On the Level and Incidence of Interchange Fees Charged by Competing Payment Networks*

Robert M. Hunt† Konstantinos Serfes‡ Yin Zhang§

August 27, 2024

Abstract

We develop a two-sided model of the payment card market, featuring novel elements, such as detailed demand, merchant competition, and competing networks with ad valorem pricing for interchange fees and rewards. A network imposes a "credit card tax," influenced by these prices, which varies with merchant competition and demand characteristics. We highlight the "elasticity effect", related to demand subconvexity, and the "competition effect". Enhanced network competition, when networks are differentiated and so the elasticity effect dominates, leads to higher credit card taxes and merchant prices, reducing welfare. Conversely, with minimal differentiation, intense competition for cardholders, due to stronger network competition, lowers the tax and enhances welfare.

Keywords: credit cards, debit cards, two sided networks, interchange fees, antitrust

JEL codes: L13, L40, G28, E42

*We would like to thank Konstantinos Charistos (discussant), Eugenio Miravete, Arina Nikandrova (discussant), Fumiko Hayashi (discussant), Patrick Rey, Oz Shy (discussant) and seminar/conference participants at the 2023 Economics of Payments conference at the Federal Reserve Board, Oligo 2024, the Federal Reserve Bank of Philadelphia, the Dialogues in Competition Law & Economics workshop, CRESSE 2024, CRETE 2024 and SPR 2024 for useful comments. **Disclaimer:** The views expressed in this research paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve Board of Governors, or the Federal Reserve System. Nothing in the text should be construed as an endorsement of any organization or its products or services. Any errors or omissions are the responsibility of the author. No statements here should be treated as legal advice.

†Consumer Finance Institute, Federal Reserve Bank of Philadelphia, Philadelphia, PA 19106. e-mail: bob.hunt@phil.frb.org.

‡School of Economics, LeBow College of Business, Drexel University, Philadelphia PA 19104. E-mail: ks346@drexel.edu.

§School of Economics, LeBow College of Business, Drexel University, Philadelphia PA 19104. E-mail: yz864@drexel.edu.

1 Introduction

This paper examines the level and incidence of a wholesale price-the interchange fee-typically set by a payment card network that influences the distribution of acceptance costs and benefits incurred by merchants and consumers. Interchange fees have been controversial for half a century, ever since the credit card networks, now known as Visa and MasterCard, began to dominate purchases at the point-of-sale in the early 1970s.

One of the complications in understanding the role and consequences of interchange fees is the consensus among economists that payments represent a classic example of a two-sided market that is intermediated by a platform-the payment card network. As such, the adoption and pricing decisions that affect one side of the market (cardholders) affect the comparable decisions made by participants (merchants) on the other side of the market. Competition and welfare implications in these markets can be very different than in the traditional markets studied by economists. Over the last three decades, an extensive theoretical literature on two-sided markets has emerged and many of those papers examine the implications for payment systems.[1]

In this paper, we are interested in the level and incidence and the welfare implications of interchange fees in a new model that allows for varying degrees of competition both amongst payment networks and amongst merchants. Why develop a new model? We think there are a number of important elements of the problem that are not adequately addressed by the existing literature, and we attempt to address those limitations with a different approach. We will illustrate with several observations.

First, the literature on networks and two-sided markets has rightly focused on the chicken-and-egg problem associated with establishing a successful payment network amidst a host of adoption externalities and coordination problems. Success has often required that users on one side of the network be subsidized at the expense, at least relatively so, of users on the other side. Once widespread adoption and coordination is established, however, those same phenomenon may lock-in an equilibrium with one or a few payment networks that are both durable and difficult to displace. It is when networks are well established that concerns about antitrust, inequality, and welfare losses emerge. For example, a major theme of the current policy debate in the United States is the extent to which networks raise interchange fees paid by merchants to cover the costs of increasing rewards (e.g., cash back or air miles) paid to cardholders. This, in turn, raises questions about welfare and inequality.

The models that are well suited for studying the emergence and adoption of a network in a two-sided market are less well suited for understanding the questions policy makers face when confronting an equilibrium with established networks. This is because most of the models developed hitherto abstract away from important elements we observe in the real world payments market.

For simplicity, the prevailing literature often assumes that consumer decisions are binary: individuals

[1] See Rochet and Tirole (2006), Rysman (2009) and Jullien, Pavan, and Rysman (2021) for reviews of this literature.

either make a purchase or they do not, with the purchased quantity being predetermined. The model's extensive margin arises from less urgent consumers who buy only when prices drop sufficiently, yet they too purchase the predetermined quantity. In scenarios where demand is elastic, a constant elasticity is assumed and this assumption rules out a host of potentially empirically relevant outcomes. Moreover, there is typically no explicit connection between the product and payment markets in the literature; for instance, the prices paid by consumers are usually not influenced directly by the rewards card users gain from transactions.

At the same time, in most of the literature, the fees charged to enable payments are specific, i.e., an absolute fee. Ad valorem pricing is rarely studied even-though it is the dominant component of interchange fees and consumer rewards set by payment card networks.[2] And we know from the public economics literature that the implications of specific vs. ad valorem taxes are usually different. In addition, the implications of ad valorem pricing cannot be thoroughly examined in models of inelastic individual demand.

Similarly, while the market structure and nature of competition among banks and networks are modeled with detail and variation in the literature, the nature of competition and the extent of market power in the retail sector is relatively undeveloped. The combination of these assumptions in much of the literature limits the kinds of welfare statements that can be made about interchange fees using these models. To address these limitations, we propose a different approach that follows from the literature on tax incidence. We can establish several general results. However, because we move away from several of the simplifying assumptions described above, additional results are established using examples and numerical simulations. We believe those results can be generalized with additional work.

We parameterize the retail sector in three dimensions: the number of firms (n), the elasticity of demand (ε), and the conjectural variation (λ) retailers use when interpreting how their competitors will respond to their pricing decisions. The number of retailers and the conjectural variation combined determine the conduct parameter (γ) that describes the amount of competition in the final goods market; on the one extreme $\gamma = 0$ means perfect competition (Bertrand), while on the other extreme $\gamma = 1$ implies a monopoly or a perfect cartel. This set-up permits us to understand more generally how consumer preferences (as captured by the shape of the demand function) and merchant competition endogenously determine retail prices. Interchange fees and the consumer response to rewards affect consumer demand which, in turn, affects retail prices.

The baseline model establishes an equilibrium in which aggregate demand and merchant profits are determined by what we call the "credit card tax", $z \equiv \frac{1-r}{1-i} > 1$, which determines the wedge between the price consumers pay and the price merchants receive for every unit of the final good; the difference between these two prices is the per dollar of transactions network revenue. The amount of this tax

[2] Shy and Wang (2011) along with Wang and Wright (2017, 2018) consider models with ad valorem pricing. However, they either assume demand with constant elasticity or model merchant competition as Bertrand. Such assumptions lead to perfect tax pass through on consumers, which may restrict the applicability of any derived policy insights.

depends negatively on the relative generosity of cardholder rewards, r, and positively on the interchange fee, i, paid by merchants, and emerges naturally when we link the product market with the network fees and rewards. For any fixed value of the level of this tax, there are a continuum of pairs of interchange fees and rewards that will generate the same level of consumption and merchant profits.

A standard result follows from demand and taxation theory: if there is perfect competition in the retail market or demand is of the constant elasticity form, then there is perfect pass through of any credit card tax to cardholders. In other words, the incidence of any markup resulting from using payment cards is borne entirely by consumers. Further, the more intense the competition in the retail sector (lower γ), the higher will be consumer surplus and the profits earned by the payment network. This follows from (1) the greater aggregate consumption that occurs when the goods market is more competitive and (2) the use of ad valorem pricing by the network. In addition, reliance on ad valorem pricing for interchange and rewards makes the network sensitive to the effects of the credit card tax on aggregate demand. This is an important element of the implications that follow.

If the price elasticity is not constant, then merchants do not raise prices one-for-one as the credit card taxes rise-so long as competition is not perfect (i.e., $\gamma > 0$). The tax incidence depends on the sign of the derivative of the demand elasticity with respect to aggregate output, which affects what we term *the elasticity effect*. If the sign is positive, i.e., market demand becomes more elastic when aggregate output falls (equivalently, subconvex demand, e.g., Mrázová and Neary (2017), Mrázová and Neary (2019)), then the incidence of any credit card tax is borne by merchants and consumers. More intense competition in the product market increases the credit card tax and the fraction of the tax that is borne by consumers. If, on the other hand, the sign is negative, i.e., market demand becomes less elastic when aggregate output falls (equivalently, superconvex demand, e.g., Mrázová and Neary (2017), Mrázová and Neary (2019)), then there is credit card tax over-shifting: consumers may pay more than 100% of the tax. More intense competition in the product market decreases the credit card tax and the fraction of the tax that is borne by consumers.

Next, we allow for competition between two differentiated payment networks. This permits us to explore the complicated question of whether network competition raises costs by competing for cardholders via more generous rewards, (see Guthrie and Wright (2007) for an example; for a general discussion, see Hayashi et al. (2009)). First, consider the case of a second payment network establishing a "toehold" in the market. This has the effect of reducing the aggregate demand of consumers that use the incumbent card brand. If market demand becomes more elastic when aggregate output decreases (subconvex demand), then the incumbent network responds by increasing its credit card tax. This is the elasticity effect and it can be understood in more detail as follows. A higher tax reduces aggregate output, and if market demand becomes more elastic merchants lower the price they charge resulting in losses for the network, due to lower total value of transactions. When the incumbent network has a lower market share, due to entry of a second network, the effect of the incumbent's higher tax on aggregate output is smaller and this, in turn, diminishes the negative impact of the elasticity effect. Hence, the incumbent is less reluctant to increase its tax, which leads to a higher tax.

However, when we endogenize payment network competition for cardholders the results can change. We do this using a Hotelling model, which allows us to parameterize the extent of network differentiation. Because we assume there are no fixed costs of adoption for retailers, so long as merchants cannot price discriminate (i.e. steer) with respect to the brand of payment card presented, in equilibrium merchants will multi-home in card acceptance. Consumers, in turn, have no incentive to multi-home.[3]

Relative to a monopoly network, we find that in the case of constant elasticity demand or Bertrand competition in the product market the presence of a second network lowers the credit card tax, increases merchant profits, raises consumer surplus and reduces the incidence of the tax paid by consumers. The effect of network competition for cardholders (what we call *the competition effect*) is at work since the elasticity effect is absent when the elasticity is constant or competition is Bertrand. However, when competition in the product market is imperfect and the elasticity of demand is not constant, the elasticity effect appears. As the payment networks become more differentiated, the competition effect is attenuated, reducing the extent of competition for cardholders. With sufficient network differentiation, the elasticity effect dominates, so entry of a second network increases the credit card tax and the product price consumers pay and reduces welfare. The elasticity effect can be easily overlooked, as it has been the case in the two-sided market literature, if the analysis is based on oversimplified assumptions regarding the product market and consumer demand.

This finding has important policy implications regarding the effect of network competition and its interaction with product market competition on prices and welfare. Consider, for instance, the recent proposed merger between Capital One and Discover. This is a case of two banks who use different networks to clear and settle credit card transactions that might combine and use one network. This, in turn, would alter the market shares of the dominant credit card networks. What is the effect of such a merger on the credit card tax and hence on prices consumers pay? Our theory suggests that, even in the absence of economies of scale, the credit card tax and hence merchant prices can go up or down as a result of changes in network competition. Key factors are how differentiated will networks be in the new equilibrium and the shape of product demand.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of the literature; it is not intended to be comprehensive. Section 3 lays out the fundamental structure of the model, while Section 4 analyzes merchant competition and equilibrium with a monopoly network under different assumptions about product demand. Section 5 analyses the case of two competing networks, each having a fixed number of cardholders. Section 6 introduces competition between the networks for cardholders and compares those results to the ones in Section 5. Section 7 introduces cash as an alternative mode of payment. Section 8 concludes and presents a number of policy implications.

[3] In the text we discuss the implications for merchant surcharging - a form of steering - on equilibrium outcomes.

2 Literature review

The literature on two-sided markets, especially as applied to payment systems is voluminous, spanning over four decades. Our purpose here is simply to sketch the main findings and highlight our contributions.

Baxter (1983) developed the first formal model of interchange fees in a payment scheme. In that paper, Baxter makes three key assumptions: i) Issuers and acquirers make no profit (perfect competition), ii) Merchants do not use card acceptance strategically, i.e., to attract consumers from rival merchants who do not accept a card and iii) there is no merchant heterogeneity in the benefit of accepting cards. Schmalensee (2002) develops a model that explores the double marginalization problem that emerges when networks set an interchange fee that is paid to card issuers and the acquiring bank for the merchant separately sets a merchant discount for processing transactions.

Rochet and Tirole (2006) offers a definition of a two-sided market: *A two-sided market is one in which the volume of transactions between end-users depends on the structure and not only on the overall level of the fees charged.* This can be contrasted with Rysman (2009), where he draws a parallel to the literature on network effects in which demand for a given good depends on the supply of a complementary good.

In Rochet and Tirole (2002), the consumers receive a different benefit from transacting using cards rather than paying cash. There is a single payment network that sets an interchange fee, but it is not ad valorem. The retail market is a Hotelling model, but consumers face the choice of purchasing a fixed quantity of their preferred good. Rochet and Tirole (2011) develop a similar model, but here there is competition between networks modelled using the Hotelling framework. Wright (2004) uses a model similar to Rochet and Tirole (2002) and relaxes all three assumptions of Baxter (1983). Bedre-Defolie and Calvano (2013) show that networks oversubsidize card usage and overtax merchants.

Guthrie and Wright (2007) show that network competition can increase the interchange fees. This result is reminiscent of our result, where entry of a network can increase the credit card tax. However, the two results are qualitatively different and result from different underlying factors. In Guthrie and Wright (2007) network entry can induce networks to compete more vigorously on the rewards they offer to users and to compensate their profits they also increase the interchange fees. To put differently, while entry can increase the interchange fee it need not increase the overall credit card tax. In contrast, in our model network competition can increase the credit card tax.

The prevailing methodology in the majority of the above papers is twofold: (1) it posits that consumer choice is binary–either purchasing a single unit or none at all, and (2) it considers fees that are specific rather than ad valorem. While these simplifications are made to facilitate analysis, they diverge significantly from real-world observations. Furthermore, there is a common presumption that the benefits and costs experienced by both parties-merchants and consumers-due to card usage are independent of each other and of the product price.

Our study seeks to bridge this gap by offering a more generalized approach, albeit at the expense of some analytical simplicity.

There are papers that employ assumptions that are closer to ours. Shy and Wang (2011) adopt a constant elasticity demand and compare "proportional" versus "fixed" transaction fees in a model where both retailers and a payment network enjoy market power. Consumers and networks do better under proportional fees; merchants do worse. The less competitive the retail market is, the greater the benefit to the network of charging proportional as to fixed fees. In related work, Wang and Wright (2017, 2018) assume Bertrand competition among sellers in a market with many different goods that vary widely in their costs and values. By assumption there is perfect pass through of any taxes to buyers. The authors show that ad valorem fees and taxes represent an efficient form of price discrimination relative to uniform fees that disadvantage low-cost, low-value goods. **?** develops a structural approach to a two-sided market of payments. He finds that interchange fee caps increase welfare by reducing rewards, retail prices, and credit card use. In the absence of regulation, because consumers are reward-sensitive, but merchants are fee-insensitive, entry of private credit card network raises rewards without cutting fees, lowering welfare.

Edelman and Wright (2015) shows that an intermediary always chooses to impose price coherence if it has the ability to do so. Doing so increases its profit even though it leads to excessive intermediation, excessive investment in buyer-side benefits, and indeed harms buyers, making them worse off in aggregate compared to the case without any intermediation. They show these effects persist and even grow when multiple intermediaries compete. Such outcomes can be overcome if merchants have a means of steering consumers to a preferred network or form of payment (e.g. surcharging). An important qualification is that the fees examined are not ad valorem.

There is a growing body of literature that employs two-sided market models to study competition in digital markets, such as Jeon and Rey (2022) and Bisceglia and Tirole (2023). A significant concern in these markets is the high commissions charged by platforms, like Apple's App Store and Google's Play Store, to app developers. Jeon and Rey (2022) demonstrate that competition between platforms exacerbates rather than alleviates high commission fees. While this result is similar to ours in its implications, the underlying mechanism and model differ significantly.

Our elasticity effect is related to the shape of the demand function and in particular whether it is superconvex or subconvex. Mrázová and Neary (2019) define a demand function as superconvex if $\log p$ is convex in $\log x$. This is equivalent to the demand function being more convex than a constant elasticity CES demand function, and to one whose (absolute) elasticity of demand is increasing in output. Subconvexity is equivalent to the demand becoming less elastic as output increases. Super- or subconvexity determine competition effects and relative pass-through. We quote from Mrázová and Neary (2017) on page 3840: "Hence, if globalization reduces incumbent firms' sales in their home markets, it is associated with a higher elasticity and so a lower markup if and only if demand is subconvex." Rephrasing the quote: entry in the product market increases mark-ups if and only if demand is superconvex. Superconvexity also implies a more than 100% pass-through, which has implications in our analysis for the credit card tax incidence. Subconvexity (which encompasses the linear demand) is sometimes called "Marshall's Second Law of Demand", but superconvexity cannot be ruled out either theoretically or empirically (for more details see the discussion in footnote 10 in Mrázová and Neary (2017)).

Our analysis uses the concepts of subconvexity and superconvexity to examine how changes in competition at the network level, and its interaction with the product market structure, affect the credit card tax and its incidence. We show that more intense network competition can *increase* the credit card tax when demand is subconvex, whereas under subconvexity any potential strengthening of competition in the product market *lowers* merchant mark-ups. This difference underscores the nuanced impact of the structure of preferences and demand on competition in two-sided and vertically related markets.

3 Structure of the market

We consider an industry consisting of n firms (merchants), $j = 1, ..., n$, producing a single homogeneous product. The output of firm j is denoted by x_j and the industry output by $X = \sum_j x_j$. All the merchants have the same cost structure $C(x) = cx$, where $c > 0$ is a constant marginal cost. The consumer price is given by an inverse demand function $P(X)$, with derivative $P_X(X) < 0$ and elasticity $\varepsilon \equiv \frac{P}{XP_X} < 0$.

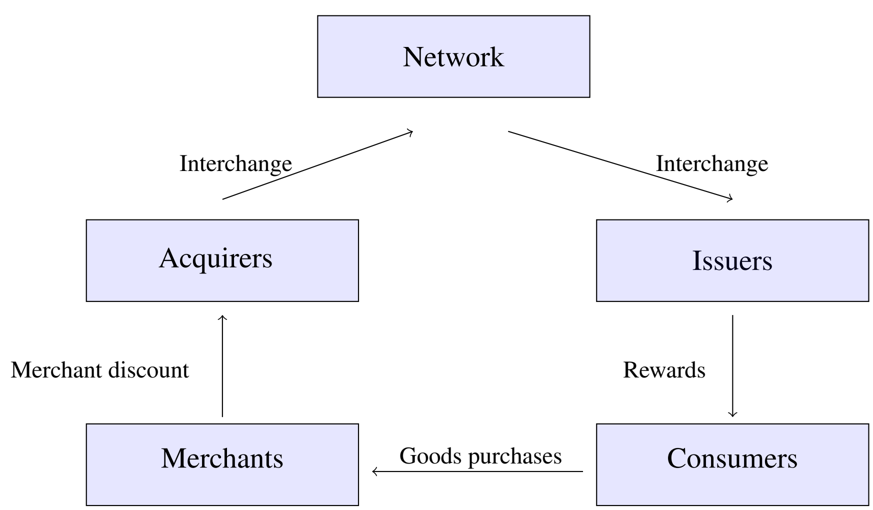


Figure 1: Payment flows in a Network

In addition to the product market, there are two competing payment networks, indexed by $l = 1, 2$. We can also allow for cash as an alternative payment mode, see section 7 where cash is introduced. Each network issues its own credit card and consumers must make purchases using one of the two available modes of payments.

More specifically, in a given network, there are $N_A < n$ acquiring and $N_I < n$ issuing banks, that are homogeneous and compete à la Bertrand. We assume that networks do not compete to attract banks. In Figure 1 we present the payment flows in one network. Each acquiring bank chooses its merchant discount m to attract merchants and each issuing bank chooses the reward r to attract users/consumers and to influence the value of consumer transactions. The network chooses the interchange fee, i, which becomes each acquirer's marginal cost. Part of this fee, f, goes to the issuing banks to fund the rewards and the rest is kept by the network. For simplicity, all other costs to process a transaction are assumed to be zero.

The reward for credit card l is denoted by $r_l \in [0, 1]$ and is a percent of the value of the transaction that a consumer who uses credit card l receives as a cash back from the issuing bank. The acquiring bank charges a merchant discount fee $m_l \in [0, 1]$, which is a percent of the value of the transaction that is paid by the merchant to the acquiring bank when a consumer uses credit card l. We assume that a network cannot price discriminate across banks, i.e., all acquirers pay the same interchange fee to the network and all issuers receive the same fraction of the interchange fee from the network.

We assume that networks are horizontally differentiated (more on this in Section 6) and are competing for users. The network fees determine the value of transactions in a network and can also be used to attract users from the rival network. Each consumer incurs a small cost ϵ if he multi-homes, i.e., holds a second credit card.

We analyze a five stage game with simultaneous and independent moves in each stage. In stage 1, networks set their interchange fees, i_l. In stage 2, the networks choose how much of the interchange fee, f_l, will be given to each issuing bank. In stage 3, acquiring banks set the merchant discounts m_l and issuing banks set the rewards, r_l. In stage 4, each merchant chooses whether to accept both credit cards or only one and its product quantity. In stage 5, each consumer chooses whether to hold one or both credit cards and makes purchases.

4 Analysis with a monopoly network

As a benchmark, we begin the analysis by assuming the existence of a monopoly network and no other modes of payment. We solve the game backwards, looking for a subgame-perfect Nash equilibrium.

4.1 Stage 4: Merchant competition

We denote the monopoly network by 1. Since there is no other mode of payment, merchants must accept the network's credit card and consumers must use it. Because market demand is elastic, and hence the value of transactions is affected by the network fees, the network, as we will show shortly, has no incentive to set its fees to levels that would make the merchant profits or the consumer utility zero. Thus, the absence of an outside option in terms of an alternative payment mode does not affect the analysis significantly.

Since the reward and interchange fee are ad valorem, the price consumers pay is $P \cdot (1-r)$, making the inverse demand $\frac{P(X)}{1-r}$ and the price merchants receive $\frac{P\cdot(1-i)}{1-r}$.[4] The profit function of merchant j is

$$\pi_j = \frac{(1-i)}{(1-r)} P(X) x_j - c x_j.$$

In selecting its output each merchant j conjectures that other merchants' responses will be such that $\frac{dX}{dx_j} = \lambda$, the conjectural variation λ being taken as a fixed constant throughout. The case $\lambda = 1$ corresponds to the Cournot conjecture. When $\lambda = 0$, conjectures are 'competitive' and we obtain the Bertrand outcome. When $\lambda = n$, each firm believes that all other active firms will behave exactly as it does; tacit collusion among incumbent firms then being perfect (in the sense that aggregate profits are maximized conditional on the number of firms). It will be assumed throughout that $\lambda \in [0, n]$.[5]

The first order condition of the representative merchant is (omitting arguments)

$$\frac{\partial \pi_j}{\partial x_j} = \frac{(1-i)}{(1-r)} \left(P_X \frac{dX}{dx_j} x_j + P \right) - c = 0. \tag{4.1}$$

Restricting attention to symmetric equilibria, this becomes

$$\begin{aligned} \frac{(1-i)}{(1-r)} \left(P_X X \gamma + P \right) = c \quad &\Rightarrow \quad P \cdot \left(1 + \frac{\gamma}{\varepsilon} \right) = \frac{c \cdot (1-r)}{(1-i)} \\ &\Rightarrow \quad P = \frac{1}{\left(1 + \frac{\gamma}{\varepsilon}\right)} \frac{(1-r)}{(1-i)} c, \end{aligned} \tag{4.2}$$

where $\gamma \equiv \frac{\lambda}{n} \in [0, 1]$.[6] As $\frac{\gamma}{\varepsilon}$ increases, the merchant market becomes more competitive (either because the conduct parameter γ decreases, or because the market demand becomes more elastic) and price approaches marginal cost c (adjusted by the 'credit card tax'). A higher reward or a lower interchange fee has the same effect on price as a reduction in a merchant's marginal cost.

Let

$$E \equiv -\frac{P_{XX} X}{P_X} \tag{4.3}$$

denote the elasticity of the slope of inverse demand. In addition, ε' is the derivative of ε with respect to X.[7] The second order condition is $2 - \gamma E > 0$. The stability condition requires that $1 + \gamma \cdot (1 - E) > 0$

[4] We use $\cdot$ to distinguish between multiplication, i.e., $P \cdot (1-i)$, and a function $P(X)$, when the two are not immediately distinguishable from the context.

[5] See Seade (1980), Bresnahan (1981) and Delipalla and Keen (1992) for similar modeling frameworks.

[6] Note that γ is similar to the conduct parameter θ in Weyl and Fabinger (2013).

[7] When (inverse) demand is strictly concave, $E < 0$, when it is strictly convex, $E > 0$ and when demand is linear $E = 0$. Moreover, when demand is linear $\varepsilon' > 0$. For example, for the constant elasticity demand, $P = \left(\frac{1}{X}\right)^{1/k}$, $\varepsilon = -k$ and $E = \frac{1+k}{k}$ and so $\varepsilon' = 0$ and $E' = 0$.

(see, for example, Seade (1980) and Delipalla and Keen (1992)). Since $2 - \gamma E > 1 + \gamma \cdot (1 - E)$, the stability condition (which we assume is satisfied) is stronger than the second order condition. From the first-order condition (4.2), a non-negative price-cost margin implies that the elasticity must satisfy $\gamma + \varepsilon < 0$. Therefore, ε and E must fall in the admissible region $\varepsilon < -\gamma$ and $E < \frac{2}{\gamma}$, see also Figure 1 in Mrázová and Neary (2017).[8]

From (4.2), it follows that the price merchants receive is $P^m \equiv \frac{(1-i)}{(1-r)} P = \frac{c}{(1+\gamma/\varepsilon)}$; the price merchants charge is $\frac{P}{1-r} = \frac{1}{(1+\gamma/\varepsilon)} \frac{c}{(1-i)}$; and the price consumers (buyers) pay (after the reward is applied) is $P^b = \frac{1}{(1+\gamma/\varepsilon)} \frac{(1-r)}{(1-i)} c$.

From the price consumers pay, P^b, it becomes clear that equilibrium aggregate demand and merchant profits depend on the 'credit card tax', denoted by $z \equiv \frac{1-r}{1-i}$. The distribution of r and i, while z is fixed, has no effect on the equilibrium variables. Moreover, since the elasticity is a function of X, the elasticity is also a function of z. Hence, the equilibrium price merchants receive is

$$P^m(z) = \frac{c}{1 + \frac{\gamma}{\varepsilon(z)}} \tag{4.4}$$

and the equilibrium price consumers (buyers) pay is

$$P^b(z) = \frac{cz}{1 + \frac{\gamma}{\varepsilon(z)}} = zP^m(z). \tag{4.5}$$

The price merchants receive, (4.4), is not affected by the presence of the credit card, if competition in the product market is Bertrand, $\gamma = 0$, or the elasticity is constant. The credit card tax in these cases is passed on 100% to consumers. In all other cases, the price merchants receive decreases as z increases, if and only if $\frac{d\varepsilon}{dX} \equiv \varepsilon' > 0$. This is because the credit card tax lowers aggregate output, see (A.3), making demand more elastic, which induces the merchants to lower the price.

Efficiency dictates that $P = c$. We can have $P > c$ either because $\frac{\gamma}{\varepsilon(z)} < 0$, or $z > 1$, or both. The first source of inefficiency arises when competition in the merchant market is imperfect ($\gamma > 0$). Also note that the effect of the elasticity on $P^b(z)$ becomes stronger as the merchant market becomes less competitive, i.e., γ increases. The second source of inefficiency is due to the credit card tax levied by the payment network. Since $z > 1$, or equivalently $i > r$ (otherwise network profit cannot be positive), there is a double-marginalization: the first mark-up is from the merchants when they have market power and the second mark-up is from the payment networks (that have market power). A key issue we address with our analysis is how the distortions from the network side of the market interact with the distortions from the merchant side of the market.

The credit card tax z is an ad valorem tax that creates a wedge between the price consumers pay and

[8] In Mrázová and Neary (2017) the elasticity of the slope of the inverse demand is denoted by ρ and the elasticity ε is a positive number.

the price merchants receive. Hence, some of the results we derive, in particular the ones regarding credit card tax pass-through and incidence, are well-known in the public economics literature, e.g., Delipalla and Keen (1992) and Auerbach and Hines Jr (2002).

4.2 Stage 3: Acquiring and issuing banks' decisions

Acquiring banks compete in merchant fees m. The network interchange fee i is each acquiring bank's marginal cost. Given that acquiring banks are homogeneous and compete for merchants à la Bertrand the equilibrium merchant fee m will be equal to i. Acquiring banks earn zero profits in equilibrium.

Issuing banks compete in rewards. Each issuing bank receives from the network part of the interchange fee $f < i$. This is the maximum amount, per dollar of transactions, that each issuing bank can give to users as a reward. Given that issuing banks are homogeneous and compete à la Bertrand for users the equilibrium reward r will be equal to f. Issuing banks earn zero profits in equilibrium.

4.3 Stages 1 & 2: Network's decisions

Given that in equilibrium in stage 3, $m = i$ and $r = f$, the network profit consists of the interchange fee minus the rewards to consumers (both in percentages) times the total value of transactions in the market, $(i - r)\frac{P(X(z))}{1-r}X(z)$. Equivalently, the network profit is the difference between aggregate revenue from the product market minus the aggregate revenue captured by the merchants. Using (4.4) and (4.5) the network profit can be expressed as follows

$$\pi_1(z) = (P^b(z) - P^m(z))X(z) = \frac{c \cdot (z-1)}{1 + \frac{\gamma}{\varepsilon(z)}}X(z). \tag{4.6}$$

All variables depend on z but for brevity in what follows we omit this dependence. The network chooses z to maximize its profits.

The Proposition below summarizes the subgame-perfect equilibrium in the market.[9]

Proposition 1 *The subgame-perfect equilibrium credit card tax must (implicitly) satisfy*

$$z^* = \frac{\gamma X \varepsilon' + \varepsilon \cdot (\varepsilon + \gamma)}{\gamma X \varepsilon' + \varepsilon \cdot (1 + \varepsilon + \gamma \cdot (2 - E))}. \tag{4.7}$$

and (4.5).

The above expression reveals that the various parameters of the demand function, i.e., ε, ε' and E,

[9] We label Lemma or Proposition the results of which there is a general proof. We label Result the results that are derived from a numerical analysis.

together with the parameter that captures the intensity of product market competition, γ, are crucial in determining the equilibrium credit card tax. Moreover, the model so far predicts the equilibrium credit card tax z, but does not yield a unique i, r pair.

It can be verified that

$$\varepsilon' = \frac{1}{X}\left(1 - \varepsilon \cdot (1 - E)\right), \tag{4.8}$$

so $\varepsilon' > 0 \Leftrightarrow \varepsilon < \frac{1}{1-E}$.[10] The locus $\varepsilon = \frac{1}{1-E}$ is an increasing and concave function in the admissible region ($\varepsilon < -\gamma$ and $E < \frac{2}{\gamma}$). This is the SC curve in Figures 2-4 in Mrázová and Neary (2017), which determines whether a demand is superconvex or subconvex.[11] Any point above the locus corresponds to $\varepsilon' > 0$, or subconvex demand, while any point below the locus to $\varepsilon' < 0$, or superconvex demand.[12]

In the public economics literature the tax is usually taken as a parameter, while in our model it is endogenously determined. Moreover, how it is affected by competition in the merchant market is one of the issues we are interested in.[13]

4.4 Specific demands

Given the complicated nature of the problem, even before we introduce network competition, we utilize specific examples of functional forms in order to shed more light on how the merchant market structure interacts with the network in affecting the equilibrium variables and price distortions.

Recall that the parameter $\gamma \equiv \frac{\lambda}{n} \in [0, 1]$ measures the competitiveness of the merchant market either due to the mode of competition or the number of merchants. The value of $\gamma = 0$ corresponds to Bertrand competition, while $\gamma = 1$ corresponds to a monopoly merchant or to n merchants who have formed a perfect cartel; $\gamma = \frac{1}{2}$ can correspond to Cournot competition, $\lambda = 1$, between two merchants $n = 2$, and so on. Hence, as γ increases the product market becomes less competitive. In Table 1 we list the signs of the various parameters for the demand functions we use in the analysis that follows.

The elasticity, ε, is always negative and it can be constant, increasing or decreasing in aggregate output. How aggregate output affects the elasticity, ε', will be very important for the subsequent analysis as it determines the sign of the elasticity effect that will be introduced shortly. The Generalized Pareto demand is quite flexible as it allows for ε' to take on any sign, i.e., demand can be either superconvex or

[10] Mrázová and Neary (2017) introduce the concept of a 'demand manifold' and show that, for all demand functions–other than the CES–that satisfy some mild conditions, the manifold is represented by a smooth curve in the (ε, E) space. The usefulness of this result lies in demonstrating that knowing the values of elasticity and convexity of demand a firm faces is sufficient to predict its responses to a wide range of exogenous shocks, such as changes in taxes.

[11] The SC curve is decreasing and convex because Mrázová and Neary (2017) use the absolute value of the elasticity, while in our paper ε is a negative number.

[12] Superconvexity of the inverse demand function is equivalent to superconvexity of the direct demand function, and implies log-convexity of the inverse demand function, which implies log-convexity of the direct demand function, which implies convexity of both demand functions; but the converses do not hold, see Lemma D.1 in Mrázová and Neary (2019).

[13] In the public economics literature, even if the tax is chosen to maximize an objective, the impact of product market competition is rarely an issue of interest.

Types of demand functions	ε	E	ε'
Constant elasticity	$-$	$+$	0
Linear	$-$	0	$+$
Generalized Pareto	$-$	$-,0,+$	$-,0,+$

Table 1: The demand and inverse demand slope elasticities and how they change with aggregate output for the three types demands we use in the examples

subconvex. The elasticity of the slope of the inverse demand, E, can be zero, positive or negative. All demand functions we use are within the class of constant elasticity of the inverse demand, i.e., $E' = 0$.

The Generalized Pareto demand we consider in Section 4.4.3 encompasses the constant elasticity and linear demands as special cases. In our analysis, we initially present findings utilizing both constant elasticity and linear demand functions. Subsequently, we expand our examination to include results derived from the Generalized Pareto demand function. Each scenario contributes uniquely to our comprehension of the market dynamics at play.

4.4.1 Constant elasticity demand

We consider a constant elasticity demand (as in footnote 7). Using (4.2), the equilibrium aggregate quantity as a function of the credit card tax is

$$X(z) = \frac{(k-\gamma)^k}{(ck)^k}\frac{1}{z^k}.$$

The price merchants receive in the presence of a credit card is the same as the price without a credit card (i.e., P^m below is not a function of z),

$$P^m = \frac{ck}{k-\gamma} \geq c.$$

The profits of merchants are positive when $\gamma > 0$. The profit function of the network is

$$\pi_1(z) = \left(\frac{k-\gamma}{ck}\right)^{k-1}\frac{(z-1)}{z^k}. \tag{4.9}$$

The solution to the first order condition with respect to z is

$$z = \frac{k}{k-1}. \tag{4.10}$$

Observe that the profit-maximizing z is not a function of competition, γ. This can be best understood by inspecting the network profit function, (4.6). When ε is constant, and hence not a function of z, then γ only affects the level of the price and profits, but there is no interaction between γ and z.

As demand becomes more elastic, $k \to \infty$, the tax burden decreases, $z \to 1$, implying that the network increases the reward and/or decreases the interchange fee.

The equilibrium profit of merchant j is

$$\pi_j = \frac{c\gamma}{k-\gamma}\left(\frac{(k-\gamma)(k-1)}{ck^2}\right)^k. \tag{4.11}$$

It can be verified that π_j is increasing in γ. The price merchants receive is not affected by the presence of the credit card, but profits are lower because aggregate output X decreases since consumers pay a higher price.

After substituting (4.10) into (4.9) the profit function of the network becomes

$$\pi_1 = \left(\frac{(k-\gamma)(k-1)}{ck^2}\right)^{k-1}\frac{1}{k}. \tag{4.12}$$

Consumers bear the entire burden of the credit card tax by paying the following price

$$P^b = \frac{ck^2}{(k-1)(k-\gamma)}.$$

We summarize in the Proposition below.

Proposition 2 *Suppose the market demand is of the constant elasticity form. More intense competition in the product market, i.e., lower γ, does not affect the credit card tax, lowers merchant profits, but increases consumer surplus and the monopoly network profits. Consumers pay the entire burden of the tax, regardless of the intensity of competition in the product market.*

As the product market becomes more competitive, the price merchants charge decreases, their profits decrease, but the value of transactions increases. Consequently, the profits of the network, which are an increasing function of the value of transactions, increase as well.

4.4.2 Linear demand

When we depart from constant elasticity, the equilibrium credit card tax is a function of product market competition. From (4.6) it is clear that γ interacts with z through the elasticity ε, when the latter is not constant. The influence of the elasticity on the equilibrium price is stronger the less competitive the merchant market is. For example, when the market is perfectly competitive demand elasticity has no effect on pricing; on the other hand, for monopoly pricing market elasticity matters a lot. Furthermore, the effect of γ on z crucially depends on the sign of ε', or whether demand is superconvex or subconvex. We term this the *elasticity effect*, that is described below.

When $\varepsilon' > 0$ (subconvex demand), as aggregate output X decreases demand becomes more elastic. In this case there is a negative effect on marginal profitability when the network increases its z. This effect is coming through the elasticity: a higher tax z lowers aggregate output X which makes the demand more elastic and hence the price merchants charge drops, lowering the network revenue, all else equal. So, when the elasticity effect becomes stronger the network is more reluctant to increase its tax. A higher γ, i.e., weaker competition in the product market, makes this effect stronger. The network then optimally responds by lowering its tax.

The simplest case where $\varepsilon' > 0$ is the linear demand, $P = 1 - X$. Then, we have $\varepsilon = -\frac{1-X}{X}$, $E = 0$ and $\varepsilon' = \frac{1}{X^2} > 0$. Using (4.2), the equilibrium aggregate quantity is

$$X(z) = \frac{1 - cz}{1 + \gamma}.$$

The price merchants receive is

$$P^m(z) = \frac{cz + \gamma}{z \cdot (1 + \gamma)},$$

which is a function of z, unless $\gamma = 0$. The price buyers pay is

$$P^b(z) = \frac{cz + \gamma}{1 + \gamma}.$$

The profit function of the network is

$$\pi_1(z) = \frac{(\gamma + cz)(z - 1)(1 - cz)}{z(1 + \gamma)^2}. \tag{4.13}$$

We begin by assuming that merchants compete à la Bertrand so that $\gamma = 0$. The network profit function becomes

$$\pi_1(z) = c \cdot (z - 1)(1 - cz). \tag{4.14}$$

The profit-maximizing z, which is not a function of γ, is

$$z = \frac{1 + c}{2c}. \tag{4.15}$$

The maximized network profit is

$$\pi_1 = \frac{(1 - c)^2}{4}. \tag{4.16}$$

Note that the network earns the monopoly profit, i.e., the profit of a monopolist merchant with marginal cost c and no credit card tax, $z = 1$. This is because $P^m(z) = c$ and as a result the network maximizes the industry profit, see (4.6).

The price merchants receive with and without the presence of credit card is c while consumers pay c when there is no credit card and pay $\frac{1+c}{2} > c$ with a credit card. As with the constant elasticity demand, consumers bear the entire burden of the tax. This is a feature of Bertrand competition and the associated lack of market power on part of the merchants.

When competition in the product market is imperfect, i.e., $\gamma > 0$, the credit card tax depends on γ. Closed form solutions for z in this case are very long, given that the network profit function, (4.13), is cubic in z. In this case, the network maximizes the residual industry profit (i.e., after subtracting the equilibrium profits of the merchants). We present the equilibrium by setting a specific marginal cost, $c = 0.8$. Table 2 depicts the equilibrium credit card tax z, the price consumer pay, P^b, the price merchants receive, P^m, and the tax incidence for various values of γ. The price consumers pay is 12.13% higher than the price merchants receive when $\gamma = 1$ and the percentage difference becomes 12.5% when $\gamma = 0$.

	$\gamma = 1$	$\gamma = 0.75$	$\gamma = 0.5$	$\gamma = 0$
Credit card tax	1.1213	1.1218	1.122	1.125
Network profits	0.005	0.006	0.007	0.01
Merchant profits	0.00236	0.0023	0.002	0
Price consumers pay	0.948	0.94	0.932	0.9
Price merchants receive	0.846	0.84	0.83	0.8
$\%$ of the 'tax' consumers pay	47.29%	54.47%	64.23%	100%

Table 2: Monopoly network: Equilibrium values, when market demand is linear, as a function of the intensity of competition in the product market, γ.

We summarize in the following Result.

Result 1 *Suppose the product demand is linear (subconvex). As competition in the product market intensifies, i.e., γ decreases, the credit card tax increases, the monopoly network and consumers become better off (but consumers bear a higher fraction of the credit card tax burden), while the merchants become worse off.*

Similar to the constant elasticity demand scenario, a more competitive product market diminishes merchant profits while augmenting those of the network. However, in contrast to the constant elasticity case, the network responds to increased market competitiveness by raising the credit card tax it imposes. This is rooted in the elasticity effect we discussed earlier. As competition intensifies among merchants–that is, as γ decreases–the influence of the credit card tax z on pricing and network profits through elasticity weakens (refer to Equation (4.6)). Consequently, if $\varepsilon' > 0$, networks become more inclined to raise their tax, leading to a higher credit card tax in equilibrium as the merchant market grows more competitive. It's noteworthy that the two primary inefficiencies-stemming from merchant pricing and network fees-trend in divergent directions as the merchant market's competitiveness wanes. This may be a new and novel finding for antitrust analysis.

4.4.3 Generalized Pareto demand

Within this class of demand functions we can choose parameters so that $\varepsilon' < 0$ (superconvex demand), and consequently the elasticity effect works in the opposite direction than in the linear demand example.

The distribution of consumer valuations v takes on the generalized Pareto distribution

$$F(v) = 1 - (1 + \xi \cdot (E-1)(v-1))^{\frac{1}{1-E}},$$

where $\xi > 0$ is the scale parameter and $E < 2$ is the shape parameter, see Bulow and Klemperer (2012) and Wang and Wright (2018). A lower ξ implies higher consumer willingness to pay in the first-order stochastic dominance sense. The generalized Pareto distribution implies the corresponding demand functions for merchants are defined by the class of demands

$$X(p) = 1 - F(p) = (1 + \xi \cdot (E-1)(p-1))^{\frac{1}{1-E}}, \tag{4.17}$$

with constant elasticity of the slope of the inverse demand given by E, see (4.3). When $E < 1$, the support of the distribution F is $\left[1, 1 + \frac{1}{\xi \cdot (1-E)}\right]$ and it has increasing hazard. Accordingly, the implied demand functions X are log-concave and include the linear demand function ($E = 0$) as a special case. Alternatively, when $1 < E < 2$, the support of F is $[1, \infty)$ and it has decreasing hazard. The implied demand functions are log-convex and include the constant elasticity demand function $(E = 1 + \frac{1}{\xi})$ as a special case. When $E = 1$, F captures the left-truncated exponential distribution $F(x) = 1 - e^{-\xi \cdot (v-1)}$ on the support $[1, \infty)$, with a constant hazard rate ξ. This implies the exponential (or log-linear) demand $X = e^{-\xi \cdot (p-1)}$.

The effect of aggregate output on the elasticity is given by

$$\varepsilon' = \frac{1 - \xi \cdot (E-1)}{X^{2-E}}, \tag{4.18}$$

which is negative if and only if $E > 1 + \frac{1}{\xi}$. Using (4.2), the equilibrium aggregate quantity as a function of the credit card tax z is

$$X(z) = \left(\frac{1 - \gamma \cdot (E-1)}{1 - \xi \cdot (E-1)(1 - cz)} \right)^{\frac{1}{E-1}}.$$

The sign of ε', which affects the elasticity effect, also determines the degree of credit card tax pass-through. We can easily show, using $X(z)$ from above, that the equilibrium price when $z = 1$, i.e., zero tax, is $P = \frac{(1-\xi\cdot(E-1))\gamma+c\xi}{(1-\gamma\cdot(E-1))\xi}$. Using this price and equation (2.10) for ad-valorem tax pass-through in Delipalla and Keen (1992), over-shifting occurs, evaluated at $t_v = 0$, when $\frac{dP}{dt_v} > P$ (where t_v is the ad-valorem tax), which holds if and only if $E > 1 + \frac{1}{\xi}$, or $\varepsilon' < 0$.[14] Therefore, when $\varepsilon' < 0$, and $\gamma > 0$, consumers pay more than 100% of the credit card tax, as expected given that demand is superconvex.

We proceed by offering a numerical example. We choose the parameters so that $\varepsilon' < 0$: $E = 1.2$ and $\xi = 40$. We also set $c = 1$. Table 3 depicts the equilibrium credit card tax z, the price consumer pay, P^b, the price merchants receive, P^m, and the tax incidence for various values of γ.

	$\gamma = 1$	$\gamma = 0.75$	$\gamma = 0.5$	$\gamma = 0$
Credit card tax	1.0315	1.03142	1.03136	1.03125
Network profits	0.00348	0.00466	0.00615	0.01024
Merchant profits	0.004	0.0023	0.003248	0
Price consumers pay	1.07062	1.0590	1.0487	1.03125
Price merchants receive	1.038	1.0268	1.0168	1
$\%$ of the 'tax' consumers pay	120.4%	114.6%	109.3%	100%

Table 3: Monopoly network: Equilibrium values, when demand is (superconvex) generalized Pareto ($E = 1.2$ and $\xi = 40$), as a function of the intensity of competition in the product market, γ.

We summarize in the following Result.

Result 2 *Suppose the product demand is (superconvex) Generalized Pareto, i.e., $\varepsilon' < 0$. As competition in the product market intensifies, i.e., γ decreases, the credit card tax decreases, the monopoly network and consumers become better off (and consumers bear a lower fraction of the credit card tax burden), while the merchants become worse off. Consumers pay more than 100% of the tax, suggesting that merchants are better off with the credit card tax than without it.*

Compared to scenarios where $\varepsilon' \geq 0$, such as with constant elasticity and linear demand functions, two notable distinctions arise. First, as the product market intensifies in competitiveness, the equilibrium tax is observed to decrease, a phenomenon attributable to the previously mentioned elasticity effect. Specifically, when $\varepsilon' < 0$, networks gain an advantage by increasing the credit card tax due to the resulting less elastic demand when aggregate output falls. This advantage wanes as the merchant market's competitiveness diminishes, i.e., as γ decreases (refer to Equation (4.6)), leading to a reduction in the equilibrium credit card tax. Second, the phenomenon of tax over-shifting occurs, whereby consumers end up paying in excess of 100% of the credit card tax. Moreover, as product market competition escalates, the proportion of the tax borne by consumers diminishes. The imposition of the credit card tax reduces aggregate output, and with demand becoming less elastic, the prices charged by merchants escalate, resulting in a post-tax price that exceeds the price in the absence of the tax.

[14] Weyl and Fabinger (2013) offer a general analysis regarding tax pass-through and tax incidence, but only under a specific tax.

As the merchant market loses competitiveness, the dual inefficiencies–originating from merchant pricing and network fees–tend to exacerbate.

5 Analysis with two competing networks and a fixed number of users

We allow for two networks, but each network has a fixed number of users. This implies that network fees affect only the volume of transactions in a given network. One of the two networks can be interpreted as cash; more on this in section 7. More precisely, suppose a fixed fraction μ of (a unit mass) consumers uses card 1 and the remaining fraction uses card 2. The assumption of a fixed number of users is reasonable for short-run competition between the networks and also allows us to disentangle the various effects of network competition when we later, in Section 6, endogenize the number of users. Hence, until Section 6, we do not allow consumers to multi-home and all merchants accept both cards. Let x_l be the individual consumer consumption using credit card l and $X_l = \mu_l x_l$ be the aggregate output purchased using credit card l. The inverse demand is $P(X, r_1, r_2)$, where $X = X_1 + X_2$ is aggregate output.[15]

5.1 Stage 4: Merchant competition

We assume that for each merchant j a fraction μ of its sales are paid with credit card 1 and a fraction $1 - \mu$ with credit card 2. Each merchant takes the rewards and the interchange fees as given and chooses its output x_j to maximize profits given by

$$\pi_j = (P(X, r_1, r_2)(\mu \cdot (1 - i_1) + (1 - \mu)(1 - i_2)) - c)\, x_j.$$

Let $i \equiv \mu i_1 + (1 - \mu)i_2$ be the weighted average interchange fee. Also let $\mathcal{P}(X, i, r_1, r_2) \equiv (1 - i)P(X, r_1, r_2)$ be the price merchants receive. Since $1 - r_l$ multiplies the price P to determine the price consumers pay, if $1 - r_1$ and $1 - r_2$ decrease by $\kappa\%$ the consumer willingness to pay will increase by the same percentage for any given aggregate output X. Therefore, $P(X, r_1, r_2)$ is homogeneous of degree one in $\frac{1}{1-r_1}$ and $\frac{1}{1-r_2}$ and $\mathcal{P}(X, i, r_1, r_2)$ becomes a function of z_l, $\mathcal{P}(X, z_1, z_2)$, where $z_l \equiv \frac{1-r_l}{1-i}$, as in the monopoly network case. The first order condition of the representative merchant is (omitting

[15] Consider the following two examples. There are two goods, x and a numeraire good y whose price is normalized to one. If each consumer has the following quasi-linear utility $U = \frac{kx(1/x)^{1/k}}{k-1} + y$, then the inverse aggregate demand is of the constant elasticity type and is given by

$$P = \left(\frac{1}{X}\right)^{1/k} \left(\frac{\mu}{(1 - r_1)^k} + \frac{1 - \mu}{(1 - r_2)^k}\right)^{1/k}. \tag{5.1}$$

Or, when $U = x - x^2/2$, the inverse aggregate demand is linear and, assuming $r_1 \geq r_2$, is given by

$$P = \begin{cases} \frac{\mu - X}{\mu(1-r_1)}, & \text{if } P \in \left(\frac{1}{1-r_2}, \frac{1}{1-r_1}\right) \\ \frac{1-X}{\mu(1-r_1)+(1-\mu)(1-r_2)}, & \text{if } P \leq \frac{1}{1-r_2}. \end{cases} \tag{5.2}$$

arguments)

$$\frac{\partial \pi_j}{\partial x_j} = \mathcal{P}_X \frac{dX}{dx_j} x_j + \mathcal{P} - c = 0. \tag{5.3}$$

Following similar steps as the ones for the derivation of (4.2) the equilibrium price merchants charge must (implicitly) satisfy (omitting arguments)

$$P = \frac{c}{\left(1 + \frac{\gamma}{\varepsilon}\right)} \frac{1}{(1 - i)}. \tag{5.4}$$

The price merchants receive is

$$P^m = \frac{c}{\left(1 + \frac{\gamma}{\varepsilon}\right)} \tag{5.5}$$

and the price consumers (buyers) who use credit card l pay is

$$P^{b_l} = \frac{c z_l}{\left(1 + \frac{\gamma}{\varepsilon}\right)}. \tag{5.6}$$

5.2 Stage 3: Acquiring and issuing banks' decisions

The equilibrium of this stage is very similar to the stage 3 equilibrium with a monopoly network, see Section 4.2. The equilibrium merchant fee is $m_l = i_l$ and the equilibrium reward is $r_l = f_l$.

5.3 Stage 2: Networks' decisions regarding f_l

Given the stage 3 equilibrium, this decision amounts to each network choosing the cardholder reward r_l and since i is fixed it implies that each network chooses z_l. Network l's profit is the difference between aggregate revenue from the product market using card l minus the aggregate revenue captured by the merchants. Using (5.5) and (5.6) it can be expressed as follows

$$\pi_l(z_l, z_{-l}) = (P^{b_l} - P^m) X_l = \frac{c \cdot (z_l - 1)}{1 + \frac{\gamma}{\varepsilon(z_l, z_{-l})}} \mu \cdot x_l(z_l). \tag{5.7}$$

The effect of network competition manifests itself through the elasticity of demand $\varepsilon(X)$, where $X = \mu x_1(z_1) + (1 - \mu) x_2(z_2)$. If the elasticity is constant, as in the next example, or when $\gamma = 0$, as in Bertrand product market competition, then network competition, relative to a monopoly network, has no effect on the equilibrium credit card tax. Only the network profits will be affected.

Next, let's assume that competition in the product market is imperfect, $\gamma > 0$, and the elasticity of product demand is not constant. Suppose the market initially is occupied by a monopoly incumbent network, $\mu = 1$, and a second network enters the market with infinitesimal fixed market share, i.e., $\mu < 1$, but in the neighborhood of one (so that aggregate output is kept fixed; otherwise there will be additional effects). If aggregate output had no effect on the elasticity, ε, then from (5.7) it follows that entry would have no effect on the incumbent's equilibrium tax (only profits would be affected). Now suppose ε does depend on z_l through X and $\varepsilon' > 0$. There is a negative effect, the *elasticity effect* we identified in Section 4.4.2, on marginal profitability when network l, the incumbent, increases its z_l, that is coming through the elasticity (i.e., higher tax z_l lowers aggregate output X which makes the demand more elastic and hence the price merchants charge drops, lowering the network revenue, all else equal). So, when the elasticity effect is at work the network is more reluctant to increase its tax. When a second network enters, the effect of the incumbent's z_l on aggregate output X, and hence on the elasticity, is diminished, because μ decreases. This in turn implies that the negative elasticity effect is weaker and hence the incumbent network would be less reluctant to increase its z_l.

Each network exerts an externality on the other when it increases its tax, even when networks do not compete for users. Higher z_l lowers aggregate output and thus the elasticity of demand if $\varepsilon' > 0$. This induces merchants to lower the price they charge, which lowers the profit of the rival network, all else equal. Hence, the externality is negative. It becomes positive if $\varepsilon' < 0$.

We summarize in the Proposition below.

Proposition 3 *Suppose the market initially is occupied by a monopoly incumbent network, product demand becomes more (less) elastic as aggregate output decreases, $\varepsilon' > 0$ ($\varepsilon' < 0$) and $\gamma > 0$. Then, entry of a second network, with an infinitesimally small and fixed number of users that is poached from the incumbent, induces the incumbent to increase (decrease) its equilibrium tax.*

The result that entry of a second network can increase the credit card tax is surprising. It has to do with the interaction between the network and the product market and it critically depends on whether the demand is subconvex ($\varepsilon' > 0$) or superconvex ($\varepsilon' < 0$).

Because what matters for the equilibrium profits is the equilibrium z_l, the choice of i_l in stage 1 becomes irrelevant. For any choice of i_1 and i_2 in stage 1, the f_1 and f_2 will be such that in stage 2 so that the market reaches the equilibrium z_1, z_2. To put differently, as we are solving the game backwards, once the equilibrium level of rewards is determined in stage 2 the interchange fees i_1 and i_2 disappear from the profit functions. Thus, as in the monopoly network case, the model cannot pin down a unique equilibrium in terms of i_l and r_l.

The specific examples that are presented next demonstrate that the surprising result of Proposition 3 can hold even if the market share of the entrant is significant.

5.4 Specific demands

5.4.1 Constant elasticity demand

We assume a constant elasticity demand as in footnote 15. Using (5.1) and (5.4) the equilibrium aggregate output is

$$X(z_l, z_{-l}) = \frac{(k-\gamma)^k}{(ck)^k}\left(\frac{\mu}{z_l^k} + \frac{1-\mu}{z_{-l}^k}\right). \tag{5.8}$$

The profit function of network l is

$$\pi_l(z_l) = \left(\frac{k-\gamma}{ck}\right)^{k-1}\frac{\mu\cdot(z_l-1)}{z_l^k}. \tag{5.9}$$

Note that the profit function (5.9) is the same as in the case of a monopoly network, (4.9), (with the exception of the μ term) and hence the equilibrium tax z_l will not change, so $z_l^* = \frac{k}{k-1}$. Equilibrium network profits will be lower than in the monopoly case as long as $\mu < 1$. Because the number of users of each network is fixed and the elasticity effect is mute ($\varepsilon' = 0$), entry of a second network has no effect on the equilibrium credit card taxes.

5.4.2 Linear demand

When demand is linear the elasticity effect is in force and $\varepsilon' > 0$. We assume a linear demand as in footnote 15. Suppose w.l.o.g. that $r_1 \geq r_2$, so that $z_1 \leq z_2$. It follows from the inverse demand, (5.2), that both groups of consumers make purchases if $P \leq \frac{1}{1-r_2}$. Then, the equilibrium aggregate output is

$$X(z_1, z_2) = X_1(z_1) + X_2(z_2) = \mu\cdot\left(\frac{1-cz_1}{1+\gamma}\right) + (1-\mu)\cdot\left(\frac{1-cz_2}{1+\gamma}\right). \tag{5.10}$$

For consistency, we need to ensure that P is indeed less than $\frac{1}{1-r_2}$. This holds if and only if

$$z_1 > \frac{((-1+(1+\gamma)\mu) - cz_2\cdot(\mu-1))z_2}{\mu((1+\gamma)-cz_2)}. \tag{5.11}$$

Consumers who use card 2 are effectively willing to pay a lower price than those who use card 1 because their reward is lower. Condition (5.11) is necessary for $X_2 > 0$. If (5.11) is not satisfied, network 1 corners the market; it offers a high reward to consumers who make transactions with its credit card (relative to the reward offered by the rival network), which drives the market price up and results in zero transactions for the consumers who use the rival credit card. In a symmetric equilibrium, $z_1 = z_2 = z$, both groups of consumers make purchases as long as $z < \frac{1}{c}$. We search for symmetric equilibria.

When $1 < z_1 \leq z_2 < \frac{1}{c}$ and (5.11) is satisfied, the profit functions of the networks are

$$\begin{aligned}\pi_1(z_1, z_2) &= \frac{(1 - cz_1)\mu \cdot (z_1 - 1)(c \cdot (z_1 - z_2)\mu + \gamma + cz_2)}{(1 + \gamma)^2((z_1 - z_2)\mu + z_2)} \\ \pi_2(z_1, z_2) &= \frac{(1 - cz_2)(1 - \mu)(z_2 - 1)(c \cdot (1 - \mu)z_2 + c\mu z_1 + \gamma)}{((1 - \mu)z_2 + \mu z_1)(1 + \gamma)^2}.\end{aligned} \tag{5.12}$$

We differentiate π_1 with respect to z_2 to obtain

$$\frac{\partial \pi_1}{\partial z_2} = \frac{\mu \cdot (1 - \mu)(z_1 - 1)(cz_1 - 1)\gamma}{((z_1 - z_2)\mu + z_2)^2(1 + \gamma)^2} \leq 0, \tag{5.13}$$

given that $z_1 \in \left(1, \frac{1}{c}\right)$. A similar result holds for the effect of z_1 on π_2. As we have discussed before, each network imposes a negative externality (unless $\gamma = 0$) on the competing network.

When $\gamma = 0$, the equilibrium profit functions can be expressed as follows

$$\pi_1(z_1) = \mu c \cdot (z_1 - 1)(1 - cz_1) \text{ and } \pi_2(z_2) = (1 - \mu)c \cdot (z_2 - 1)(1 - cz_2).$$

The profit functions are the same, with the exception of μ, as in the monopoly network case, see (4.14). The equilibrium taxes are given by

$$z_1^* = z_2^* = \frac{1 + c}{2c}. \tag{5.14}$$

Hence, entry of a second network, when the product market is characterized by Bertrand competition, leaves the rewards and interchange fees unchanged and only lowers individual network profits (as in the constant elasticity demand example of Section 5.4.1).

Let's now assume that competition in the product market is imperfect, $\gamma > 0$. The equilibrium expressions are very long and hence we present the equilibrium by setting $c = 0.8$ and $\mu = 0.5$. Table 4 depicts the equilibrium tax z_l, the price consumers pay, P^{b_l}, the price merchants receive, P^m, and the tax incidence for various values of γ.

	$\gamma = 1$	$\gamma = 0.75$	$\gamma = 0.5$	$\gamma = 0$
Credit card tax	1.1231	1.1234	1.1237	1.125
Network profits	0.0026	0.003	0.0034	0.005
Merchant profits	0.0023	0.0022	0.002	0
Price consumers pay	0.949	0.942	0.933	0.9
Price merchants receive	0.845	0.839	0.83	0.8
% of the 'tax' consumers pay	47.32%	54.51%	64.26%	100%

Table 4: Duopoly network: Equilibrium values, when market demand is linear and network shares are fixed, as a function of the intensity of competition in the product market, γ

The result below is derived by comparing the results in Table 2 (where $\mu = 1$) with those in Table 4.

These results are consistent with Proposition 3.

Result 3 *Suppose the product demand is linear. Entry of a second network that captures exogenously 50% of the market share, holding the intensity of competition in the product market fixed, increases the equilibrium credit card tax and the price consumers pay and decreases merchant profits, consumer surplus and network profits. In sum, it decreases social welfare. Consumers bear a higher fraction of the credit card tax burden, than when there is a monopoly network.*

5.4.3 Generalized Pareto demand

Demand is Generalized Pareto as in Section 4.4.3. As we have discussed before, an interesting aspect of the Pareto demand is its flexibility, at the expense of having a complex functional form and hence we must resort exclusively to a numerical analysis. Depending on the value of the shape parameter E relative to a function of the scale parameter ξ, the slope of the elasticity of demand with respect to aggregate output can be positive (subconvexity) or negative (superconvexity). Specifically, $\varepsilon' > 0 \Leftrightarrow E < 1 + \frac{1}{\xi}$.

As in the numerical example of Table 3 under a monopoly network, we set $c = 1$, $\xi = 40$ and $E = 1.2$. Unlike that example, we now assume a symmetric duopoly network structure, i.e., $\mu = 0.5$. Under the assumed values of the parameters, $\varepsilon' < 0$. Table 5 presents the equilibrium values.[16] Comparing the duopoly equilibrium with the monopoly one presented in Table 3, we conclude that network competition increases the equilibrium tax and the price consumers pay, although consumers pay a lower fraction of the credit card tax. Furthermore, more intense competition in the merchant market decreases the equilibrium tax.[17]

	$\gamma = 1$	$\gamma = 0.5$	$\gamma = 0$
Credit card tax	1.036	1.034	1.032
Network profits	0.00172	0.003066	0.005119
Merchant profits	0.00359	0.0030297	0
Price consumers pay	1.07625	1.05166	1.032
Price merchants receive	1.0388	1.01708	1
% of the 'tax' consumers pay	120.325%	109.244%	100%

Table 5: Duopoly network with fixed shares: Equilibrium values, when demand is generalized Pareto with $E = 1.2$ and $\xi = 40$, as a function of the intensity of competition in the product market, γ.

To demonstrate the critical role of the elasticity of demand for the effect of network competition on consumer welfare, we change the scale parameter to $\xi = 4$ so that $\varepsilon' > 0$. Table 6 presents the equilibrium under a monopoly network, while Table 7 presents the equilibrium under a duopoly network.

[16]The equilibrium was calculated using the Matlab software. We performed a grid search with an increment of 0.001.

[17]Note that network competition increases the equilibrium credit card tax under a linear demand where $\varepsilon' > 0$ and under a Generalized Pareto with $\varepsilon' < 0$. Hence, while the sign of ε' matters greatly for the equilibrium, it is, in general (i.e., for any μ), neither sufficient nor necessary for the effect of network competition on the equilibrium credit card tax.

	$\gamma = 1$	$\gamma = 0.5$	$\gamma = 0$
Credit card tax	1.309	1.311	1.3125
Network profits	0.04354	0.06846	0.1024
Merchant profits	0.0323	0.0257	0
Price consumers pay	1.6989	1.4842	1.3125
Price merchants receive	1.2978	1.1323	1
% of the 'tax' consumers pay	96.32%	98.13%	100%

Table 6: Monopoly network with fixed shares: Equilibrium values, when demand is generalized Pareto with $E = 1.2$ and $\xi = 4$, as a function of the intensity of competition in the product market, γ.

	$\gamma = 1$	$\gamma = 0.5$	$\gamma = 0$
Credit card tax	1.304	1.308	1.3125
Network profits	0.02177	0.03423	0.0512
Merchant profits	0.03287	0.02598	0
Price consumers pay	1.6925	1.4811	1.3125
Price merchants receive	1.298	1.1323	1
% of the 'tax' consumers pay	96.30%	98.12%	100%

Table 7: Duopoly network with fixed shares: Equilibrium values, when demand is generalized Pareto with $E = 1.2$ and $\xi = 4$, as a function of the intensity of competition in the product market, γ.

By comparing Tables 6 and 7 we can see, in contrast to the previous example, that network competition, in this case, lowers the equilibrium credit card tax. Also, more intense competition in the merchant market increases the equilibrium tax, both under a monopoly and under a duopoly network. Consumers are better off under a duopoly network market structure and pay a lower fraction of the credit card tax.

Result 4 *Suppose the product demand is Generalized Pareto. Entry of a second network that captures exogenously 50% of the market share, holding the intensity of competition in the product market fixed, increases (decreases) the equilibrium credit card tax and decreases (increases) welfare when $\varepsilon' < 0$ ($\varepsilon' > 0$).*

5.5 Price discrimination

We digress momentarily from our main analysis to examine the effect of price discrimination. Suppose merchants charge different prices based on the credit card a consumer uses to make purchases. Under this assumption, there are two inverse demand functions $P_1(X_1)$ and $P_2(X_2)$. Network l's profit function is given by (5.7), but the elasticity is only a function of own tax, $\varepsilon(z_l)$. Entry of a second network does not affect the strength of the elasticity effect.

Proposition 4 *Suppose merchants price discriminate based on the credit card consumers use to make purchases. Then, entry of a second network that poaches a fixed number of users from the incumbent has* *no effect on the equilibrium tax.*

6 Analysis with two competing networks and an endogenous number of credit card users

We focus on an equilibrium where all merchants accept both cards and consumers single-home. Recall that all merchants face the same credit card tax from a given network. Because product demand is elastic, in a symmetric equilibrium where all merchants accept both cards, merchants earn strictly positive profits. Given that consumers single-home and merchants sell a homogeneous product, if a merchant deviates and stops accepting one card, it will lose all consumers who use that card and hence its profits will drop, making such deviation unprofitable. Given that all merchants accept both cards and the existence of a small cost of holding a second card, each consumer will single-home, i.e., use his most-preferred card.

The network tax affects not only the intensive but also the extensive margin. We assume that the two networks are horizontally differentiated. We model credit card differentiation using the spatial duopoly model of Hotelling (1929).[18] In particular, on the $[0, 1]$ interval, network 1 is located at 0, network 2 is located at 1 and users/consumers are uniformly distributed with density one. We assume that each consumer receives a gross benefit $V > 0$, pays a price P^{b_l} from making transactions with credit card l and incurs a linear per-unit of distance to a credit card transportation cost $t > 0$. The parameter t captures the degree of network differentiation, with a higher value indicating more differentiated networks. For tractability, we assume that V is independent of the volume of transactions. Thus, the consumer located at $x \in [0, 1]$, if he uses credit card 1 obtains a net utility $V - P^{b_1} - tx$ and if he uses credit card 2 obtains a net utility $V - P^{b_2} - t(1 - x)$. If network l offers higher rewards to its cardholders, then the price consumers pay using card l will decrease, which in turn will attract more users to network l. Hence, the mass of consumers who uses credit card 1 is given by

$$\mu = \frac{1}{2} + \frac{P^{b_2} - P^{b_1}}{2t} = \frac{1}{2} + \frac{c \cdot (z_2 - z_1)}{2t \cdot \left(1 + \frac{\gamma}{\varepsilon}\right)}, \tag{6.1}$$

where the elasticity, ε, is a function of aggregate output $X = \mu x_1(z_1) + (1 - \mu)x_2(z_2)$. Using implicit differentiation, the effect of own tax on own market share is

$$\frac{d\mu}{dz_1} = -\frac{\left(\gamma\mu \frac{dx_1}{dz_1} \cdot (z_1 - z_2)\varepsilon' + \varepsilon\right)(\varepsilon + \gamma)c}{c\gamma \cdot (x_1 - x_2)(z_1 - z_2)\varepsilon' + 2t \cdot (\varepsilon + \gamma)^2}.$$

Observe that as networks become sufficiently differentiated, $t \to \infty$, z_1 has no effect on market share, i.e., $\frac{d\mu}{dz_1} \to 0$. In this case, competition between networks would be equivalent to the case when μ

[18] See Jeon and Rey (2022) for a similar assumption regarding differentiation between two platforms.

is fixed. Using (6.1), the network profit function (5.7) can be expressed as follows

$$\pi_l(z_l, z_{-l}) = (P^{b_l} - P^m)X_l = \frac{c \cdot (z_l - 1)}{1 + \frac{\gamma}{\varepsilon(z_l, z_{-l})}} \mu(z_l, z_{-l}) x_l(z_l). \tag{6.2}$$

Aggregate output purchased with own credit card, X_l, depends not only on own tax z_l but also, through μ, on rival network tax z_{-l}.

When $\gamma = 0$, $\frac{d\mu}{dz_1} = -\frac{c}{2t} < 0$, and when the elasticity is constant ($\varepsilon' = 0$), $\frac{d\mu}{dz_1} = -\frac{\varepsilon c}{2t \cdot (\varepsilon + \gamma)} < 0$. In these cases, each network exerts a positive externality on the other when it increases its tax, because users switch networks. Recall that when each network has an exogenously fixed number of users, and $\gamma = 0$ or the elasticity is constant, there are no externalities, and when $\gamma > 0$ and $\varepsilon' > 0$ the externality is negative. The externality when $\gamma > 0$, $\varepsilon' > 0$ and networks compete for users depends on the relative strength of the above two extreme cases and it may change signs.

6.1 Specific demands

6.1.1 Constant elasticity demand

Recall from the analysis in Section 5.4.1 that, when the number of users μ is fixed, entry of a second network has no effect on the equilibrium tax and hence on consumer welfare. This is no longer true when networks compete for users. Using (5.9) and (6.1) the profit function of network l becomes

$$\pi_l(z_l, z_{-l}) = \left(\frac{k - \gamma}{ck}\right)^{k-1} \frac{(kc \cdot (z_{-l} - z_l) + t \cdot (k - \gamma)))\,(z_l - 1)}{2 z_l^k t \cdot (k - \gamma)}. \tag{6.3}$$

The symmetric equilibrium tax is

$$z_1^* = z_2^* = \frac{\sqrt{(k-1)^2(k-\gamma)^2 t^2 + 2ck \cdot (k+1)(k-\gamma)t + c^2k^2} - k^2 t + ((1+\gamma)t + c)k - t\gamma}{2ck}. \tag{6.4}$$

We summarize in the Proposition below.

Proposition 5 *Suppose the product demand is of the constant elasticity form and the number of users is endogenously determined. The symmetric equilibrium tax is given by* (6.4)*. Entry of a second network lowers the equilibrium tax,* $z_1^* = z_2^* < \frac{k}{k-1}$*, lowers the price consumers pay and increases social welfare. Finally,*

$$\lim_{t \to \infty} z_1^* = z_2^* = \frac{k}{k-1} \tag{6.5}$$

indicating that as networks become sufficiently differentiated, the equilibrium tax converges to the equilibrium tax when the number of users is fixed, or when there is a monopoly network.

When networks compete for users, so both the extensive and the intensive margins are flexible, there is a *competition effect*, that tends to lower the equilibrium taxes. When demand exhibits constant elasticity, only the competition effect is in force since the elasticity effect is absent. That is why entry of a second network decreases the credit card tax.

The linear demand that we examine next is the simplest case in which both the elasticity and the competition effects are operative.

6.1.2 Linear demand

We substitute $\varepsilon(X) = -\frac{1-X}{X}$ into (6.1) and solve for μ to derive the number of users as a function of z_1 and z_2 only (and parameters). This yields

$$\mu = \frac{\sqrt{(1+\gamma)^2(z_2+z_1)^2t^2 - 2(z_1-z_2)^2(1+\gamma)(c\cdot(z_1+z_2)+4\gamma)t + c^2(z_1-z_2)^4} - cz_1^2 + (2cz_2 + t\cdot(1+\gamma))z_1 - cz_2^2 - 3(1+\gamma)tz_2}{4t\cdot(1+\gamma)(z_1-z_2)}. \tag{6.6}$$

The network profit functions then are given by (5.12), with μ given by (6.6).

We begin the analysis by assuming Bertrand competition in the product market, i.e., $\gamma = 0$. The networks' profit functions become

$$\pi_1(z_1, z_2) = \frac{(t-(z_1-z_2)c)(z_1-1)(1-cz_1)c}{2t} \text{ and } \pi_2(z_1, z_2) = \frac{(t+(z_1-z_2)c)(z_2-1)(1-cz_2)c}{2t}.$$

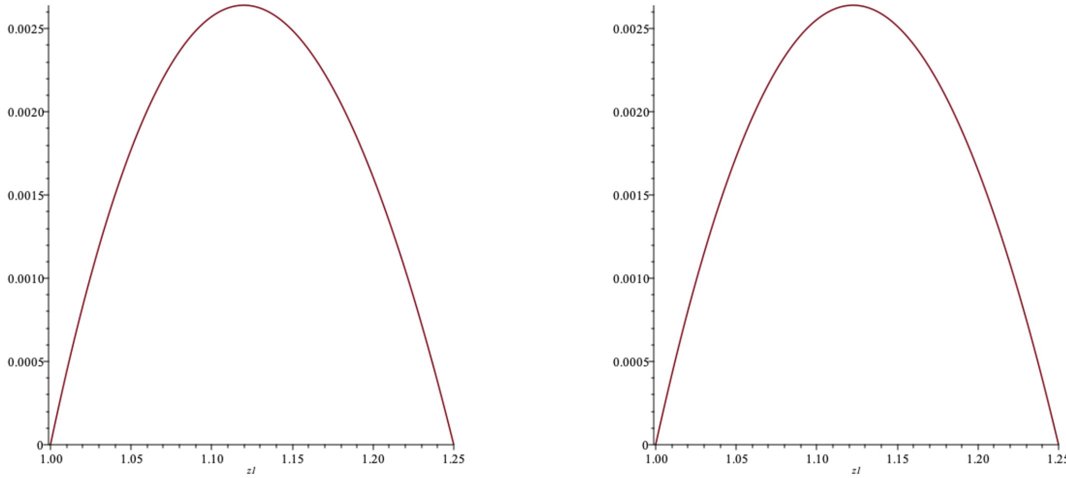
The symmetric equilibrium tax is

$$z_1^* = z_2^* = \frac{1+c+2t-\sqrt{(1-c)^2+4t^2}}{2c}. \tag{6.7}$$

It can be easily verified that (6.7) is less than the equilibrium tax when the number of users is fixed, $\frac{1+c}{2c}$ and $\lim_{t\to\infty} z_1^* = z_2^* = \frac{1+c}{2c}$. When $\gamma = 0$ the elasticity effect is mute as in the constant elasticity demand example of Section 6.1.1.

According to Proposition 3, when $\gamma > 0$ and $\varepsilon' > 0$ (which holds for the linear demand), entry of a second network increases the equilibrium tax when the extensive margin is fixed (fixed number of users). This is due to the elasticity effect. On the other hand, the competition effect tends to lower the equilibrium taxes. The degree of network differentiation affects the relative strength of these two effects. If the entrant network is sufficiently differentiated from the incumbent's network, then the competition effect is weak. In this case entry can lead to higher equilibrium taxes and lower consumer welfare. This is confirmed next.

We illustrate the result using the following numerical example. We set $\gamma = 1$ and $c = 0.8$. From Table 2 we see that when $\gamma = 1$, $z^* = 1.1213$ under a monopoly network. When μ is not fixed and $t = 2$, $z^* = 1.1198$, lower than 1.1213. Hence, entry of a second network and competition for users



(a) $\pi_1(z_1, z_2^* = 1.1198)$ (b) $\pi_1(z_1, z_2^* = 1.1225)$

Figure 2: Profit functions of network 1 when $t = 2$ (a) and $t = 10$ (b), holding z_2 fixed at its equilibrium values

lowers the equilibrium tax and consequently the price consumers pay. But when $t = 10$, so networks are more differentiated, the equilibrium tax is $z^* = 1.1225$, which is higher than 1.1213.[19]

The above examples lead to the following Result.

Result 5 *Suppose the product demand is linear and networks compete for users. Entry of a second network, holding the intensity of competition in the product market fixed, lowers the equilibrium taxes and increases social welfare if networks are not too differentiated, while it increases the equilibrium taxes and decreases social welfare if networks are sufficiently differentiated.*

In most two-sided market models, the prices each side pays are disconnected. Hence, more intense network (or platform) competition can increase one price but decrease the other, as in Guthrie and Wright (2007). Overall welfare implications are ambiguous as some groups benefit while others are hurt. In our model the two prices, interchange fees and rewards, are linked through the price consumers pay and the price merchants receive, and is captured by the credit card tax. More intense network competition can increase the credit card tax and yield unambiguous welfare predictions.

[19]Figure 2 depicts the profit functions of network 1 when z_2 is set at the the equilibrium values. The profit functions of network 2 are symmetric. These profits functions are quasi-concave, continuous over a compact set and illustrate that networks have no incentive to deviate from the solutions to the first order conditions. Therefore, the numerical solutions to the system of first order conditions constitute a Nash equilibrium.

7 One network and cash users

Our analysis so far can be used to allow for cash as an alternative payment mode. We begin by assuming that consumers either use a credit card or cash. (Then, we allow for two competing networks and cash.) Preferences about these two modes of payments are modeled using the Hotelling model, as in section 6, where network 2 assumes the role of cash. Hence, $z_1 = \frac{1-r}{1-i}$ and $z_2 = \frac{1}{1-i}$. We assume that consumers located in $[k, 1]$ do not have access to credit and hence must use cash, with $k > \frac{1}{2}$ meaning that more than 50% of consumers have access to credit. The 'mobile' consumers in terms of the mode of payment are in $[0, k]$. The network chooses the interchange fee i and the reward r, so essentially it chooses z_1 and z_2, to maximize its profit $\pi_1(z_1, z_2)$.

Moreover, we assume that the product market demand is linear, so we follow the analysis of section 6.1.2. We begin by setting $\gamma = 0$. The share of credit card users, using (6.1), is given by $\mu = \frac{1}{2} + \frac{c\cdot(z_2 - z_1)}{2t}$. The network profit function is

$$\pi_1(z_1, z_2) = \frac{(t - (z_1 - z_2)c)(z_1 - 1)(1 - cz_1)c}{2t}.$$

We have $\frac{\partial \pi_1}{\partial z_2} = \frac{c^2(z_1-1)(1-cz_1)}{2t} > 0$, given that $z_1 \in \left(1, \frac{1}{c}\right)$, implying that the network increases its profit by increasing z_2, or increasing i, holding z_1 fixed (which implies that the reward r is also increasing so that cash users are being converted into credit card users). This can continue until all mobile consumers use the credit card in equilibrium: $\mu = k \Rightarrow z_2(z_1) = z_1 + \frac{t}{c}(2k-1)$. Substituting $z_2(z_1)$ into $\pi_1(z_1, z_2)$ and maximizing $\pi_1(z_1, z_2(z_1))$ with respect to z_1 we obtain

$$z_1^* = \frac{1 + c + 4t - \sqrt{(1-c)^2 + 16k^2t^2}}{2c}$$

and so z_2 becomes

$$z_2^* = \frac{1 + c + (8k-2)t - \sqrt{(1-c)^2 + 16k^2t^2}}{2c}.$$

Using z_1^* and z_2^* from above and $z_1 = \frac{1-r}{1-i}$ and $z_2 = \frac{1}{1-i}$, the reward and interchange fee are uniquely determined

$$r^* = \frac{2t(2k-1)}{1 + c - 2t(1-4k) - \sqrt{(1-c)^2 + 16k^2t^2}} \text{ and } i^* = \frac{1 - c - 2t(1-4k) - \sqrt{(1-c)^2 + 16k^2t^2}}{1 + c - 2t(1-4k) - \sqrt{(1-c)^2 + 16k^2t^2}}.$$

It can be readily verified that $i^* > r^* > 0$. The credit card tax for cash users is z_2^* and for credit card users it is z_1^*. Because $r^* > 0$, $z_2^* > z_1^* > 1$, indicating that cash users pay a higher credit card tax. Credit card users are subsidized through rewards and pay a lower price than cash users. However, because $\gamma = 0$, the credit card tax burden is borne entirely by consumers regardless of the mode of payment. The proof of the following proposition is straightforward and is omitted.

Proposition 6 *As more consumers gain access to credit, i.e., higher k, the interchange fee i^* and the reward r^* increase. Moreover, both credit card taxes, z_1^* and z_2^*, increase.*

Hence, credit card users exert a negative externality on cash users, but also on themselves.

The model can simulate reality where the average reward is 1.30% and the average merchant discount is 2.25%, see Wang (2023). If we assume that $k = 0.8$ (20% of consumers does not have a credit card) and set $c = 0.9774$ and $t = 0.0217$, then $i^* = 0.0225$ and $r^* = 0.013$.

NOTE: When $\gamma = 0$ the elasticity effect is absent. When $\gamma > 0$, it may be the case that a higher k, lowers the credit card tax...

Interchange fee cap. Suppose i is capped at a level $\bar{i} < i^*$. Hence z_2 is also capped at a level below z_2^* and the network can only affect z_1 by varying the reward. The optimal reward when the interchange fee is capped at $\bar{i}$ is maximizing $\pi_1(z_1, \bar{z}_2)$ and is given by

$$\bar{r} = \frac{(c+t+1)\bar{i} + c - t - 1 + \sqrt{(1-\bar{i})^2 t^2 + ((1-\bar{i}^2)c - (1-\bar{i})^2)t + (\bar{i}^2 - \bar{i} + 1)c^2 + (3\bar{i} - \bar{i}^2 - 2)c + (1-\bar{i})^2}}{3c}.$$

It can be verified that $\bar{r}$ is monotonic in $\bar{i}$, hence when the interchange fee is capped, the network lowers the reward. While the tax cash users are paying, z_2, is clearly decreasing, the effect on the credit card tax credit card users are paying, z_1, is ambiguous since both r and i are decreasing.

The optimal credit card tax is z_1 is

$$\bar{z}_1 = \frac{1 + c\bar{z}_2 + c + t - \sqrt{(\bar{z}_2^2 - \bar{z}_2 + 1)c^2 + ((2t-1)\bar{z}_2 - 1 - t)c + t^2 - t + 1}}{3c}.$$

It can be verified that $\bar{z}_1$ is monotonic in z_2. This suggests that when i decreases, the network will respond by decreasing its tax as well. This is achieved by not lowering the reward as fast as the interchange fee. Since both taxes are decreasing as the cap on i is becoming tighter, all consumers are becoming better off.

8 Conclusion

We develop a model to study payment networks that differs in a number of assumptions from classic two-sided models that have been introduced in the literature. One of our main goals is to be more flexible in modeling the product demand and competition amongst merchants, while also allowing for ad valorem pricing. All these very important and realistic elements have been missing from most of the two-sided models that study payment systems. We know from the taxation literature that ad valorem taxation yields different predictions than specific taxation. There is no reason why this is not the case in payment networks. Hence, for sound antitrust recommendations these elements ought to be incorporated in a model.

We construct a model that initially considers a monopoly payment network which later extends to a duopoly framework. This network determines the interchange fees that merchants incur and the rewards (or fees) allocated to consumers. Subsequently, n merchants engage in competition within the product market, offering a homogeneous good and operating under a 'general' demand function. The focal point of each network is to maximize its total profit. The 'aggregate' credit card tax is pivotal for determining prices, profits, and welfare–not the specific allocation between the interchange fee and the reward. The rationale is that both the interchange fee and the reward influence the consumer's final price, which in turn affects all equilibrium variables. Drawing parallels from the taxation literature, factors such as demand elasticity, demand curvature, and the intensity of product market competition play crucial roles. They are instrumental in ascertaining the extent of credit card tax pass-through, the incidence of credit card tax, and the overall welfare impact.

We are interested in the effects of competition both amongst networks and amongst merchants. Under a monopoly network, as competition in the product market intensifies, the credit card tax and the fraction of the tax borne by consumers can increase or decrease, depending on the the slope of the demand elasticity with respect to aggregate output. At the same time, network and consumers become better off, while merchants become worse off. (These results continue to hold when there are two competing networks in the market).

Then, we assume that a second network enters. If networks do not compete to attract users, then entry results in higher equilibrium credit card taxes if and only if demand becomes more elastic as aggregate output decreases. This is due to the elasticity effect. A higher network tax lowers aggregate output, and if demand becomes more elastic (subconvex demand) the market price decreases and so does the network revenue. With two networks, own network tax has a smaller impact on aggregate output, so the negative elasticity effect weakens, making the networks less reluctant to increase their own tax. As a result, entry increases the credit card tax while reducing profits and welfare. The reverse is true when market demand becomes less elastic as aggregate output decreases (superconvex demand).

We then allow networks to compete for users, assuming networks are horizontally differentiated. Each network in order to attract users lowers its tax, and consequently the price users pay to make purchases with the network's credit card; the competition effect. When networks are not much differentiated, the competition effect dominates the elasticity effect, and entry of a second network lowers the equilibrium tax and increases welfare, if demand becomes more elastic when aggregate output decreases (which is true when, for example, demand is linear). When, however, the two networks are sufficiently differentiated, entry results in higher credit card taxes, higher product prices and lower welfare.

In our analysis, we categorize demand functions based on how demand elasticity varies with aggregate output, which corresponds to subconvexity or superconvexity. This property is crucial as it influences competition and credit card tax pass-through at both the merchant and network levels. While prior literature has explored the impact of these properties on competition and pass-through in the product market, our paper is the first to highlight the complexities and novelties that emerge in two-sided or vertically related markets.

There are a number of novel policy implications that follow from our model. First, the model illustrates when and why having additional payment networks can increase or decrease welfare. Two networks competing for essentially the same transactions will be pro-competitive, relative to monopoly, as long as the networks are competing intensely for card users, due to, for instance, lack of differentiation. Otherwise, merchants and consumers can be made worse off by the presence of two networks that are not competing aggressively.

Second, merchants typically worry about the potential shift of profits away from their businesses towards networks and consumers, especially in less competitive retail markets. When the slope of the elasticity of the product demand with respect to aggregate output is positive, a more competitive retail environment tends to result in a higher credit card tax, primarily borne by consumers. Conversely, a negative slope leads to a lower tax, with merchants bearing a higher fraction of the cost. Policymakers should consider that increased competition among merchants could lead to higher credit card taxes, potentially diminishing some of the advantages that come with intensified competition in the product market.

Finally, it is crucial to note that within our model, implementing a cap on interchange fees alone may be partially undermined by the networks' downward adjustment of consumer rewards.[20] Nevertheless, the overall credit card taxes will decrease and all consumers will become better off.

For the sake of simplicity and analytical convenience, we have operated under the assumption that both issuing and acquiring banks lack market power, which is concentrated at the level of the network. A potential direction for future research could involve attributing a degree of market power to issuing banks.

[20]This phenomenon aligns with the 'waterbed effect' observed in regulatory theory, which suggests that regulation of only a subset of prices can lead to compensatory adjustments in unregulated prices. This concept is supported by the findings in the studies by Genakos and Valletti (2011) and Hong et al. (2023), which provide empirical evidence of such regulatory outcomes.

A Appendix: Proofs of Lemmas and Propositions and Some Further Results

A.1 Proof of Proposition 1

We differentiate (4.6) with respect to z. The resulting first order condition (omitting arguments) is

$$\frac{d\pi_1}{dz} = \frac{\left(\left(\gamma X\varepsilon' + \gamma\varepsilon + \varepsilon^2\right)(z-1)\frac{dX}{dz} + X\varepsilon(\varepsilon+\gamma)\right)c}{(\varepsilon+\gamma)^2} = 0. \tag{A.1}$$

The first order condition (4.1) can be written as follows

$$\frac{1}{z}\left(P_X X\gamma + P\right) - c = 0.$$

We invoke the implicit function theorem to derive the effect of z on aggregate output X

$$\frac{dX}{dz} = \frac{P_X X\gamma + P}{z(P_X(1+\gamma) + P_{XX}X\gamma)} = \frac{X(\varepsilon+\gamma)}{z(1+\gamma(1-E))}. \tag{A.2}$$

Thus, the elasticity of aggregate output X with respect to the tax z is

$$e \equiv \frac{dX}{dz}\frac{z}{X} = \frac{\varepsilon+\gamma}{1+\gamma(1-E)} < 0, \tag{A.3}$$

a function of product demand elasticity, the elasticity of the product demand slope and product market competition. We substitute (A.2) into (A.1) and after simple algebra we obtain

$$z = \frac{\gamma X\varepsilon' + \varepsilon(\varepsilon+\gamma)}{\gamma X\varepsilon' + \varepsilon(1+\varepsilon+\gamma(2-E))}.$$

References

Auerbach, Alan J and James R Hines Jr (2002), "Taxation and economic efficiency." In *Handbook of Public Economics*, volume 3, 1347–1421, Elsevier.

Baxter, William F (1983), "Bank interchange of transactional paper: Legal and economic perspectives." *The Journal of Law and Economics*, 26, 541–588.

Bedre-Defolie, Özlem and Emilio Calvano (2013), "Pricing payment cards." *American Economic Journal: Microeconomics*, 5, 206–231.

Bisceglia, Michele and Jean Tirole (2023), "Fair gatekeeping in digital ecosystems."

Bresnahan, Timothy F (1981), “Duopoly models with consistent conjectures.” *American Economic Review*, 71, 934–945.

Bulow, Jeremy and Paul Klemperer (2012), “Regulated prices, rent seeking, and consumer surplus.” *Journal of Political Economy*, 120, 160–186.

Delipalla, Sofia and Michael Keen (1992), “The comparison between ad valorem and specific taxation under imperfect competition.” *Journal of Public Economics*, 49, 351–367.

Edelman, Benjamin and Julian Wright (2015), “Price coherence and excessive intermediation.” *The Quarterly Journal of Economics*, 130, 1283–1328.

Genakos, Christos and Tommaso Valletti (2011), “Testing the âwaterbedâ effect in mobile telephony.” *Journal of the European Economic Association*, 9, 1114–1142.

Guthrie, Graeme and Julian Wright (2007), “Competing payment schemes.” *The Journal of Industrial Economics*, 55, 37–67.

Hayashi, Fumiko et al. (2009), “Do US consumers really benefit from payment card rewards?” *Economic Review*, 94, 37–63.

Hong, Suting, Robert M Hunt, and Konstantinos Serfes (2023), “Dynamic pricing of credit cards and the effects of regulation.” *Journal of Financial Services Research*, 64, 81–131.

Hotelling, Harold (1929), “Stability in competition.” *Economic Journal*, 39, 41–57.

Jeon, Doh-Shin and Patrick Rey (2022), “Platform competition and app development.” *Unpublished manuscript, Toulouse Sch. Econ., Toulouse.*

Jullien, Bruno, Alessandro Pavan, and Marc Rysman (2021), “Two-sided markets, pricing, and network effects.” In *Handbook of Industrial Organization*, volume 4, 485–592, Elsevier.

Mrázová, Monika and J Peter Neary (2017), “Not so demanding: Demand structure and firm behavior.” *American Economic Review*, 107, 3835–3874.

Mrázová, Monika and J Peter Neary (2019), “Selection effects with heterogeneous firms.” *Journal of the European Economic Association*, 17, 1294–1334.

Rochet, Jean-Charles and Jean Tirole (2002), “Cooperation among competitors: Some economics of payment card associations.” *Rand Journal of Economics*, 549–570.

Rochet, Jean-Charles and Jean Tirole (2006), “Two-sided markets: a progress report.” *The RAND Journal of Economics*, 37, 645–667.

Rochet, Jean-Charles and Jean Tirole (2011), “Must-take cards: Merchant discounts and avoided costs.” *Journal of the European Economic Association*, 9, 462–495.

Rysman, Marc (2009), "The economics of two-sided markets." *Journal of Economic Perspectives*, 23, 125–143.

Schmalensee, Richard (2002), "Payment systems and interchange fees." *The Journal of Industrial Economics*, 50, 103–122.

Seade, Jesus (1980), "On the effects of entry." *Econometrica*, 479–489.

Shy, Oz and Zhu Wang (2011), "Why do payment card networks charge proportional fees?" *American Economic Review*, 101, 1575–1590.

Wang, Lulu (2023), "Regulating competing payment networks." *URL: https://luluywang. github. io/PaperRepository/payment_jmp. pdf (last accessed on 10 November 2023)*.

Wang, Zhu and Julian Wright (2017), "Ad valorem platform fees, indirect taxes, and efficient price discrimination." *The RAND Journal of Economics*, 48, 467–484.

Wang, Zhu and Julian Wright (2018), "Should platforms be allowed to charge ad valorem fees?" *The Journal of Industrial Economics*, 66, 739–760.

Weyl, E Glen and Michal Fabinger (2013), "Pass-through as an economic tool: Principles of incidence under imperfect competition." *Journal of Political Economy*, 121, 528–583.

Wright, Julian (2004), "The determinants of optimal interchange fees in payment systems." *The Journal of Industrial Economics*, 52, 1–26.