange fee coincides with the social optimum.[34]

[32] Because of the assumption that consumers' transaction benefit of card usage in the l sector follows a degenerate distribution, we abstract from distortion in that sector *per se*. However, the regulator and the private network still have different objectives for internalizing cross-sector card demand externalities.

[33] As mentioned before, the analysis in this paper can be carried over to the framework of Rochet and Tirole (2002), where heterogenous consumers differ systematically in their transaction benefits of using cards. In that framework, price coherence implies that cash-paying consumers are subsidizing those who use cards.

[34] In reality, a card network typically has a large number of issuers. Therefore, it is likely that the privately determined interchange fee exceeds the socially optimal level.

In spite of the result that the privately determined interchange fee in the h sector may exceed the socially optimal level, we find that under the same set of assumptions, the social planner behaves similar to the private network by setting differentiated interchange fees to serve card transactions in both the h and l sectors. Essentially, both the social planner and the private network treat the transactions in the l sector as a loss leader. In doing so, they subsidize the l-sector card transactions in order to internalize the positive externalities of card usage between the l and h sectors.

A similar analysis can be done if we assume that the social planner maximizes total user surplus instead of social welfare. Total user surplus is the sum of consumer surplus and merchants' profit (but not issuers' profit).[35] Focusing on total user surplus is legitimate when card issuers' profit is dismissed by competition authorities. In this case, the results turn out to be even stronger. Under plausible parameter values, I again find that the social planner who maximizes total user surplus sets differentiated interchange fees to serve card transactions in both the h and l sectors. Moreover, the resulting interchange fees in both the h and l sectors are lower than those maximizing total issuers' profit or social welfare. The proof of the results can be found in the Appendix.

5.2 Alternative regulation

The analysis above suggests that privately determined interchange fees tend to be too high (based on the criterion of social welfare maximization or total user surplus maximization), a finding consistent with previous studies. This implies that payment cards could be overused at equilibrium. Therefore, lowering interchange fees may potentially improve payments efficiency, which provides some justification for the interchange cap regulation.

Moreover, Eq (35) points out that the socially optimal h-sector interchange fee is determined by multiple factors including merchant-and-consumer net benefits $Z^h(\chi_l)$, merchant transaction benefit b_S^h, issuer competition n, and the acquirers' cost c_S. The finding suggests that the Durbin regulation that requests interchange fees to be capped by issuers' marginal cost c_B may lack theoretical foundation.

[35] Maximizing total user surplus is the criterion Rochet and Tirole (2011) used to derive the optimal interchange fee regulation based on the "merchant avoided-cost test."

In addition, the analysis suggests that in the presence of card demand externalities across merchant sectors, capping the maximum interchange fee may adversely affect small-ticket merchants and yields a suboptimal outcome. According to the model, under such a regulation, the card network maximizing issuers' profit may set a uniform interchange fee and drop small-ticket merchants.

To restore the social optimum, one possibility might be to regulate all card fees (i.e. interchange fees and consumer fees) to ensure that all merchant sectors receive card services at appropriate levels. However, this approach risks being too heavy-handed. Alternatively, policymakers may be interested in the following question: In case a regulator can only set a single interchange fee cap, would it still be possible to restore the social optimum? The answer is indeed *yes*. The following discussion illustrate how this can be achieved conceptually.

Proposition 4 *In the presence of card demand externalities across merchant sectors, capping the weighted average interchange fee, instead of the maximum interchange fee, may restore the social optimum.*

Proof. Consider a regulator who maximizes social welfare (a similar analysis can be done using the criterion of total user surplus). When $n > 2$, the privately determined interchange fee $a^h(\chi_l = 1)$ given by (26) exceeds the welfare-maximizing level $\tilde{a}^h(\chi_l = 1)$ given by (35). Assume that Assumptions A1-A3 and Condition A4 hold, so both the social planner and the private network would want to serve card transactions in the h and l sectors. Following the analysis in Section 2, I focus on the scenario where $\tilde{a}^h(\chi_l = 1) < a^h(\chi_l = 0)$, that is, the welfare-maximizing h-sector interchange fee when both the h and l sectors use cards is lower than the privately determined h-sector interchange fee when only the h sector uses cards.[36] In this case, as suggested by Proposition 2, capping the maximum interchange fee would not restore the social optimum because the card network and issuers would stop subsidizing the l sector. Instead, the regulator could consider capping the weighted average interchange fee as follows.

Capping the weighted average interchange fee means that the regulator restricts in-

[36] Eqs (21) and (35) imply that $\tilde{a}^h(\chi_l = 1) < a^h(\chi_l = 0)$ whenever $Z^h(1) < \frac{n^2}{n+2} Z^h(0)$.

terchange fees chosen by the card network, a^h and a^l, such that

$$\lambda a^h + (1-\lambda)a^l \leq \bar{a}, \tag{44}$$

where $\bar{a}$ is the cap and $0 \leq \lambda \leq 1$ is the weight chosen by the regulator.

Recall that the welfare-maximizing interchange fees are given by (35) and (36) that

$$\begin{aligned} \tilde{a}^h &= \frac{Z^h(1)}{n} + b_S^h - c_S, & (45) \\ \tilde{a}^l &= b_S^l - c_S. & (46) \end{aligned}$$

Note that $\tilde{a}^h > \tilde{a}^l$ because $b_S^h > b_S^l$ and $Z^h(1) > 0$. The corresponding total issuers' profit is determined by (16), (45) and (46) as

$$\Pi^{h+l}(\tilde{a}^h, \tilde{a}^l) = \frac{n[\mu^h(1) + \gamma + \tilde{a}^h - c_B]^2}{2\gamma(n+1)^2} + Z^l. \tag{47}$$

Denote $\Pi^h(\tilde{a}^l)$ as the total issuer's profit for only serving the h sector at the interchange fee level $\tilde{a}^l = b_S^l - c_S$, so

$$\Pi^h(\tilde{a}^l) = \frac{n[\mu^h(0) + \gamma + \tilde{a}^l - c_B]^2}{2\gamma(n+1)^2}.$$

Under plausible parameter values, we have that[37]

$$\Pi^{h+l}(\tilde{a}^h, \tilde{a}^l) > \Pi^h(\tilde{a}^l) > 0. \tag{48}$$

Given that $\tilde{a}^h > \tilde{a}^l$, (48) implies that there exists a non-empty set of values that $\bar{a}$ can take, which satisfy $\tilde{a}^h > \bar{a} > \tilde{a}^l$ and

$$\Pi^{h+l}(\tilde{a}^h, \tilde{a}^l) > \Pi^h(\bar{a}) > \Pi^h(\tilde{a}^l) > 0. \tag{49}$$

The regulator can then choose any cap value $\bar{a}$ from this set and determine the corre-

[37]Note that the first inequality in (48) holds if $\left[\frac{n}{n+1}Z^h(1)\right]^2 - \left[Z^h(0) - (b_S^h - b_S^l)\right]^2 > \frac{n(-Z^l)}{2\gamma(n+1)^2}$, and the second inequality holds if $Z^h(0) > b_S^h - b_S^l$.

sponding weight λ by solving

$$\lambda\tilde{a}^h + (1-\lambda)\tilde{a}^l = \bar{a}. \tag{50}$$

Given $\bar{a}$ and λ determined by (49)-(50), if the card network sets a uniform interchange fee and only serves card transactions in the h sector, the highest interchange fee that it can set is $\bar{a}$. In contrast, if the card network sets differentiated interchange fees to serve both the h and l sectors, (50) implies that the profit-maximizing interchange fees are $\tilde{a}^h$ and $\tilde{a}^l$. As (49) shows, the card network receives a higher profit by choosing $\tilde{a}^h$ and $\tilde{a}^l$, so the outcome restores the social optimum. ■

Proposition 4 shows that capping the weighted average interchange fee may restore the social optimum. Essentially, this alternative cap regulation penalizes card networks if they set a uniform interchange fee and stop subsidizing the small-ticket transactions. Instead, card networks are allowed to set a higher-than-the-cap interchange fee to the large-ticket merchants only if they continue to subsidize the small-ticket merchants. As long as the cap and weight are appropriately chosen (cf conditions (49) and (50)), card networks are indued to set differentiated interchange fees to serve both merchant sectors in a way that coincides with the social optimum.

6 Conclusion

The recent U.S. debit card regulation introduced by the Durbin Amendment of the Dodd-Frank Act has generated some unintended consequences. While the regulation was intended to lower merchant card acceptance costs by capping the maximum interchange fee, small-ticket merchants find their fees instead higher because card networks subsequently charge the maximum cap amount on small-ticket transactions that used to pay less.

In this paper, I consider a two-sided payment card market and allow for card demand externalities across merchant sectors. In the model, merchant sectors are charged with differentiated interchange fees due to their (observable) heterogenous benefits of card acceptance and usage. In addition, consumers' benefits of using cards in a merchant sector are positively affected by their card usage in other sectors, which I call "ubiquity externalities." This type of demand externalities is shown to drive card networks' response to the cap regulation: Before the regulation, card networks were willing to offer discounted interchange fees to small-ticket merchants because their card acceptance boosts consumers' card usage for large-ticket purchases from which card issuers can collect higher interchange fees. After the regulation, however, card issuers profit less from this kind of externality, so they discontinued the discounts.

Based on the model, I then study socially desirable interchange fees and optimal interchange regulation. The analysis shows that both the social planner and the private card network would want to subsidize small-ticket transactions in order to internalize card demand externalities across merchant sectors, even though the interchange fees set by the private card network tend to exceed the socially optimal level. However, in the presence of card demand externalities, we find that simply capping the maximum interchange fee would not restore the social optimum because of the side effects on small-ticket transactions. Alternatively, a cap regulation based on the weighted average interchange fee may restore the social optimum and avoid the unintended consequence on small-ticket merchants.[38]

There are several avenues for further research. First, it would be useful to take the qualitative analysis to data and quantify the card demand externalities across merchant sectors. This would be an important step for assessing the empirical impact of current as well as alternative interchange regulations. Second, it would be useful to consider policy options other than interchange fee regulation. For instance, in theory, if merchants can set different retail prices conditioning on payment means, the interchange fee becomes less of an issue. However, those policy options may also have their own limitations, so some cautions need to be taken.[39] Finally and more broadly, the analysis can be extended to industries beyond payment cards or two-sided markets. In fact, policymakers may always want to be alert to cross-sector externalities when designing regulatory policies so that unintended consequences can be reduced or avoided.

[38] This paper proposes a theory of optimal interchange fee regulation. A successful implementation of the theory will require additional quantitative and empirical work.

[39] For example, in countries where card surcharging is allowed, few merchants choose to do so. Moreover, for some merchants who are indeed surcharging, they are found surcharging more than card acceptance costs or imposing surcharging in nontransparent ways. See Hayashi (2012).

Appendix: Total user surplus maximization

Total user surplus is generated whenever consumers use cards for payment at retailers provided consumer-and-merchant joint transaction benefits exceed the joint fees that they pay, namely $b_S^i + b_B^i > f_B^i + f_S^i$. In other words, total user surplus is the sum of consumer surplus and merchants' profit (but not issuers' profit). The expression of total user surplus can be derived from (17) and (33), as shown below. Again, I assume that the social planner can collectively set card fees $(a^l,\ f_B^l)$ for small-ticket transactions to internalize card demand externalities. Therefore, the social planner sets card fees a^h, a^l, f_B^l to maximize total user surplus as follows:

$$TUS = \max_{a^h, a^l, f_B^l} \frac{\chi_h}{2\gamma}\left([b_S^h - c_B - c_S][\mu^h(\chi_l) + \gamma - f_B^h] + \frac{[(\mu^h(\chi_l) + \gamma]^2 - f_B^{h2}}{2}\right) \tag{51}$$

$$- \frac{\chi_h}{2\gamma}\left(\frac{n[\mu^h(\chi_l) + \gamma + a^h - c_B]^2}{(n+1)^2}\right) + (b_S^l - c_S - a^l)\chi_l$$

$$s.t. \quad \frac{n[\mu^h(\chi_l) + \gamma + a^h - c_B]^2}{2\gamma(n+1)^2}\chi_h + (f_B^l + a^l - c_B)\chi_l \geq 0, \tag{52}$$

and (9), (14), (18), (19).

The constraint (52) requires that card issuers make a non-negative profit.

The following proposition characterizes the solution to the problem. The results show that under plausible parameter values, the social planner who maximizes total user surplus would also set differentiated interchange fees to serve both the h and l sectors, but the interchange fees in both sectors are lower than those maximizing total issuers' profit or social welfare.

Proposition 5 *The social planner who maximizes the total user surplus sets differentiated interchange fees to serve card transactions in both the h and l sectors, and the interchange fees in both sectors are lower than those maximizing total issuers' profit or social welfare if*

$$[Z^h(1)]^2 > \frac{2\gamma(n+2)^2(-Z^l)}{n}. \tag{A5}$$

Proof. Consider that the card network is run by a social planner who maximizes total user surplus. The constraint (18) never binds and the first order condition with regard to f_B^h yields

$$\tilde{f}_B^h = c_B + c_S - b_S^h + \frac{2}{n+2} Z^h(\chi_l). \tag{53}$$

Eqs (14) and (53) then determine the interchange fee

$$\tilde{a}^h = b_S^h - c_S - \frac{Z^h(\chi_l)}{n+2}. \tag{54}$$

Therefore, if the social planner only provides card services to the h sector but not the l sector, it is optimal to set

$$\tilde{a}^h(\chi_l = 0) = b_S^h - c_S - \frac{Z^h(0)}{n+2}. \tag{55}$$

The total user surplus is

$$TUS^h = \frac{n}{4\gamma(n+2)}[Z^h(0)]^2, \tag{56}$$

and issuers make the total profit

$$\Pi^h = \frac{n}{2\gamma(n+2)^2}[Z^h(0)]^2. \tag{57}$$

Alternatively, if the social planner also provides card services to the l sector, it is optimal to set

$$\tilde{a}^h(\chi_l = 1) = b_S^h - c_S - \frac{Z^h(1)}{n+2}, \tag{58}$$

and set the lowest interchange fee $\tilde{a}^l$ that satisfies the constraint (52):

$$\tilde{a}^l = b_S^l - c_S - \left(\frac{n}{2\gamma(n+2)^2}[Z^h(1)]^2 + Z^l \right). \tag{59}$$

The total user surplus is

$$TUS^{h+l} = \frac{n^2+4n}{4\gamma(n+2)^2}[Z^h(1)]^2 + Z^l, \tag{60}$$

and the total issuers' profit is

$$\Pi^{h+l} = 0. \tag{61}$$

Under Condition A5, $TUS^{h+l} > TUS^{h}$, so the social planner achieves higher total user surplus by setting differentiated interchange fees to serve card transactions in both the h and l sectors. Moreover, under Condition A5, Eqs (58) and (59) confirm that the interchange fees in both the h and l sectors are lower than those maximizing total issuers' profit (given by (26) and (27)) or those maximizing social welfare (given by (35) and (36) when $n > 2$ or (39) and (41) when $n \leq 2$). ■

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] Bolt, Wilko and Sujit Chakravorti (2008), "Consumer Choice and Merchant Acceptance of Payment Media," DNB Working Papers 197, Netherlands Central Bank.

[4] Cabral, Luis (2005), "Market Power and Efficiency in Card Payment Systems: A Comment," *Review of Network Economics*, 5: 15-25.

[5] Carlton, Dennis and Alan Frankel (1995), "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, 63: 903-915.

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

[7] Hayashi, Fumiko (2012), "Discounts and Surcharges: Implications for Consumer Payment Choice," *Payments System Research Briefings,* June, Federal Reserve Bank of Kansas City.

[8] Hayashi, Fumiko (2013), "The New Debit Card Regulations: Effects on Merchants, Consumers, and Payments System Efficiency," *Federal Reserve Bank of Kansas City Economic Review*, Q1: 89-118.

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

[10] McAndrews, James and Zhu Wang (2012), "The Economics of Two-Sided Payment Card Markets: Pricing, Adoption and Usage." Research Working Paper 12-08, Federal Reserve Bank of Richmond.

[11] 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.

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

[13] Rochet, Jean-Charles and Jean Tirole (2006), "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, 35: 645-667.

[14] 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.

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

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

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

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

[19] Shy, Oz (2012), "Who Gains and Who Loses from the 2011 Debit Card Interchange Fee Reform?" *Public Policy Discussion Paper,* No. 12-6, Federal Reserve Bank of Boston.

[20] 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.

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

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

[23] Wang, Zhu (2012), "Debit Card Interchange Fee Regulation: Some Assessments and Considerations," *Federal Reserve Bank of Richmond Economic Quarterly*, 98 (3): 159-183.

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

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

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

[27] Wright, Julian (2010), "Why Do Merchants Accept Payment Cards?" *Review of Network Economics*, 9, Article 1.

[28] Wright, Julian (2012), "Why Payment Card Fees Are Biased Against Retailers," *RAND Journal of Economics*, 43: 761-80.