Working Paper Series

Price Cap Regulation in a Two-sided Market: Intended and Unintended Consequences

WP 13-06R

Zhu Wang
Federal Reserve Bank of Richmond

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Price Cap Regulation in a Two-sided Market: Intended and Unintended Consequences

Zhu Wang*

December 2015
Working Paper No. 13-06R

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.

Keywords: Price cap regulation; Two-sided market; Demand externalities
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, 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.

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 percent of U.S. consumers have debit cards and 70 percent have credit cards. In a typical month, 31 percent of consumer payments are made with debit cards, and 18 percent 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 in-

[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).

terchange 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, 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

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

[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.

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.

[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.

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 percent 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 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 provide card services through member financial institutions (issuers and acquirers) and account for 85 percent 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 percent of the market share and MasterCard holding 25 percent. 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 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 "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

[5] Source: *Nilson Report*, December 2011. Prepaid cards are another type of general-purpose card but with much smaller volumes. They accounted for 2 percent 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.

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 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.[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 directs the Federal Reserve Board to regulate debit card interchange fees so that they are "reasonable and proportional to the cost incurred by the issuer with respect to the transaction." The

[7]For example: Baxter (1983), Carlton and Frankel (1995), Schmalensee (2002), Rochet and Tirole (2002, 2006, 2011), Gans and King (2003), Wright (2003, 2004, 2010, 2012), Schwartz and Vincent (2006), Bolt and Chakravorti (2008), Prager et al. (2009), Rochet and Wright (2010), Wang (2010), Weyl (2010), Shy and Wang (2011), McAndrews and Wang (2012), and Bedre-Defolie and Calvano (2013).

[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]Worldwide, more than 20 countries and areas have regulated or started investigating interchange fees. Primary examples include Australia, Canada, the European Union, France, Spain, and the United Kingdom.

Board subsequently issued Regulation II (Debit Card Interchange Fees and Routing), which took effect on October 1, 2011.

The regulation establishes a cap on the debit interchange fees that financial institutions with more than $10 billion in assets can charge to merchants through merchant acquirers. The permissible fees were set based on an 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 charge of 0.05 percent of the transaction value 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. Under the regulation, the interchange fee applicable to an average debit transaction would be capped at 24 cents (21 cents + ($40 × .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 in 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, Bankrate's 2012 Checking Survey shows that the average monthly fee of noninterest checking accounts rose by 25 percent compared with the year before, 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]

[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," *New York Times*, September 29, 2011.

2.3 Small-Ticket Effect

While merchants are supposed to benefit from the regulation, the distribution of the benefits turns out to be quite uneven. Particularly, an unintended consequence quickly surfaced: Interchange fees rose for small-ticket transactions.

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 for those transactions.[14] For instance, Visa and MasterCard set the small-ticket debit interchange rate at 1.55 percent of the transaction value plus 4 cents for sales of $15 and below. As a result, card issuers would only charge a 7-cent interchange fee for a $2 sale or 12 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.[15] 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 accounted for 8.3 percent of all payment card transactions (including credit, debit, and prepaid cards), and the latter accounted for 18.3 percent. Depending on their compositions of transaction sizes, merchants could be affected differently by the post-regulation debit interchange fees.[16] However, merchants who specialize in small-ticket transactions would be the most adversely affected. In response, many small-ticket merchants have tried to find ways to offset the higher interchange rates. Some raised prices or started to restrict the use of debit cards. Notable examples in the press include the DVD-rental company Redbox raising rental prices from $1 to $1.20 to cover increased debit fees; USA Technologies and Apriva, two large payment facilitators in the vending

[14] Visa and MasterCard introduced small-ticket discounted interchange fees in the early 2000s. The rates were initially applied to merchant sectors specializing in small-ticket transactions, such as Fast Food, Coffee Shops, Parking Lots, Video Rental, Vending Machines, Bus Lines, Dry Cleaners, Quick Copy, etc. Later on, the small-ticket rates were applied to general merchant sectors.

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

[16] E.g., Shy (2014) 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.

industry, stopping acceptance of MasterCard debit cards; the fast food restaurant chain Dairy Queen asking customers to pay with cash for purchases under $10.[17] Some other merchants offered customers incentives to consolidate transactions using prepaid cards or online wallets.[18]

A recent survey study conducted by the Federal Reserve Bank of Richmond provides more systematic evidence on the Durbin regulation's impact on merchants (Wang et al. 2014). The survey was performed from late 2013 through January 2014, roughly two years after the regulation was put in place. Survey respondents were merchants serving on a pre-existing research panel who sell goods and services directly to consumers and accept debit cards as a payment method. The sample comprised 420 merchants across 26 sectors in all U.S. states with various attributes.[19] The survey results show that the Durbin regulation has had an unequal impact on merchants' debit acceptance costs: Two-thirds of the survey respondents reported no change or did not know the change of debit costs post-regulation. One-fourth of the respondents reported an increase of debit costs, especially for small-ticket transactions, and less than 10 percent of the respondents reported a decrease of debit costs.[20] Among all sectors, Fast Food and Delivery Services ranked top in both small-ticket debit cost increase and total debit cost increase. Presumably, merchants in those sectors deal with mostly small-ticket transactions, so they were more likely to feel cost increases in both small-ticket and total debit transactions.

The study also finds significant merchant reactions to the post-regulation debit cost increases in terms of raising prices and debit restrictions. In the regression analysis, the

[17] See "Debit-Fee Cap Has Nasty Side Effect," *Wall Street Journal*, December 8, 2011.

[18] 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.-area parking operator Parkmobile offering discounts for customers who pay with an online wallet. See "Small-Ticket Retailers Squeezed By High Transaction Fees," *U.S. News & World Report*, October 26, 2012.

[19] The survey covers 26 merchant sectors: Apparel, Art, Auto, Casinos, Consumer Electronics, Convenience Stores, Delivery Services, Department Stores, Discount Retail, Education, Entertainment, Fast Food, Grocery Stores, Home Furnishings, Home Improvement, Hospitality, Maintenance, Medical, Office Products, Other Sector, Real Estate, Restaurants, Services, Sporting Goods, Toys, and Transportation.

[20] The unequal cost impact on merchants may result from the following factors: Merchants who reported reduced debit costs could be those who gained more from large-ticket transactions than they lost on small-ticket ones. Merchants reporting no change in debit costs could be those whose customers were primarily using debit cards from exempt issuers or whose losses from small-ticket transactions balanced out gains from large-ticket ones. Finally, merchants reporting increased debit costs could be those who specialized in small-ticket transactions.

study divides merchants into four dummy groups according to their debit cost changes: (1) small-ticket costs increased, total costs increased; (2) small-ticket costs increased, total costs did not increase; (3) small-ticket costs did not increase, total costs increased; (4) small-ticket costs did not increase, total costs did not increase. Holding other explanatory variables at their mean values, the estimation suggests that for a merchant who had no increase in small-ticket or total debit costs post-regulation, there is 5.1 percent probability that the merchant would raise prices. However, the probability rises to 59.6-74.7 percent for a merchant who did. Also, the probability that a merchant imposes a minimum amount for debit purchases rises from 7.5 percent to 30.9-45.1 percent if there was an increase in small-ticket and/or total debit costs. While the survey did not include merchants who did not accept debit cards at the time of the survey (as a result, the survey overlooked merchants who stopped accepting debit cards due to the post-regulation cost increase), its findings on changes in prices and minimum amount requirements suggest that the Durbin regulation has had a sizeable negative impact on debit use in small-ticket transactions.

What could explain the unintended consequences on small-ticket transactions? I propose the following explanation: Prior to the regulation, small-ticket transactions were loss leaders for card issuers, with low interchange fees set to get merchant card acceptance for small-ticket items, which increased the demand of cardholders with respect to larger-ticket items. Regulation reduced the card profit from large-ticket items and so reduced the incentive to loss-lead in this way. According to the recent Federal Reserve Board's survey (Federal Reserve Board 2011), the median per-transaction processing cost across covered issuers for all debit card transactions was 11 cents and the 80th percentile was 19 cents.[21] Issuers typically did not charge a fee to cardholders but incurred additional non-processing costs for handling cardholder inquiries and providing rewards (across covered issuers, the median per-transaction cost for handling cardholder inquiries was 3 cents and for providing cardholder rewards was 2-3 cents). Therefore, issuers appear to have subsidized small-ticket transactions by charging 4 cents plus 1.55 percent before the regulation, especially for those under \$5.[22] In the following sections, I provide a formal analysis.

[21] For signature debit transactions, the median issuer per-transaction cost was 13 cents and the 80th percentile was 21 cents. For PIN debit transactions, the median and 80th percentile were 8 cents and 14 cents, respectively.

[22] 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."

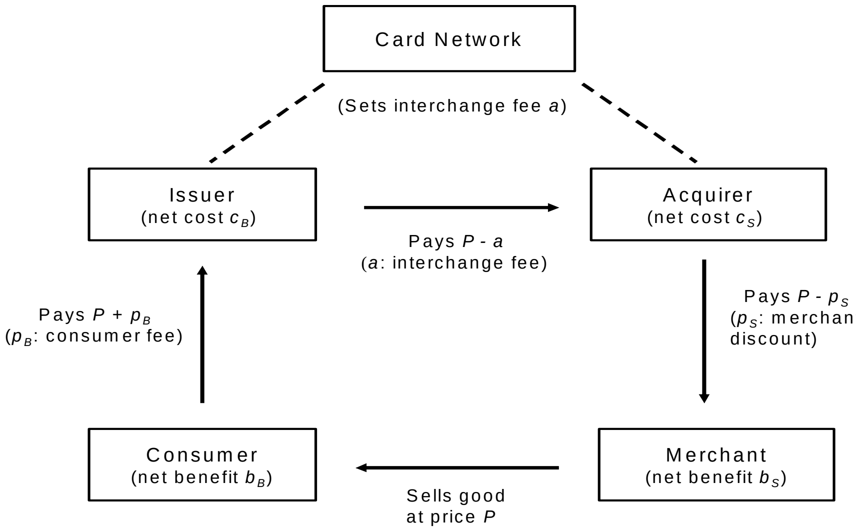

Figure 1: A Payment Card System

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. Consumers purchase two types of goods, large-ticket and small-ticket ones. For ease of exposition, I assume that merchants specialize in selling one of the two goods, but our analysis can equally apply to the scenario where some merchants sell both types of goods as long as they can restrict card use for small-ticket transactions by setting a minimum transaction amount. I will show below that the setup extends the standard two-sided market model, such as Rochet and Tirole (2002, 2011), to allow for card demand externalities across different transactions.

■ **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.[23]

See "Debit-Fee Cap Has Nasty Side Effect," *Wall Street Journal*, December 8, 2011.

[23] 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 the transaction value. Therefore, it is natural to assume that merchants and consumers benefit more from card usage in large-ticket transactions than in small-ticket transactions.

Consumers have inelastic demand and buy one good per sector. Within each sector, consumers need to decide which store to patronize. They know the stores' prices and card acceptance policies before making the choice. Once in a 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.[24] Whenever a transaction between a consumer (buyer) and a retailer (seller) is settled by card, the buyer pays a fee p_B^i to her card-issuing bank (issuer) and the seller pays a merchant discount p_S^i to her merchant-acquiring bank (acquirer). These fees, p_B^i and p_S^i, depend on the merchant sector $i \in \{h, l\}$, and p_B^i is allowed to be negative, in which case the cardholder receives a reward.[25] There are no annual fees and all consumers have a card.[26]

A consumer's transaction benefit of purchasing good i with a card instead of cash is a random variable b_B^i drawn from a cumulative distribution function H_i on the support $[\underline{b}_B^i, \overline{b}_B^i]$. It is natural to think that the mean and variance of consumer card usage benefits positively relate to the transaction value. Denoting μ^i as the mean of b_B^i, this implies that $\mu^h > \mu^l$ and $var(b_B^h) > var(b_B^l)$. For simplicity, I thereafter 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. The latter is an innocuous assumption given that the variance of H_l is sufficiently small.[27] Moreover, I assume that μ^h is positively affected by the consumer's card usage in the small-ticket sector y^l, i.e., $d\mu^h/dy^l > 0$. (I define $y^l = D^l \chi_l$, where χ_l indicates whether l-sector merchants accept cards and D^l is a consumer's

[24] 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.

[25] In reality, issuers often set card rewards as a percentage of the transaction value, so a large-ticket transaction typically delivers a higher reward than a small-ticket transaction.

[26] This model assumes a representative consumer framework developed by Wright (2004) and used in the subsequent literature. An alternative framework was used by Rochet and Tirole (2002), which assumes that heterogenous consumers differ systematically in their transaction benefits of using cards. As Rochet and Tirole (2011) show, these two alternative frameworks deliver convergent results, so our analysis and findings can be interpreted using either framework.

[27] Intuitively, we can think that a consumer's transaction benefit of paying by card relative to using cash is a random variable $b_B^i = bP^i + c$, where b is a random factor, P^i is the price of good i largely determined by the non-payment cost of the good, and c is a constant. This implies that $E(b_B^h) > E(b_B^l)$ and $var(b_B^h) > var(b_B^l)$. Moreover, given that P^l is small, $var(b_B^l)$ could be close to zero.

frequency of card usage in the l sector conditional on cards being accepted.)[28] This assumption captures the idea that ubiquity externalities shift up consumers' valuation of paying with cards in the h sector.[29]

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 $b_B^i - p_B^i$, a card payment is optimal for the consumer whenever $p_B^i \leq b_B^i$. Hence, whenever $p_B^h \leq \mu^h(y^l) + \gamma$, the proportion of card payments at an h-sector (i.e., large-ticket) store that accepts cards is

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

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

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

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

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

$$D^l(p_B^l) = \Pr(\mu^l \geq p_B^l) = 1, \tag{3}$$

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

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

Note that $D^l(p_B^l) = v^l(p_B^l) = 0$ if $p_B^l > \mu^l$.

■ **Merchants:** Merchants belong to one of the two sectors, h and l. A merchant in

[28] Under the assumption that H_l is degenerate, consumer card usage in the l sector becomes a simple binary outcome, i.e., $D^l \in \{0, 1\}$. This makes it easier to model card usage externalities between the l and h sectors. Note that if H_l is a non-degenerate distribution, we then need to specify how card demand externalities vary by various levels of card usage in the l sector, which would significantly complicate the problem without providing greater intuition.

[29] 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.

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 the 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 $D^i(p_B^i)v^i(p_B^i)$, but it faces an additional expected net cost of $D^i(p_B^i)(p_S^i - b_S^i)$ per cardholder from doing so. Here, p_S^i is the sector-specific merchant discount paid to the acquirer. Therefore, a merchant accepts cards if and only if $p_S^i \leq b_S^i + v^i(p_B^i)$. 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 holds for Cournot competition.

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

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

Note that merchants in the h and l sectors do not directly coordinate to internalize card usage externalities. This is a realistic assumption given that there could be a large number of merchants in each sector, which makes coordination too costly. Moreover, due to antitrust restrictions, merchants in reality cannot engage in group bargaining regarding interchange fees, so they typically face "take-it-or-leave-it" offers from card networks.

■ **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 p_S^i such that

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

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

■ **Issuers:** Issuers have market power. Particularly, I consider a monopoly issuer, who makes pricing decisions in h and l sectors to internalize card demand externalities.[30] The assumption of a monopoly issuer is reasonable in the case of debit cards because debit cards are issued based on cardholders' bank accounts and cardholders typically bear significant costs for switching banks.

The issuer incurs a per-transaction cost c_B and receives an interchange payment of a^i in a card transaction. The issuer also charges a consumer fee p_B^i per transaction, which can be negative if cardholders receive a reward.

For small-ticket transactions, the issuer takes the interchange fee a^l as given and sets the consumer fee p_B^l to maximize its profit. Expecting that merchants accept cards (i.e., $\chi_l = 1$), the issuer solves the following problem:

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

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

where (8) follows (3). Whenever $D^l = 1$, the highest possible consumer fee that the issuer chooses is

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

and the corresponding issuer profit in the l sector is

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

For large-ticket transactions, expecting that merchants accept cards (i.e., $\chi_h = 1$), the issuer solves the following profit maximizing problem:

$$\hat{\Pi}^h = \max_{p_B^h}(p_B^h + a^h - c_B)D^h \tag{11}$$

$$s.t.\ D^h = \begin{cases} \frac{\mu^h(y^l)+\gamma-p_B^h}{2\gamma} & \text{if } p_B^h \leq \mu^h(y^l) + \gamma \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

[30] More generally, we can think that the network serves n symmetric monopoly issuers and we simply normalize $n = 1$ in the analysis.

where (12) follows Eq (1). As a result, the card usage D^h and the consumer fee p_B^h are pinned down as follows:

$$D^h = \frac{1}{4\gamma}[\mu^h(y^l) + \gamma + a^h - c_B], \tag{13}$$

$$p_B^h = \frac{1}{2}[\mu^h(y^l) + \gamma - a^h + c_B], \tag{14}$$

and the issuer's profit in the h sector is

$$\hat{\Pi}^h = \frac{[\mu^h(y^l) + \gamma + a^h - c_B]^2}{8\gamma}. \tag{15}$$

■ **Network:** I consider a monopoly network that sets sector-specific interchange fees a^h and a^l to maximize the issuer 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 issuer profit, it makes a decision consistent with the issuer 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.

■ **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. The issuer and acquirers set fees p_B^i and p_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

In this section, I characterize the market equilibriums with and without the interchange cap regulation.

4.1 Basic Assumptions

To set the stage for the analysis, I make three basic assumptions on parameter values.

Assumption A1.

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

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 the issuer earns 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.[31] As will be shown, this implies that the issuer makes a loss for serving card transactions in the l sector *per se*.

Assumption A3.

$$\mu^h(1) - \mu^h(0) = Z^h(1) - Z^h(0) > \frac{9\gamma(-Z^l)}{2[Z^h(0) + Z^h(1)]}.$$

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

[31] It would be useful to think that Assumption 2 refers to average cardholder benefit, and also that it refers to surplus rather than just benefits. Thus, it is still consistent with some consumers and merchants being made jointly better off by using cards and accepting cards in the small-ticket sector.

4.2 Unregulated Market

I first consider the unregulated market outcome. Absent interchange regulation, the network sets sector-specific interchange fees a^i to maximize the issuer profit by solving the following problem:

$$\Pi = \max_{a^h,\, a^l} \frac{[\mu^h(\chi_l) + \gamma + a^h - c_B]^2}{8\gamma} \chi_h + (\mu^l + a^l - c_B)\chi_l \tag{17}$$

$$s.t.\ \ \chi_h = \begin{cases} 1 & \text{if } a^h \leq b_S^h - c_S + \frac{Z^h(\chi_l)}{3} \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

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

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

The findings are shown by the following proposition.

Proposition 1 *Under Assumptions A1-A3, an unregulated card network that maximizes the issuer 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, if only the h sector is served with card services (i.e., $\chi_h = 1, \chi_l = 0$), the card network maximizes the issuer profit (17) by setting the h-sector interchange fee such that the constraint (18) is binding

$$a^h(\chi_l = 0) = b_S^h - c_S + \frac{1}{3} Z^h(0). \tag{20}$$

As a result, the total number of card transactions is

$$D^h = \frac{1}{3\gamma} Z^h(0), \tag{21}$$

and the issuer profit is

$$\Pi^h = \frac{2}{9\gamma}[Z^h(0)]^2. \tag{22}$$

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

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

$$a^l = b_S^l - c_S. \tag{23}$$

Under Assumption A2, the issuer profit is

$$\Pi^l = Z^l = b_S^l + \mu^l - c_S - c_B < 0. \tag{24}$$

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

$$a^h(\chi_l = 1) = b_S^h - c_S + \frac{1}{3}Z^h(1), \tag{25}$$

$$a^l = b_S^l - c_S. \tag{26}$$

The resulting issuer profit is

$$\Pi^{h+l} = \frac{2}{9\gamma}[Z^h(1)]^2 + Z^l. \tag{27}$$

Comparing Eqs (20), (25), and (26), it is found that the interchange fee is always higher in the h sector than the l sector, i.e.,

$$a^h(\chi_l = 1) > a^h(\chi_l = 0) > a^l \tag{28}$$

given that $b_S^h > b_S^l$ and $Z^h(1) > Z^h(0) > 0$. Comparing (22) and (27), 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 (25) and (26) and serves card transactions in both the h and l sectors. ■

4.3 Interchange Cap Regulation

In comparison, I now characterize the model equilibrium under the interchange cap regulation. Once an interchange cap $\overline{a}$ is introduced by regulation, the network then solves a similar problem as above but with an additional constraint:

$$a^i \leq \overline{a} \ \text{ for } \ i \in \{h, l\}. \tag{29}$$

The goal of the analysis 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 cards, but merchants in the l sector do not.

Recall the finding (28) 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.[32] I now 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*

$$\mu^h(1) - \mu^h(0) < \frac{8\gamma(-Z^l)}{Z^h(1) + \frac{5}{3}Z^h(0)}. \tag{A4}$$

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

$$D^h = \frac{1}{4\gamma}[\mu^h(\chi_l) + \gamma + \overline{a} - c_B],$$

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

[32] 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 it is less relevant for explaining the market reality.

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

$$\Pi^{h+l} = (p_B^h + \overline{a} - c_B)D^h + Z^l = \frac{1}{8\gamma}[\mu^h(1) + \gamma + \overline{a} - c_B]^2 + Z^l.$$

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

$$\Pi^h = (p_B^h + \overline{a} - c_B)D^h = \frac{1}{8\gamma}[\mu^h(0) + \gamma + \overline{a} - c_B]^2.$$

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

$$\mu^h(1) - \mu^h(0) < \frac{8\gamma(-Z^l)}{\mu^h(1) + \mu^h(0) + 2\gamma + 2\overline{a} - 2c_B}. \tag{30}$$

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

$$\mu^h(1) - \mu^h(0) < \frac{8\gamma(-Z^l)}{\mu^h(1) + \mu^h(0) + 2\gamma + 2a^h(\chi_l = 0) - 2c_B}. \tag{31}$$

Inserting the expression of $a^h(\chi_l = 0)$ from Eq (20), the condition (31) can then be rewritten as (A4).

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

$$\frac{8\gamma(-Z^l)}{Z^h(1) + \frac{5}{3}Z^h(0)} > \frac{9\gamma(-Z^l)}{2[Z^h(0) + Z^h(1)]}.$$

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 $\overline{a}$ to serve only the h sector 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 small-ticket and large-ticket transactions could play an important role in explaining card networks' response to the regulation. Based on the model framework, I now proceed to 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.[33] 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

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

To be comparable with the analysis in the previous section, I assume that the social planner can set card fees (a^l, p_B^l) for small-ticket transactions to internalize card demand externalities (As we will show below, this outcome can be implemented as a market equilibrium). 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, p_B^l to maximize social welfare as follows:

$$\begin{aligned} W &= \max_{a^h, a^l, p_B^l} \frac{\chi_h}{2\gamma} \left([b_S^h - c_B - c_S][\mu^h(\chi_l) + \gamma - p_B^h] + \frac{[(\mu^h(\chi_l) + \gamma]^2 - p_B^{h2}}{2} \right) \qquad (33) \\ &\quad + [b_S^l + \mu^l - c_B - c_S]\chi_l \\ &\qquad s.t.\ (8),\ (14),\ (18),\ (19). \end{aligned}$$

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, and the fee levels coincide with those set by the unregulated private network.

Proposition 3 *The social planner who maximizes social welfare sets differentiated in-*

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

terchange fees to serve card transactions in both the h and l sectors, and the interchange fees set by the social planner coincide with those set by the private network.

Proof. Consider that the card network is run by a social planner who maximizes social welfare. Without considering the constraint (18), the welfare maximizing interchange fee in the h sector would be

$$\tilde{a}^h = b_S^h - c_S + Z^h(\chi_l),$$

which violates the constraint (18). Therefore, the constraint (18) has to bind, which implies

$$\tilde{a}^h = b_S^h - c_S + \frac{1}{3}Z^h(\chi_l). \tag{34}$$

Meanwhile, the social planner can choose whether to set card fees

$$\tilde{a}^l = b_S^l - c_S \quad \text{and} \quad \tilde{p}_B^l = \mu^l \tag{35}$$

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

$$W^h = \frac{2[Z^h(0)]^2}{9\gamma}. \tag{36}$$

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

$$W^{h+l} = \frac{2[Z^h(1)]^2}{9\gamma} + Z^l. \tag{37}$$

Under Assumption A3, $W^h < W^{h+l}$, so the social planner prefers the latter. Note that in this case, the welfare-maximizing interchange fees coincide with those set by the private network as shown in Proposition 1. ■

Our welfare findings are consistent with Rochet and Tirole (2002, 2011). There are two counteracting distortions in the card payment system that we consider (particularly in the h sector where the consumer transaction benefit of card usage follows a non-degenerate distribution).[34] 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. 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 of a monopoly issuer, the distortion due to issuer markup dominates. Because the welfare-maximizing interchange fee is limited by the merchant card acceptance constraint, the privately determined interchange fee coincides with the welfare-maximizing one.

However, the welfare analysis reveals important policy concerns. In fact, the coincidence between welfare maximization and profit maximization implies that the end users (i.e., merchants and consumers) essentially receive zero total surplus from using card services as stated in the following proposition.

Proposition 4 *While the interchange fees set by the private card network maximize social welfare, the total user surplus is zero.*

Proof. Propositions 1 and 3 show that issuer profit-maximizing card fees equal welfare-maximizing ones, and the corresponding issuer profit equals social welfare. Because social welfare comprises issuer profit, consumer surplus, and merchant profit, this finding implies that the sum of consumer surplus and merchant profit equals zero. ■

Because policymakers typically value the welfare of end users, our finding explains why the debit regulation was called for in the first place. Below, we study further the regulatory implications of the model.

5.2 Regulatory Implications

The analysis above shows that market-determined interchange fees could be too high from the total user surplus perspective, which provides some justification for regulatory interventions like the Durbin Amendment. Our model provides a useful framework for evaluating such a regulation.

[34] Because of the assumption that consumers' transaction benefit of card usage in the l sector follows a degenerate distribution, we abstract from distortion in the l sector *per se* (which is supposed to be sufficiently small anyway).

By definition, the 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 (i.e., $b_S^i + b_B^i > p_B^i + p_S^i$), which means

$$TUS = \sum_{i \in \{h,l\}} \left(\chi_i \int_{p_B^i}^{\overline{b}_B^i} [b_S^i + b_B^i - p_B^i - p_S^i] dH_i(b_B^i) \right). \tag{38}$$

In other words, total user surplus is the sum of consumer surplus and merchant profit, which also equals social welfare minus issuer profit. In our model context, the expression of the total user surplus is shown below. Note that the first item in the expression (39) is social welfare in the h sector as shown in (33), and the second item is issuer profit in the h sector as shown in (17).

$$\begin{aligned} TUS &= \frac{\chi_h}{2\gamma} \left([b_S^h - c_B - c_S][\mu^h(\chi_l) + \gamma - p_B^h] + \frac{[(\mu^h(\chi_l) + \gamma]^2 - p_B^{h2}}{2} \right) \\ &\quad - \frac{[\mu^h(\chi_l) + \gamma + a^h - c_B]^2}{8\gamma} \chi_h + (b_S^l + \mu^l - p_B^l - c_S - a^l)\chi_l. \end{aligned} \tag{39}$$

Under the Durbin regulation, the interchange fee is regulated down to the level of the issuer cost (c_B in our model context).[35] In spite of the negative impact on small-ticket transactions, our model suggests that such a regulation may indeed result in higher total user surplus. This is shown in the following proposition.

Proposition 5 *Regulating down interchange fees improves the total user surplus in spite of the unintended consequence on small-ticket transactions.*

Proof. Proposition 4 shows that absent regulation, the total user surplus is zero. In comparison, Proposition 2 shows that the market response to the interchange regulation can be rationalized by a cap $\overline{a}$ that $a^l < \overline{a} \leq a^h(\chi_l = 0) = b_S^h - c_S + \frac{Z^h(0)}{3}$, under which the card network sets a single interchange fee at $\overline{a}$ such that cards are used only in the h sector. (More generally, this will also happen in the case that $\overline{a} \leq a^l$ because the issuer will then

[35] In contrast, the European Commission has recently regulated interchange fees based on merchants' convenience benefits of accepting cards, an approach named the "merchant indifference test." The approach, as shown in Rochet and Tirole (2011), is intended to maximize the total user surplus under the assumption of a constant issuer markup.

have incentive to raise p_B^l so that $p_B^l > \mu^l$ to avoid serving small-ticket transactions.) We can verify that the issuer profit increases in $\overline{a}$ and equals zero for $\overline{a} = b_S^h - c_S - Z^h(0)$. We can also verify that in the range $\overline{a} \in [b_S^h - c_S - Z^h(0), b_S^h - c_S + \frac{Z^h(0)}{3}]$ where the issuer finds it profitable to operate, the total user surplus TUS^h is non-negative. In fact, TUS^h equals zero only when $\overline{a}$ takes the boundary values (i.e., $\overline{a} = b_S^h - c_S - Z^h(0)$ and $\overline{a} = b_S^h - c_S + \frac{Z^h(0)}{3}$) and is strictly positive anywhere else. Therefore, capping the interchange fee by the issuer cost in this range results in a positive total user surplus. ■

Proposition 5 provides some justification for the Durbin regulation. However, our analysis suggests that the issuer cost-based regulation lacks theoretical foundation and could result in unintended consequences. 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 fee actually improves both the total user surplus and the issuer profit, and hence increases social welfare. This is shown in the following proposition.

Proposition 6 *Depending on the value of issuer cost, the issuer cost-based regulation may under- or overshoot. Especially in the overshooting case, the cap could be set too low so that a higher interchange cap improves both the total user surplus and the issuer profit, and hence increases social welfare.*

Proof. Following the proof above in Proposition 5, we can further verify in the range $\overline{a} \in [b_S^h - c_S - Z^h(0), b_S^h - c_S + \frac{Z^h(0)}{3}]$, the total user surplus TUS^h is strictly concave (i.e., $\partial^2 TUS^h / \partial \overline{a}^2 < 0$) and reaches the maximum value when $\overline{a} = b_S^h - c_S - \frac{Z^h(0)}{3}$. We also know that in this range, the issuer profit increases in $\overline{a}$. Because the issuer cost c_B is an independent parameter, using it as a regulatory benchmark is quite arbitrary, which may under- or overshoot the regulator's preferred balance between the total user surplus and the issuer profit. Especially, in the case that $c_B \in [b_S^h - c_S - Z^h(0), b_S^h - c_S - \frac{Z^h(0)}{3})$, capping the interchange fee by the issuer cost definitely overshoots, in the sense that there exists a higher interchange cap $\overline{a} \geq b_S^h - c_S - \frac{Z^h(0)}{3}$ that improves both the total user surplus and the issuer profit, and hence increases social welfare. ■

Another unintended consequence is the side effect on small-ticket transactions as discussed before. We show in the following that an alternative cap regulation may improve the total user surplus by taking into account card demand externalities across different transactions.

In principle, policymakers could design the interchange cap in a way that penalizes the card network and the issuer in case they stop subsidizing small-ticket transactions. To see how this works conceptually, consider the following example in our model context. Denote $\overline{a}$ as the interchange cap set by the current regulation, and a^A as the cap under the alternative regulation. The alternative regulation may cap the maximum interchange fee at

$$a_t^A = \overline{a} - \delta(1 - y_{t-1}^l).$$

This implies that for period t, the cap is set at $a_t^A = \overline{a}$ if small-ticket transactions were served last period (i.e., $y_{t-1}^l = 1$); otherwise, the cap is reduced to $a_t^A = \overline{a} - \delta$. The regulation can be enforced by requiring the network to report the small-ticket transaction volume each period. As long as the penalty δ is large enough, the card network and the issuer would continue to serve small-ticket transactions. In each period, compared with the current regulation, the total user surplus would increase by $TUS^{h+l}(a^A = \overline{a}) - TUS^h(a^A = \overline{a})$, though the issuer profit would decrease by $\Pi^h(a^A = \overline{a}) - \Pi^{h+l}(a^A = \overline{a})$. A regulation like this can also be implemented in other ways. For instance, policymakers can set the cap conditional on the small-ticket fees (instead of transaction volumes) by requiring $a_t^A = \overline{a}$ if the card network and the issuer keep their small-ticket rates (a^l, p_B^l) at the pre-regulation levels; otherwise, requiring $a_t^A = \overline{a} - \delta$.

It is worth noting that while we discuss alternative regulations, the purpose is not to make a particular policy proposal, but rather to shed light on the working of the unintended consequences in the two-sided card market so that more informed regulatory decisions can be made. Ultimately, policymakers need to consider the trade-off between the total user surplus and the issuer profit as well as the implementation costs to decide on which regulatory approach to pursue.

Also, considering the two-sided nature of the market, our analysis suggests that cautions need to be taken with regard to some regulatory options. For example, setting a separate low interchange cap for small-ticket transactions (or in a similar vein, setting a percentage cap so that large and small transactions essentially face different caps in absolute term) may not solve the small-ticket problem if the issuer can raise the consumer fee p_B^l to shut out small transactions. For a similar reason, the proposal to set a very low interchange cap for all transactions may not help small-ticket transactions either. Moreover, as shown in the proof of Proposition 6, a too-low interchange cap may also run the risk of overshooting for large-ticket transactions.[36]

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 merchants' card acceptance costs by capping the maximum interchange fee, the fee for small-ticket transactions became higher instead.

In this paper, I address this puzzle by introducing card demand externalities into a two-sided market model. The findings rationalize the card networks' response to the cap regulation: Before the regulation, card networks and issuers were willing to offer subsidized interchange fees to small-ticket transactions in order to increase overall card acceptance and usage, recouping the losses via high interchange fees on large-ticket transactions. Once an interchange cap was imposed, this loss-leading strategy was no longer profitable so they discontinued the subsidy. Based on the model, I provide a welfare assessment of the regulation. The analysis shows that in spite of the negative impact on small-ticket transactions, the regulation may indeed improve the total user surplus. However, an issuer cost-based regulation lacks theoretical foundation and could result in unintended consequences such as under- or overshooting and adverse effects on small-ticket transactions.

The takeaway of the paper is that interchange fees encompass more than just the costs of processing payment card transactions. In the two-sided market, they also serve to balance demand between consumers and merchants, as well as coordinate card acceptance and usage among different transactions. The model shows that privately determined

[36] Consider in our model context, the regulator sets an interchange cap for all transactions at the pre-regulation small-ticket level (i.e., $\overline{a} = b_S^l - c_S$) in the hope of protecting small-ticket card transactions. However, the issuer may still be able to raise p_B^l to shut out small transactions. Moreover, for large-ticket transactions, in the case that $\overline{a} = b_S^l - c_S < b_S^h - c_S - \frac{Z^h(0)}{3}$, there exists a higher interchange cap that improves both the total user surplus and the issuer profit, and hence increases social welfare.

interchange fees tend to be too high from the total user surplus perspective, so regulating down interchange fees may help to improve the market outcome. But regulation that only considers one-sided market logic (setting fees equal to issuers' costs, for example) or one sector of the market (ignoring card demand externalities between large- and small-ticket transactions, for example) may result in unintended consequences.

There are several avenues for future research. First, it would be interesting to test our model against alternative hypotheses. For example, one may think of the regulation as creating a focal price equal to the maximum cap rate so that previously competing networks were able to collude.[37] Second, to overcome the shortcomings of the issuer cost-based regulation, it would be useful to investigate card usage benefits for consumers and merchants in different transactions. As our two-sided market model suggests, this will provide the critical information for designing a sound interchange policy. Finally, one may consider policy options other than price regulation. For instance, in theory, if merchants can set different retail prices conditioning on payment means, interchange fees may become less of an issue.

References

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

[2] Bedre-Defolie, Özlem and Emilio Calvano (2013). "Pricing Payment Cards." *American Economic Journal: Microeconomics*, 5(3): 206-31.

[3] Bolt, Wilko and Sujit Chakravorti (2008). "Consumer Choice and Merchant Acceptance of Payment Media." DNB Working Paper 197, De Nederlandsche Bank.

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

[37] The "focal price" argument relies on the assumption that the regulation created a focal price so that previously competing networks were able to collude. However, a potential counterargument is why networks could not collude before the regulation. The interchange fees were set by a very small number of networks and all the fees were publicly posted. Therefore, networks could easily collude even before the regulation if they wanted to. In fact, collusion has long been a major criticism that merchants and policymakers brought against card networks. This lends some support to our model for assuming a monopoly network.

[5] Federal Reserve Board (2011). "2009 Interchange Revenue, Covered Issuer Cost, and Covered Issuer and Merchant Fraud Loss Related to Debit Card Transactions." http://www.federalreserve.gov/paymentsystems/files/debitfees_costs.pdf.

[6] Foster, Kevin, Scott Schuh and Hanbing Zhang (2013). "The 2010 Survey of Consumer Payment Choice." Research Data Report No. 13-02, Federal Reserve Bank of Boston.

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

[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] 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.

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

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

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

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

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

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

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

[17] 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.

[18] Shy, Oz (2014). "Measuring Some Effects of the 2011 Debit Card Interchange Fee Reform." *Contemporary Economic Policy*, 32(4), 769–783.

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

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

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

[22] Wang, Zhu, Scarlett Schwartz and Neil Mitchell (2014). "The Impact of the Durbin Amendment on Merchants: A Survey Study." *Federal Reserve Bank of Richmond Economic Quarterly*, 100 (3): 183–208.

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

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

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

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

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