Proportional vs. Unit Fees: Incentive and Welfare

Dingwei Gu, Zhiyong Yao, and Wen Zhou*

January 28, 2019

Abstract

Oz Shy and Zhu Wang (2011) show that when a payment card network switches from unit fees to proportional fees, the network and consumers both benefit. Their results are established on a monopoly network facing a constant-elastic demand. This paper demonstrates that the two parties' interests are not always fully aligned if the demand is more general or multiple networks compete. On a monopoly network, proportional fees always benefit the network regardless of the demand function, but the welfare consequence is ambiguous. In particular, welfare improves if the demand is sub-convex; when that condition fails, social welfare is more likely to improve if the network's marginal cost is higher, the merchants' marginal costs are lower, or merchant competition is stronger. On competitive networks, the results are swapped: Welfare always improves, but networks may lose. Beyond the payment card industry, our findings strengthen or clarify existing results in public finance, technology licensing, and retailing. Common to all these activities is a "discount effect" of proportional fees, which mitigates the double markup problem and encourages output expansion.

Keywords: proportional fee, unit fee, the discount effect, agency model, wholesale model, ad valorem tax, specific tax

JEL Codes: D2, D4, L1, L4

1 Introduction

Payment cards have become ubiquitous and indispensable in modern economies. In 2017, U.S. consumers made 123.5 billion transactions with payment cards, amounting to a total value of \$6.48 trillion (Federal Reserve 2018). When a cardholder makes a purchase, the payment to the merchant is settled by a card network, which also sets and collects an interchange fee on behalf of the card issuing bank. The total card fees that U.S. merchants paid in 2016 added up to \$88.39 billion, averaging to \$1.49 for every \$100 card transaction (The Nilson Report 2017).

*Gu: HKUST and Fudan University, Dingwei.gu@gmail.com; Yao: Fudan University, yzy@fudan.edu.cn; Zhou: University of Hong Kong, wzhou@hku.hk.

For a long time, merchants have complained that the high fees had inflated the transaction costs. A particular concern is that the fees are usually proportional to the transaction value (henceforth proportional fees), even though the cost of executing each transaction does not seem to vary much with the price (Shy and Wang 2011). Partly in response to these complaints, the Federal Reserve Board imposed a cap in 2011 that effectively halved the interchange fees for most domestic debit card transactions. The legislation, however, brought mixed and unintended consequences, with banks cancelling free current accounts and encouraging cardholders to use more costly credit cards to recover the loss (Mark D. Manuszak and Krzysztof Wozniak 2017). Both the merchants and card-issuing banks have opposed this regulation, and the Financial CHOICE Act of 2017 would repeal the so-called Durbin Amendment (Darryl E. Getter 2017).[1]

Shy and Wang (2011) study the welfare consequences of proportional fees vis-à-vis a fixed, per-unit transaction fee (henceforth unit fees). They demonstrate that when a monopoly card network switches from unit fees to proportional fees, the network is better off, the merchants are worse off, while social welfare and consumer surplus always improve. These results are neat, strong and unambiguous. They explain why a network wants to adopt proportional fees and why merchants complain, but caution against regulation of the fee scheme because the network's private incentive is fully aligned with the social incentive. Shy and Wang (2011) established their findings with a constant-elastic demand on a monopoly network. A natural question is: Do they continue to hold for more general demands and on competing networks?

There are good reasons for Shy and Wang (2011) to focus on the fee schemes rather than fee levels. First, fee levels have already been studied extensively.[2] Second, in terms of policy implementation, enforcing a particular fee type is much easier than targeting a socially optimal fee level, as the latter is more demanding in information collection and subject to strategic manipulation. Third, proportional and unit fees are the two most commonly used arrangements in a wide range of economic activities well beyond payment card networks. Some are causing controversies among practitioners, scholars and regulators. For example, the government levies either an ad-valorem tax (proportional to the value) or a specific tax (a fixed amount per unit of the good or transaction). In the licensing of technologies, the royalty is commonly calculated based on either the value of the whole product or the component specific to the technology, which are equivalent to ad valorem and per-unit royalty rates respectively (Gerard Llobet and Jorge Padilla 2016).[3] A recent trend is that the court seems

[1]There are many other lawsuits concerning payment cards. In the Payment Card Interchange Fee and Merchant Discount Antitrust Litigation, merchants and trade associations alleged that Visa, MasterCard, and major credit card issuers had engaged in price fixing and other allegedly anti-competitive practices. In December 2013, a U.S. District Court Judge approved a settlement of $7.25 billion, which lowered interchange fees in return for a provision that bar future lawsuits over the same issue. Many merchants disputed the provision. In June 2016, the United States Court of Appeals for the Second Circuit overturned the settlement.

[2]For example, Willian Baxter (1983); Ozlem Bedre-Defolie and Emilio Calvano (2013); Jean-Charles Rochet and Jean Tirole (2002); Julian Wright (2004; 2013); Zhu Wang (2010), and many others.

[3]One more reason to focus on the fee type rather than fee level is that only the fee type is contractable

to favor the per-unit regime, which worries some scholars. In retailing and supply chains, agency and wholesale are the two most common business models (Andrei Hagiu and Julian Wright 2015). In the agency model, a retailer sets the commission rate, and suppliers set retail prices; this is similar to proportional fees where the retailer acts like a card network and suppliers act like merchants. In the wholesale model, suppliers set the unit prices at which products or contents are sold to the retailer, who in turn sets the final retail prices. This corresponds to unit fees with a twist that the two parties' roles are swapped.

In recent years, the agency model has become increasingly popular among platforms for the sale of digital contents such as music, electronic books, hotel and airline booking, and cell phone applications, and has been at the center of some high profile antitrust lawsuits (Germain Gaudin and Alexander White 2014). The Office of Fair Trade in the UK had investigated some major firms in the hotel online booking industry for tax issues and anticompetitive conducts, and a key issues is whether an online booking company should be treated as an agent or a distributor.[4] Another example is the 2012 ebook price-fixing case involving Apple Inc. Prior to the introduction of Apple's iPad in 2010, Amazon was the dominant player in the ebook market with its Kindle, which used the wholesale model. Apple entered the competition by signing agency contracts with five major publishers, who in turn forced Amazon to switch to the agency model. Apple was eventually convicted of price-fixing, and its appeals to the circuit and supreme courts both failed. In a more recent case in November 2018, again involving Apple, the Supreme Court of the United States is considering whether consumers have the right to sue Apple Inc. for anticompetitive conduct. At issue is Apple's practice of charging 30 per cent commission to any software sold in Apple's exclusive Apple Store.

In this research, we revisit the private and social desirability of proportional fees by expanding Shy and Wang's (2011) framework to first a more general demand, and then to competing networks. The major findings are the following. A monopoly network always benefits from switching to proportional fees regardless of the demand function, but the welfare consequence is ambiguous. When multiple networks compete, the results are swapped: Welfare always improves, but the networks may be better or worse off. For a monopoly network, therefore, the regulatory concern is an over-utilization of proportional fees, as a social planner would have preferred unit fees in some situations. For competing networks, however, an under-utilization of proportional fees needs not worry policy makers, as each network's individual incentive can be strong even if every network earns lower profits in equilibrium. That is, competition among networks may cause an across-the-board adoption of proportional fees to the benefits of consumers and social welfare.

(Llobet and Padilla 2016).

[4] "An agreement between a principal and an agent/distributor has received different treatment in competition law. The EU Guidelines on Vertical Restraints state that the agent is a separate undertaking from the principal and exclusive agency provisions will in general not lead to anti-competitive effects. This is markedly different from distribution agreements, which have received intense antitrust scrutiny." (Ioannis Kokkoris 2013)

The intuition can be understood as follows, starting with a monopoly network. From the network's viewpoint, merchants produce too little due to merchants' own markup. This is true for both unit and proportional fees. Compared to unit fees, however, proportional fees encourage each merchant to produce a little more at any given (absolute) fee level, as the loss from price drop (due to output expansion) will be shared with the network. Such a "discount effect" mitigates the underproduction and therefore raises the network's profit. The result is very robust because the discount effect holds for any given fee paid to the network and, therefore, is independent of the particular equilibrium characterization. For a monopoly network, the superior profitability of proportional fees is unconditional.

All the remaining results depend on equilibrium characterization. Starting from the equilibrium under unit fees, the discount effect implies that a monopoly network can use proportional fees to raise the output without changing the fee level, or raise the fee without changing the output level. In fact it may even reduce the output if the gain in the fees outweighs the loss in output. We find that a sub-convex demand (i.e., the demand becomes less elastic when price is lower, according to Monika Mrázová and J. Peter Neary, 2017) is sufficient for proportional fees to raise social welfare. When the condition fails so that the demand is super-convex, social welfare is more likely to improve if the network's marginal cost is higher, the merchants' marginal costs are lower, or merchant competition is stronger. It turns out that the constant-elastic demand used by Shy and Wang (2011) is a borderline case between sub- and super-convex demands.

When networks compete, the discount effect continues to encourage output expansion under proportional fees, but each network's individual incentive to expand is further strengthened, as part of a network's benefits comes from stealing its rival networks' businesses. If all networks simultaneously adopt proportional fees, the equilibrium output always rises, leading to higher social welfare unconditionally. However, a network's expansion hurts competing networks, and all networks may be worse off in equilibrium.

The idea that proportional fees can mitigate double markup is not new, but our analysis clarifies that the effect is unconditional (as long as merchants have market power) and therefore much more robust than other results, which involve particular equilibrium choices. The explicit expression of the discount effect enables us to completely characterize the equilibrium comparison between the two fee schemes, going beyond the literature's common focus on demand properties. For network competition, we develop an analytical tool within the standard oligopoly framework, which explicitly models and endogenizes strategic interactions both within and across inter-related oligopolies. The methodology seems useful more broadly for any study that involves multiple oligopolies.

Our findings shed new light to other economic activities, as the mathematical models are almost identical. Taxation is usually thought to be carried out by a monopoly government who does not seek to maximize tax revenue. Then the relevant finding in our paper is the discount effect as it, too, is independent of equilibrium characterizations and hence the monopolist's objective. The major finding in public finance, starting with D. B. Suits and R.

A. Musgrave (1953) and followed by Robert L. Bishop (1968), Sofia Delipalla and Michael Keen (1992), Simon Anderson et al. (2001), is that ad valorem tax Pareto dominates specific tax. This is an immediate corollary to our discount effect, which is stronger because we point out that the desirability of ad valorem tax holds even if the government incurs positive costs in tax collection.

Technology is licensed by a monopoly patent holder typically assumed to incur zero cost, who does seek to maximize its profits. Then all of our findings about monopoly network apply directly. The licensing literature has rarely studied ad valorem royalty even though it is much more common than unit royalty in practice.[5] Recent court rulings, especially those involving standard-setting organizations (SSOs) such as the Institute of Electrical and Electronics Engineers (IEEE), tend to favor the per-unit regime. This is worrying according to Llobet and Padilla (2016), who demonstrate that ad valorem royalty welfare dominates unit royalty if the demand is sub-convex. Our research shows that as long as the demand is sub-convex, the benign welfare effect of ad valorem royalty continues to hold even when technology is transferred at positive costs and to multiple produces. In addition, we point out that if the demand is super-convex, the conclusion may be reversed depending on all the remaining parameters.

In retailing, Gaudin and White's (2014) study of the ebook industry finds that the agency model raises welfare if and only if demand is sub-convex (in their non-essential case). That is reached in bilateral monopoly with zero retailing cost. We demonstrate that welfare may still rise for super-convex demands, therefore sub-convexity is only sufficient, but not necessary, for welfare improvement when retailing cost is positive. Justin Johnson (2017) finds that in both bilateral monopoly and bilateral oligopoly, the agency model improves welfare if the demand is sub-convex. For monopoly, our research establishes what happens when the demand condition fails. For bilateral oligopoly, Johnson gives a sufficient condition that is independent of network competition, but our analysis shows that the welfare improvement is unconditional so long as there are at least two networks competing. This is because Johnson adopts a constant conduct-parameter approach, which assumes identical conduct under both fee schemes and precludes horizontal strategic interaction. By contrast, our standard oligopoly approach fully endogenizes firms' behavior, under which conduct parameters are necessarily different between the two schemes, and horizontal strategic interaction becomes the driving force of the unconditional welfare consequence. A more detailed comparison with these two retailing papers is carried out later.

In addition to monopoly and competing networks, the paper also analyzes mixed fees, Bertrand competition, and two-sided markets. It provides discussions concerning robustness, outside options, tax competition and system competition, policy implications, and alternative ways to mitigate double markup. We also highlight a peculiar feature of proportional fees that may dampen merchants' innovation incentives and, hence, bring additional, negative

[5] Alain Bousquet et al. (1998) mentioned that out of 225 licensing contracts that specified royalties, only nine used per-unit royalties, and all the remaining adopted the ad valorem scheme.

welfare impacts in a more dynamic setting.

2 Monopoly Network

The monopoly model is based on Shy and Wang (2011) but with a more general demand, and will be further extended later to allow for network competition. Suppose that $m < \infty$ merchants compete à la Cournot by selling a homogeneous good to consumers. All merchants have the same constant marginal cost of production, κ_M. The inverse demand for the good, $p(Q)$, satisfies $p(Q) > 0$ and $p'(Q) < 0$, where Q is the total sales quantity of all the merchants. Assume that the elasticity of demand $\varepsilon \equiv -\frac{p(Q)}{p'(Q)Q} \geq 1$ so that marginal revenue is non-negative; and the concavity of demand $\rho \equiv \frac{p''(Q)Q}{p'(Q)} > -2$ so that marginal revenue decreases with output.[6]

Consumers use payment cards to buy the good, and all the cards belong to a monopoly card network, which incurs a constant marginal cost of κ_N for each unit of goods transacted. For card services, consumers have to pay the network a transaction fee. The fee can take two forms: it can be proportional to the transaction price, hereby referred to as a proportional fee, with the coefficient being denoted as τ_C; or it can be a fixed amount for each transacted unit, thus referred to as a unit fee and denoted as t_C. Similarly, the network also charges the merchants a fee, either proportional (τ_M) or unit (t_M).

Assume that the network uses the same type of fees for both consumers and merchants (mixed fees will be analyzed later). For every unit of transaction, consumers pay $p(Q)$ from their pockets. Under unit fees, the network receives t_C from consumers and t_M from merchants, leading to an average revenue of $t_C + t_M$, and the merchants receive the remaining $p(Q) - (t_C + t_M)$. Under proportional fees, the transaction price is $\frac{p(Q)}{1+\tau_C}$. The network receives $\frac{\tau_C}{1+\tau_C}p(Q)$ from consumers and $\frac{\tau_M}{1+\tau_C}p(Q)$ from merchants, so its average revenue is $\frac{\tau_C+\tau_M}{1+\tau_C}p(Q)$, and the merchants receive the remaining $p(Q) - \frac{\tau_C+\tau_M}{1+\tau_C}p(Q) = \frac{1-\tau_M}{1+\tau_C}p(Q)$.

The game proceeds as follows. For any given fee scheme, the network chooses (τ_C, τ_M) under proportional fees, or (t_C, t_M) under unit fees. Taking these two fees as given, each merchant chooses its output in Cournot competition, and consumers purchase the good. Finally the network collects fees from merchants and consumers based on the transaction price and quantity.

[6]Both ε and ρ are local properties depending on the output Q and do not have to be constant. The two properties are assumed only in the neighborhood of any relevant equilibrium.

2.1 Unit fees

When the network charges unit fees, the profit of merchant i is $\pi_i^U = [p(Q) - t_C - t_M - \kappa_M]q_i$ if it produces q_i. The merchant first-order condition (FOC) leads to:[7]

$$\frac{1}{m}p'(Q)Q + p(Q) = \kappa_M + t_M + t_C. \tag{1}$$

As a result, the network's average revenue is:[8]

$$\begin{aligned} D^U(Q) &= t_C + t_M \\ &= p(Q) - \kappa_M + p'(Q)\frac{Q}{m}. \end{aligned} \tag{2}$$

Given the average revenue, the network's profit is:

$$\begin{aligned} \Pi_N^U(Q) &= (t_C + t_M)Q - \kappa_N Q \\ &= [p(Q) - \kappa_M - \kappa_N]\, Q + \frac{1}{m}p'(Q)Q^2. \end{aligned} \tag{3}$$

Although the network has two instruments (i.e., t_M and t_C), only $t_C + t_M$ matters for its profit. Because choosing (t_C, t_M) is equivalent to choosing Q, we will treat Q as the network's choice variable, and express its average revenue, $t_C + t_M$, as a dependent variable. Assuming $\frac{\partial^2 \Pi_N^U(Q)}{\partial Q^2} < 0$,[9] the network's FOC:

$$\kappa_M + \kappa_N = p'(Q)Q + p(Q) + \frac{1}{m}\left[p''(Q)Q + 2p'(Q)\right] Q, \tag{4}$$

will determine a unique optimal output, Q^U, based on which other endogenous variables can be calculated.

2.2 Proportional fees

When the network charges proportional fees, the profit of merchant i is $\pi_i^P = \left[\frac{1-\tau_M}{1+\tau_C}p(Q) - \kappa_M\right] q_i$. The merchant FOC is:

$$\frac{1-\tau_M}{1+\tau_C}\left[\frac{1}{m}p'(Q)Q + p(Q)\right] = \kappa_M. \tag{5}$$

As a result, the network's average revenue under proportional fees is:

$$\begin{aligned} D^P(Q) &= \frac{\tau_C + \tau_M}{1+\tau_C}p(Q) = \left(1 - \frac{1-\tau_M}{1+\tau_C}\right)p(Q) \\ &= p(Q) - \kappa_M + \frac{\kappa_M}{p'(Q)\frac{Q}{m} + p(Q)}p'(Q)\frac{Q}{m}. \end{aligned} \tag{6}$$

[7] $\varepsilon \geq 1$ and $\rho > -2$ are sufficient to guarantee merchants' profit functions are well behaved, i.e., the equilibirum is unique and stable.

[8] $D^U(Q)$ links the network's per-unit revenue to the total output, Q, and can therefore be regarded as its (inverse) demand function. However, since the revenue is collected from both consumers and merchants, it is unclear who is demanding the service. For this reason, we will use the term *average revenue* rather than inverse demand.

[9] This condition relates to the third derivative of the final demand function ($p'''(Q)$) and is very mild.

Again, there exists price neutrality such that there is a unique one-to-one mapping between Q and $\frac{\tau_C+\tau_M}{1+\tau_C}$. The network's profit is:

$$\begin{aligned}\Pi_N^P(Q) &= \left[D^P(Q)-\kappa_N\right]Q \\ &= \left[p(Q)-\kappa_M-\kappa_N\right]Q+\frac{\kappa_M}{p'(Q)Q+mp(Q)}p'(Q)Q^2 \end{aligned} \tag{7}$$

Treating Q as the choice variable, we have the network's FOC:

$$\kappa_M+\kappa_N = p'(Q)Q+p(Q)-\frac{\kappa_M\left[p''(Q)Q+(m+1)p'(Q)\right]p'(Q)Q^2}{\left[p'(Q)Q+mp(Q)\right]^2}+\frac{\kappa_M\left[p''(Q)Q+2p'(Q)\right]Q}{p'(Q)Q+mp(Q)}. \tag{8}$$

A unique optimal output, Q^P, can be determined from condition (8) if $\frac{\partial^2\Pi_N^P(Q)}{\partial Q^2}<0$, which is assumed.

2.3 Network's average revenue: the discount effect

Under unit fees, the term $-p'(Q)\frac{Q}{m}>0$ in equation (2) captures the merchant's markup, and hence the magnitude of distortion from the network's viewpoint. Under proportional fees, let $\beta=\frac{\tau_M+\tau_C}{1+\tau_C}$ denote the share of consumer payment that goes to the network, hereafter referred to as the network's *commission rate* (under proportional fees). For any given Q, merchant FOC (5) implies that:

$$\beta(Q)=1-\frac{\kappa_M}{p'(Q)\frac{Q}{m}+p(Q)}.$$

Then, network average revenue (6) can be re-written as

$$D^P(Q)=p(Q)-\kappa_M+[1-\beta(Q)]p'(Q)\frac{Q}{m}. \tag{9}$$

The commission rate, $\beta(Q)$, must be between 0 and 1 even though the individual τ_M or τ_C can be negative (i.e., a subsidy to one side). Under proportional fees, then, the merchant markup is $-[1-\beta(Q)]p'(Q)\frac{Q}{m}>0$.

For any given Q, the merchant markup under proportional fees is only a fraction of that under unit fees, which immediately implies that the network's average revenue is higher under proportional fees:[10]

$$D^P(Q)-D^U(Q)=\underbrace{-\beta(Q)p'(Q)\frac{Q}{m}}_{\text{Discount Effect }(+)}>0.$$

To understand why $D^P(Q)>D^U(Q)$, note that each merchant's output choice balances its marginal revenue (from the final demand) with its marginal cost, which includes the fees paid to the network. Under unit fees, if a merchant sells more, a one dollar drop in the final sales price is fully shouldered by the merchant. Under proportional fees, however, a one-dollar

[10] $D^P(Q)=D^U(Q)$ only when $Q=0$ or $p=0$ (see Figure 1), which never happens in equilibrium. For this reason, subsequent discussions ignore the possibility of $D^P(Q)=D^U(Q)$.

drop of the price will result in only a $1-\beta(Q)<1$ dollar loss for the merchant. Fixing the output level, therefore, proportional fees reduce a merchant's marginal cost without affecting its marginal revenue. The saved marginal cost, which equals the commission rate, $\beta(Q)$, multiplied by the merchant's markup, $-p'(Q)\frac{Q}{m}$, can then be taken away by the network without reducing the total output.

Therefore, proportional fees give the network a larger average revenue for any given total output or, alternatively and equivalently, a larger total output for any given average revenue (see Figure 1). This property of proportional fees, referred to as *the discount effect,* has a few features we would like to highlight. First, the discount effect is valid for any arbitrary Q, and is therefore independent of particular characterization of the equilibrium. Second, it is not due to competition among merchants, and therefore exists even when there is only one merchant, or when the downstream competition is Bertrand (with differentiated products), as shown in section 4.1. Third, the driving force is that a smaller markup can generate the same output under proportional fees, and thus induces a smaller distortion. As a result, the discount effect exists if and only if individual merchants have market power, i.e., $-p'(Q)\frac{Q}{m}\neq 0$, which holds given the assumption of $m\neq\infty$. If there is perfect competition among merchants (i.e., $\lim\limits_{m\to\infty}-p'(Q)\frac{Q}{m}=0$), the effect disappears completely, and the two fee schemes will lead to exactly the same outcome.

Proposition 1 *For any given Q, the network earns a strictly larger revenue under proportional fees than under unit fees.*

Figure 1 shows the network's average revenue in the two fee schemes, $D^P(Q)$ and $D^U(Q)$. They are constructed from the consumer final demand, $p(Q)$, which is moved down first by κ_M to take account of merchants' marginal production cost, and then further down due to merchants' market power.

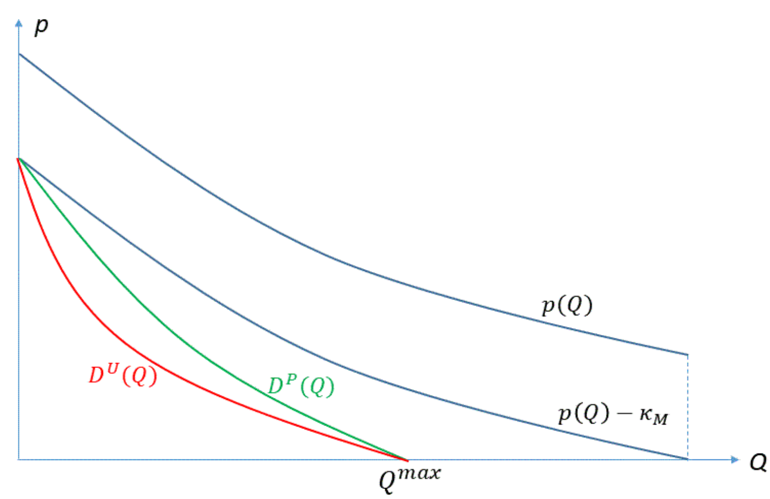

Figure 1: The network's average revenue under unit and proportional fees

2.4 The network's profit

Given $D^P(Q) > D^U(Q)$, it is straightforward to establish:

$$\begin{aligned}\Pi_N^P(Q^P) &\geq \Pi_N^P(Q^U) \text{ (as } Q^P \text{ is chosen optimally to maximize } \Pi_N^P(Q)) \\ &> \Pi_N^U(Q^U) \text{ (as } D^P(Q) > D^U(Q) \text{ due to the discount effect)}\end{aligned}$$

Proposition 2 *The network earns a strictly greater profit in equilibrium under proportional fees than under unit fees.*

The proposition is a direct consequence of the discount effect: proportional fees do better for any given Q and, by revealed preference, must also do better in equilibrium, i.e., when Q is allowed to differ between the two schemes. The result is established without any need to analyze the particular equilibrium choices, and is therefore unconditional as long as merchant competition is imperfect.

2.5 Consumer surplus and social welfare

Since the merchants have identical cost and produce a homogeneous product, consumer surplus and social welfare can be both evaluated by a single variable, the equilibrium total output, Q. A larger Q leads to greater consumer surplus and social welfare.

To compare Q^U and Q^P, rewrite network FOC under proportional fees, (8), as:

$$\begin{aligned}\kappa_M + \kappa_N &= p'(Q)Q + p(Q) + \frac{1}{m}\left[p''(Q)Q + 2p'(Q)\right]Q \\ &+ \left[\underbrace{\beta(Q)\frac{p''(Q)Q + 2p'(Q)}{p'(Q)}}_{\text{Expansion Effect } (+)} + \underbrace{\beta'(Q)Q}_{\text{Concession Effect } (-)}\right] \times \left[-p'(Q)\frac{Q}{m}\right]\end{aligned} \tag{10}$$

Now compare (10) with the corresponding FOC under unit fees, (4). At any given Q, the network's average revenue is higher under proportional fees due to the discount effect, but its marginal revenue can be higher or lower depending on the tradeoff between two effects, as shown in (10). Fixing the commission rate, $\beta(Q)$, proportional fees give the network a higher total revenue and therefore tends to increase its marginal revenue. This is the *expansion effect.* On the other hand, a larger Q is induced only through a larger share to the merchants, and consequently a smaller commission rate $\beta(Q)$ to the network. Thus the marginal revenue tends to decrease. This is the *concession effect.*

The two effects always move in opposite directions. To determine the net effect, we solve κ_M from (4) and substitute it into (10), and find that $Q^P > Q^U$ (i.e., the net effect is positive) if and only if at the equilibrium Q^U,

$$\kappa_N > -p(Q^U)\frac{\left[\varepsilon(\rho+1)+1\right](\rho+m+1)}{\varepsilon\left\{m\left[\varepsilon(\rho+2)+1\right]-1\right\}} \equiv \widehat{\kappa}_N, \tag{11}$$

where ε and ρ are demand elasticity and concavity. Notice that Q^U, ε and ρ are endogenous and therefore dependent on m, κ_N and κ_M. In addition, both $\rho + m + 1$ and $\varepsilon \{m [\varepsilon (\rho + 2) + 1] - 1\}$ are positive, but the sign of $\varepsilon (\rho + 1) + 1$ is ambiguous. According to Mrázová and Neary (2013, 2017), a demand function is sub-convex if $\varepsilon(\rho + 1) + 1 > 0$, constant-elastic if $\varepsilon(\rho + 1) + 1 = 0$, and super-convex if $\varepsilon(\rho + 1) + 1 < 0$.[11]

For sub-convex demand,

$$\varepsilon(\rho + 1) + 1 > 0 \Longrightarrow \widehat{\kappa}_N < 0 \Longrightarrow Q^P > Q^U.$$

Therefore, demand sub-convexity is sufficient for proportional fees to strictly improve welfare. To understand the intuition, recall that in order to expand the total output, the network has to lower the fees, waiting for the merchants to pass (part of) the concessions to final consumers. A sub-convex demand becomes less elastic when Q increases (Mrázová and Neary, 2017), which means the final price will become increasingly sensitive to changes in the merchants' effective marginal cost.[12] That is, a given increase in Q can be achieved through a smaller concession, so the net effect tends to be positive.

For constant-elastic demand,

$$\varepsilon(\rho + 1) + 1 = 0 \Longrightarrow \widehat{\kappa}_N = 0 \Longrightarrow \begin{cases} \text{if } \kappa_N > 0, \text{ then } Q^P > Q^U \\ \text{if } \kappa_N = 0, \text{ then } Q^P = Q^U \end{cases}$$

When the demand elasticity is constant, merchant markup maintains a constant ratio between the two schemes, which leads to a constant ratio of the network's average revenue for any given Q (i.e., $D^P(Q) = \frac{m\varepsilon}{m\varepsilon - 1} D^U(Q)$). As a result, the network's marginal revenue is always weakly larger under proportional fees than under unit fees, with equality if and only if $MR^P = MR^U = 0$, which leads to $Q^P = Q^U$ in equilibrium when $\kappa_N = 0$.

Finally, for super-convex demand,

$$\varepsilon(\rho + 1) + 1 < 0 \Longrightarrow \widehat{\kappa}_N > 0 \Longrightarrow \begin{cases} \text{if } \kappa_N = 0, \text{ then } Q^P < Q^U \\ \text{if } \kappa_N > 0, \text{ see below} \end{cases}$$

The second case is analyzed by comparing the network's marginal revenues from the two schemes (i.e., (4) and (8)):

$$MR^P(Q) - MR^U(Q) = \left\{ \underbrace{\beta(Q)\,(\rho + 2)}_{\text{Expansion Effect } (+)} + \underbrace{\left[-\left(1 - \beta(Q)\right) \frac{\rho + m + 1}{m\varepsilon - 1} \right]}_{\text{Concession Effect } (-)} \right\} \times \left[-p'(Q) \frac{Q}{m} \right]. \tag{12}$$

[11] In fact, the definition concerns how the elasticity changes with price. For example, a demand is sub-convex if and only if the demand becomes less elastic when the price is lower, sometimes also referred to as Marshall's Second Law of Demand.

[12] Roughly, $\frac{p-c}{p} = \frac{1}{\varepsilon}$, where c is a merchant's effective marginal cost, which includes its own cost κ_M and the fees paid to the network. Then $p = \left(1 + \frac{1}{\varepsilon - 1}\right) c$. When ε is smaller, the coefficient $1 + \frac{1}{\varepsilon - 1}$ becomes larger, meaning that a small drop in c can lead to a large drop in p.

Then $Q^P > Q^U$ if $MR^P(Q) > MR^U(Q)$, i.e.,

$$\kappa_M < p(Q)\frac{(\rho+2)(m\varepsilon-1)^2}{m\varepsilon\{m[\varepsilon(\rho+2)+1]-1\}} \equiv \widehat{\kappa}_M. \tag{13}$$

which can be shown (Appendix 7.1) to be more likely when κ_M is smaller, m is larger, or κ_N is larger when ρ is constant.

Proposition 3 *Compared to unit fees, proportional fees improve consumer surplus and social welfare if and only if the total output increases (i.e., $Q^P > Q^U$). In addition.*

(i) If the demand is sub-convex, then $Q^P > Q^U$ always holds.

(ii) If the demand is constant-elastic, then $Q^P > Q^U$ if $\kappa_N > 0$, and $Q^P = Q^U$ if $\kappa_N = 0$.

(iii) If the demand is super-convex, then

- *a) if $\kappa_N = 0$, then $Q^P < Q^U$;*
- *b) if $\kappa_N > 0$ and $\kappa_M = 0$, then $Q^P > Q^U$.*
- *c) if $\kappa_N > 0$ and $\kappa_M > 0$, then $Q^P > Q^U$ is more likely if κ_M is smaller, m is larger, or κ_N is larger when ρ is constant.*

Proposition 3 relates proportional fees' welfare consequence to all the underlying conditions in terms of demand (as captured by ε and ρ, which are themselves inter-related), cost (as captured by κ_N and κ_M), and the degree of merchants competition (as captured by m). The intuitions for sub-convex and constant-elastic demands have been explained.

If the demand is super-convex, the equilibrium comparison depends on the remaining three parameters: m, κ_M, and κ_N. Note that the tradeoff between the expansion and concession effects depends crucially on the endogenous $\beta(Q)$ under proportional fees. For any given Q, a larger $\beta(Q)$ means the network gets a larger share of the additional profit brought by output expansion, and therefore has a stronger incentive to expand.[13] Now look at how m and κ_M affect $\beta(Q)$. Under proportional fees, a merchant under-produces (i.e., distorts output) for two reasons: the merchant's own markup, and inflated cost due to revenue sharing.[14] If more merchants compete (i.e., m is larger), the merchant markup is smaller. If merchant costs are smaller (i.e., κ_M is smaller), the cost inflation is dampened. In both cases, the distortion is smaller for any given β, which in turn means that the network can raise β (to achieve any given Q) without worrying too much about merchant distortion under proportional fees. As a result, $Q^P > Q^U$ is more likely when there are more merchants or their costs are smaller.

[13] From the decomposition (12), the net effect is more likely to be positive if a given Q can be achieved through a larger $\beta(Q)$.

[14] Under proportional fees, a merchant maximizes $[(1-\beta)p(Q)-\kappa_M]q_i = (1-\beta)\left[p(Q)-\frac{\kappa_M}{1-\beta}\right]q_i$, which leads to merchant FOC: $p'(Q)\frac{Q}{m}+p(Q)=\frac{\kappa_M}{1-\beta}$. By contrast, the network would like the merchant to behave according to $p(Q)=\kappa_M$. Therefore, the first sources of distortion is $-p'(Q)\frac{Q}{m}$, i.e., the merchant's markup, and the second is the presence of $1-\beta$, i.e., revenue sharing inflates the merchant's true cost.

These two parameters' impacts are established for any given Q and therefore does not involve the characterization of the equilibrium choice. By contrast, how the network's cost, κ_N, affects output comparison is an equilibrium effect. A larger κ_N reduces the equilibrium output, which in return indicates that the network obtains a larger share of additional profit (i.e., a larger $\beta(Q)$). At the same time, the smaller output raises demand elasticity (due to super-convexity), which further reduces merchants' share of the additional profit (given constant ρ). Both effects tend to strengthen the expansion effect and reduce the concession effect.

Note that although κ_N and κ_M both reduce the equilibrium Q, they have opposite impacts on the comparison between Q^P and Q^U. The network's cost (κ_N) mainly determines how large the equilibrium output is under unit fees, while the merchants' cost (κ_M) determines how much the merchants' market power distorts the output choice.

Shy and Wang (2011) focus on constant-elastic demand, which corresponds to part (ii) of Proposition 3. They also explain that the driving force for welfare improvement is $\kappa_N > 0$ (i.e., double markup is imposed on κ_M but not κ_N). It is now clear that their results depend crucially on the special demand function. If demand is sub-convex, proportional fees still improve welfare even when $\kappa_N = 0$. More importantly, if the demand is super-convex, proportional fees may hurt welfare.

2.6 Merchants' profits

Under unit fees, the total profit of all merchants is $\Pi_M^U(Q) \equiv [p(Q) - t_C - t_M - \kappa_M]Q$. Given the network FOC in (4), $p(Q) - t_C - t_M - \kappa_M = -\frac{1}{m}p'(Q)Q$ at the equilibrium Q^U, so

$$\Pi_M^U(Q^U) = -\frac{1}{m}p'(Q^U)(Q^U)^2.$$

Under proportional fees, the total profit is $\Pi_M^P(Q) \equiv \{[1-\beta(Q)]p(Q) - \kappa_M\}Q$. By equation (5), $[1-\beta(Q)]\left[\frac{1}{m}p'(Q)Q + p(Q)\right] = \kappa_M$. Then at the equilibrium Q^P,

$$\Pi_M^P(Q^P) = m\pi_i^P = -\left[1-\beta(Q^P)\right]\frac{1}{m}p'(Q^P)(Q^P)^2.$$

Evaluated at the same Q, the merchants' profits are apparently smaller under proportional fees than under unit fees: $\Pi_M^P(Q) \le \Pi_M^U(Q)$. This is the mirror relation of the network's profit being higher under proportional fees, as the two parties' profits add up to the supply chain's total profit, which is invariant to the fee structure for any given Q: $\Pi(Q) = \Pi_M^P(Q) + \Pi_N^P(Q) = \Pi_M^U(Q) + \Pi_N^U(Q)$. However, different fee structure leads to different equilibrium Q. It is easy to show $\frac{\partial \Pi_M^P}{\partial Q} > 0$ and $\frac{\partial \Pi_M^U}{\partial Q} > 0$, meaning that given a fee scheme, the merchants' profit always increases with the total output. Therefore, when the network shifts from unit fees to proportional fees, the merchants' profit changes through two channels. First, fixing Q^U, the network gains, implying that the merchants lose. Second, moving from the equilibrium Q^U to the equilibrium Q^P, the merchants may gain or lose depending on whether the output increases or decreases. If $Q^P < Q^U$, as shown in the left panel of Figure 2, the two effects move in the same direction, and merchants are unambiguously worse off under proportional fees. If $Q^P > Q^U$, as shown in the right panel of Figure 2, the two effects move in opposite directions; the impacts on merchants are ambiguous.

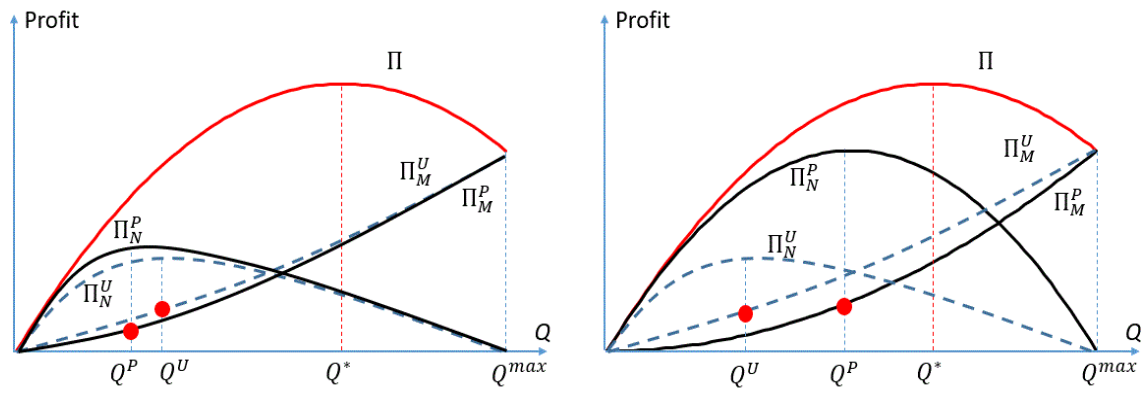

Figure 2: Network's and merchants' profits

The joint profit of the network and merchants, $\Pi(Q) = [p(Q) - \kappa_M - \kappa_N]Q$, is maximized by Q^*, which is characterized by

$$\kappa_M + \kappa_N = p'(Q^*)Q^* + p(Q^*). \tag{14}$$

Comparing (14) with (4), the extra term in the right-hand side of (4) indicates how the network's incentive under unit fees differs from joint profit maximization. The term is negative, indicating that the network produces too little from the viewpoint of joint profit, i.e., $Q^U < Q^*$. That is because when Q increases, the merchants will benefit as long as they have positive markup, but the network ignores such externality (Dingwei Gu, et al., forthcoming). By the same logic, $Q^P < Q^*$, as can be seen from the comparison between (14) and (8).

For any $Q < Q^*$, the joint profit of the network and merchants increases with Q, as Figure 2 shows. Given $Q^U < Q^*$ and $Q^P < Q^*$, we immediately conclude that proportional fees raise the joint profit if and only if $Q^P > Q^U$, which is also the condition for proportional fees to improve social welfare and consumer surplus (Proposition 3). Therefore,

Proposition 4 *The network and merchants' joint profit increases if and only if the equilibrium output increases. In particular,*

(i) If $Q^P > Q^U$, then proportional fees raise the joint profit.

(ii) If $Q^P = Q^U$, then proportional fees do not change the joint profit. The merchants are worse off.

(iii) If $Q^P < Q^U$, then proportional fees reduce the joint profit. The merchants are worse off.

Propositions 3 and 4 combine to conclude that the network and merchants' joint profit is perfectly synchronized with consumer surplus and social welfare. The three parties, i.e., the network, consumers, and merchants, are affected by proportional fees with increasing degrees of ambiguity: Improved network profitability is unconditional; improved consumer surplus is conditional; and improved merchant profitability is conditional even more strongly. If merchants are better off, then consumers must also be better off; if consumers are worse off, then merchants must also be worse off. The following table summarizes the impacts.

Table 1: Impacts of switching from unit fees to proportional fees

	network	consumer	merchant	network+merchant	welfare
If $Q^P > Q^U$	↑	↑	↑ or ↓	↑	↑
If $Q^P = Q^U$	↑	=	↓	=	=
If $Q^P < Q^U$	↑	↓	↓	↓	↓

Shy and Wang (2011) find that proportional fees always hurt merchants, based on a constant-elastic demand. Our analysis shows that the effect is ambiguous for more general demand.[15] Proposition 4 highlights the possibility for proportional fees to hurt merchants and social welfare simultaneously. The coexistence is excluded by Shy and Wang (2011), who then conclude that policy makers should not heed the complaints by merchants. We demonstrate that the relationship between merchants' profits and social welfare is ambiguous, and more factors need to be considered for policy making.

3 Competing networks

The analysis so far deals with a monopoly network. Now suppose there are $n \geq 2$ competing networks, each offering exclusive payment services to $\frac{m}{n} \geq 1$ merchants, so there are m merchants in total competing a la Cournot. Assume all networks use the same type of fee structures, either unit or proportional, to both consumers and merchants. Price neutrality then implies that consumer fees can be set to zero without any loss of generality.[16]

3.1 Unit fees

Under unit fees, the profit of merchant $i \in \{1, 2, .., \frac{m}{n}\}$ in network $j \in \{1, 2, .., n\}$ is $\pi_{ij}^U = [p(Q) - t_j - \kappa_M]q_{ij}$. The merchant FOC, $\frac{\partial \pi_{ij}^U}{\partial q_{ij}} = 0$, leads to:

$$\frac{n}{m}p'(Q)Q_j^U + p(Q) = \kappa_M + t_j, \tag{15}$$

[15] Indeed Johnson (2017, footnote 18) provides an example in which merchants benefit from proportional fees, based on a demand generated from the beta function.

[16] Consumers are not affiliated with any particular network. By the model setting, the out-of-pocket price that consumers pay, $p(Q)$, is the same regardless of which merchant a consumer buys from. As a result, consumers do not choose between merchants that belong to different networks even when networks charge differential fees to consumers—a larger fee paid to the network will be exactly offset by a smaller fee paid to the merchant so that the consumer's out-of-pocket fee remains the same.

where $Q_j^U = \sum_{i=1}^{\frac{m}{n}} q_{ij}$ is the total output of merchants in network j, and $Q = \sum_{j=1}^{n} Q_j^U$. Summing up (15) over j to have:

$$\frac{1}{m}p'(Q)Q + p(Q) = \kappa_M + \frac{1}{n}\sum_{j=1}^{n} t_j. \tag{16}$$

Network j's profit is $\Pi_j^U = (t_j - \kappa_N)Q_j^U$. Unlike a monopoly network for which choosing t is equivalent to choosing Q, here network j must choose t_j. We take derivative of (15) and (16) with respect to t_j to obtain expressions of $\frac{\partial Q}{\partial t_j}$ and $\frac{\partial Q_j}{\partial t_j}$, which are substituted into networks' FOCs (see Appendix 7.2 for details). Symmetry leads to:

$$\kappa_M + \kappa_N = p'Q + p - \frac{p'Q\left\{\left[(m-1)(n-1) - n\right]\left[p''Q + (m+1)p'\right] + (m-1)p'\right\}}{m\left[(n-1)p''Q + (mn+n-m)p'\right]}, \tag{17}$$

which will determine a unique equilibrium total output Q^{UC}. Note that p, p', p'' are all functions of Q. From (17), it can be established that $Q^{UC} > Q^*$ if $n > 2$, or if $m \geq 4$ when $n = 2$.

3.2 Proportional fees

Under proportional fees, the profit of merchant i in network j is $\pi_{ij}^P = \left[(1-\tau_j)\,p(Q) - \kappa_M\right] q_{ij}$. Merchant FOC leads to:

$$(1-\tau_j)\left[\frac{n}{m}p'(Q)Q_j^P + p(Q)\right] = \kappa_M. \tag{18}$$

Summing up the above equation over j to have:

$$\frac{1}{m}p'(Q)Q + p(Q) = \frac{\kappa_M}{n}\sum_{j=1}^{n}\frac{1}{1-\tau_j}. \tag{19}$$

Network j's profit is $\Pi_j^P = \left[\tau_j p(Q) - \kappa_N\right] Q_j^P$. By a similar method as for the unit fees, the equilibrium condition here can be derived as:

$$\begin{aligned}
\kappa_M + \kappa_N \quad = \quad & p'Q + p - \frac{(n-1)p'Q\left[p''Q + (m+1)p'\right]}{(n-1)p''Q + (mn+n-m)p'} \\
& + \kappa_M p'Q\frac{(n-1)\left[p''Q + (m+1)p'\right] - (m-1)p'}{\left[p'Q + mp\right]\left[(n-1)p''Q + (mn+n-m)p'\right]} \\
& + \kappa_M p'Q\frac{mnp\left[p''Q + (m+1)p'\right]}{\left[p'Q + mp\right]^2\left[(n-1)p''Q + (mn+n-m)p'\right]}
\end{aligned} \tag{20}$$

which will determine a unique equilibrium output Q^{PC}.

3.3 The discount effect

Under unit fees, network j's average revenue is $t_j = p - \kappa_M + \frac{n}{m}p'Q_j$, where Q_j is the total outputs of all merchants that belong to network j. Under proportional fees, it is

$$\begin{aligned} \tau_j p &= p - \kappa_M + \frac{n}{m}p'Q_j - \left[1 - \frac{\kappa_M}{\frac{n}{m}p'Q_j + p}\right]\frac{n}{m}p'Q_j \\ &= t_j + \underbrace{\left[-\beta_j \frac{n}{m}p'Q_j\right]}_{\text{Discount Effect } (+)}, \end{aligned}$$

where $\tau_j = \beta_j$ is network j's commission rate.

For given Q_j and Q_{-j} (i.e., output of all networks other than j), if network j switches from unit fees to proportional fees, its average revenue will increase. This is the same discount effect as in the monopoly network case. It captures how an individual network's merchants respond to its fee scheme, with the caveat that the outputs of other networks are fixed. To merchants that are affiliated with network j, other networks matter only through their respective merchants' outputs. Since these outputs are fixed, the competing networks' fee schemes are irrelevant for the discount effect that network j is experiencing.

3.4 Consumer surplus and social welfare

The welfare effect is again fully aligned with the equilibrium total output. To compare Q^{PC} with Q^{UC}, note that equation (17) can be viewed as being derived from $\frac{\partial \Pi_j^U/\partial t_j}{\partial Q_j^U/\partial t_j} = 0$, while (20) is derived from $\frac{\partial \Pi_j^P/\partial \tau_j}{\partial Q_j^P/\partial \tau_j} = 0$. Appendix 7.3 shows that $Q^{PC} > Q^{UC}$ if and only if (when evaluated at Q^{UC}):

$$\underbrace{\beta_j(\rho+2)}_{\text{Expansion Effect } (+)} + \underbrace{\beta_j(\rho+2)\frac{n-1}{n}\left[1 + \frac{2(m-1)}{\rho+2}\right]}_{\text{(Additional) Business-stealing Effect } (+)} + \underbrace{\left[-\left(1-\beta_j\right)\frac{\rho+m+1}{m\varepsilon-1}\right]}_{\text{Concession Effect } (-)} > 0.$$

The output comparison between the two fee schemes in oligopoly is driven by three effects. The concession and expansion effects are exactly the same as in monopoly, but there appears a third effect of (additional) business stealing, which tends to increase the output. To see the intuition, look at network j's additional revenue from proportional fees, roughly $\tau_j Q_j(\tau_j)$, and consider what happens when j lower τ_j, fixing other networks' τ.[17] Fixing the output, Q_j, a lower τ_j reduces the revenue; this is the concession effect. Because Q_j is fixed, other networks do not matter, and the effect is the same as in monopoly. On the other hand, fixing τ_j, a lower fee expands Q_j. There is not only a direct, expansion effect (j's merchants produce more due to lower τ_j, fixing other networks' outputs, which is the same expansion effect as in monopoly), but also an indirect, business-stealing effect, as the more competitive

[17] Here an individual network's incentive to expand must be analyzed through its choice of the fees rather than output quantities. In the monopoly network case, by contrast, choosing the fee and choosing the output are equivalent.

merchants affiliated with j (due to lower τ_j) grab market shares from competing networks, forcing them to reduce their outputs. The business-stealing effect is "additional" because business-stealing exists on competitive networks under both fee schemes, and here we are talking about the additional effect due to proportional fees.[18] In addition, it increases with the number of networks.

To see the net effect, rearrange the above expression to conclude that $Q^{PC} > Q^{UC}$ if and only if:[19]

$$\begin{aligned} \kappa_N \quad > \quad & -\frac{p(Q^{UC})}{[(n-1)(\rho+m+1)+1]} \\ & \times \frac{n(\rho+m+1)\left\{\varepsilon\left[n\left(\rho+1\right)+m\left(n-1\right)\right]+1\right\}}{\varepsilon\left\{\left[\left(m\varepsilon-1\right)\left(2n-1\right)+n\right]\left(\rho+m+1\right)-\left(m-1\right)\left(m\varepsilon-1\right)\right\}}. \end{aligned} \tag{21}$$

When $n = 1$, (21) degenerates into (11). When $n > 1$ (and, of course, $m > 1$, too), we have

$$\varepsilon\left[n\left(\rho+1\right)+m\left(n-1\right)\right]+1 = \varepsilon\left[n\left(\rho+2\right)+\left(m-1\right)\left(n-1\right)-1\right]+1 > 0,$$

regardless of demand convexity. Therefore, the right hand side of (21) is always negative, meaning that the condition always holds (see Appendix 7.3 for details).

Proposition 5 *When symmetric networks compete (i.e., $n \geq 2$), proportional fees always increase equilibrium output (i.e., $Q^{PC} > Q^{UC}$) and, therefore, always raise consumer surplus and social welfare.*

For a monopoly network, proportional fees may increase or decrease the equilibrium output depending on the comparison between the expansion and concession effects. Here when networks compete, the concession effect remains the same, but the expansion effect is amplified by the business-stealing effect. As a result, the equilibrium output always increases.

3.5 Profits

For a monopoly network, the discount effect is sufficient to determine the network's profitability. When networks compete, however, this is no longer the case, as a network's profit also depends on other networks' equilibrium choices. If other networks' outputs are fixed, a network benefits from proportional fees. However, all networks expand outputs simultaneously in equilibrium (Proposition 5), which hurts one another.

To investigate the impacts of proportional fees on network profitability, it helps to first look at the impacts on the joint profit of all networks and merchants, which is easier because

[18] The sign of the additional business stealing effect is independent of the nature of competition among merchants, whether Cournot or Bertrand.

[19] The condition can be equivalently expressed in terms of κ_M:

$$\kappa_M < p(Q^{UC})\frac{(m\varepsilon-1)^2}{m\varepsilon}\frac{(2n-1)(\rho+m+1)-(m-1)}{\left[\left(m\varepsilon-1\right)\left(2n-1\right)+n\right]\left(\rho+m+1\right)-\left(m-1\right)\left(m\varepsilon-1\right)}.$$

it is closely related to the comparison between Q^{PC}, Q^{UC}, and Q^* (the one that maximizes the joint profit). If $n > 2$, or if $m \geq 4$ when $n = 2$, then $Q^{UC} > Q^*$. In that case, unit fees already lead to over-production for the networks and merchants collectively. By encouraging output expansion ($Q^{PC} > Q^{UC}$), proportional fees exacerbate the situation, and the joint profit can only be lower.

Proposition 6 *When there are more than two networks (i.e., $n > 2$), or no fewer than four merchants in duopoly networks (i.e., $n = 2$ and $m \geq 4$), proportional fees reduce the joint profit of all networks and merchants.*

As long as there is a little competition among networks or merchants, proportional fees hurt the two parties as a whole. It is impossible for both to gain (and they may both lose). Switching to proportional fees at a given Q will benefit networks at the expense of merchants, but the resulting greater Q will benefit merchants at the expense of networks.[20] For either party, the net effect is ambiguous.

Both Propositions 5 and 6 concern the equilibrium output, but the former is unconditional while the latter needs a condition. This is because when m and n are both very small, $Q^{UC} < Q^*$ is possible so that merchants under-produce under unit fees. Then the output expansion brought by proportional fees may mitigate the under-production and generate a greater joint profit. Appendix 7.4 shows such an example.

For a monopoly network, the unconditional higher profit directly implies the incentive to adopt proportional fees. When networks compete, they may be collectively worse off, as argued above. Even so, each network may still have an individual incentive to adopt proportional fees. Appendix 7.4 shows an examples in which proportional fee is each network's dominant strategy even though they are collectively worse off. In other words, networks may face a Prisoners' Dilemma when choosing between the two fee schemes.

3.6 Constant-elastic demand

For the general demand, the impacts of proportional fees on network profitability are intractable, but we may gain some insights from more specific settings. Consider constant-elastic demand $p(Q) = \alpha Q^{-\frac{1}{\varepsilon}}$ with $\varepsilon > 1$. Based on closed-form solutions (Appendix 7.5), a sufficient (but not necessary) condition for networks to gain from proportional fees, i.e., $\Pi_N^P > \Pi_N^U$, is

$$\frac{\kappa_N}{\kappa_M} < \frac{mn\varepsilon^2}{(m\varepsilon - 1)\left[(n-1)m\varepsilon - n\right]\left[(n-1)(m\varepsilon - 1) + \varepsilon\right]}, \tag{22}$$

which is easier to hold when κ_M is larger, or κ_N, ε, m, or n is smaller. In particular, $\kappa_N = 0$ is sufficient to guarantee $\Pi_N^P > \Pi_N^U$. When $n = 1$, the condition becomes $\frac{\kappa_N}{\kappa_M} > -\frac{m\varepsilon}{m\varepsilon - 1}$,

[20] These can be seen from Figure 2 (focusing on the right half of both curves for output quantities greater than Q^*). Given the symmetric setting, the only difference between monopoly network and and competitive networks is the equilibirum output. The profit curves as functions of Q are similar, so we can use the same graph to discuss both cases.

which always holds.

On the other hand, if:

$$\frac{\kappa_N}{\kappa_M} > \frac{mn}{(m-1)\left[(n-1)^2(m-1)^2 - 1\right]}, \tag{23}$$

then there always exist demands that sufficiently inelastic ($\varepsilon \to 1$) for which $\Pi_N^P < \Pi_N^U$. Note that (23) holds when κ_M is sufficiently small, or κ_N, m, or n is sufficiently large.[21]

(22) and (23) are both sufficient conditions. Appendix 7.5 offers a sufficient and necessary condition for $\varepsilon = 2$. Calculation shows $\Pi_N^P > \Pi_N^U$ is more likely if κ_M is larger, or κ_N, m, or n is smaller.

Proposition 7 *Suppose competing networks face constant-elastic demand:* $p(Q) = \alpha Q^{-\frac{1}{\varepsilon}}$ *with* $\varepsilon > 1$.

(i) If condition (22) holds, then proportional fees raise networks' profits.

(ii) If condition (23) holds and the demand is sufficiently inelastic, then proportional fees reduce networks' profits.

For monopoly network, network profitability is unconditional, while welfare improvement is conditional. For competing networks, network profitability is conditional, while welfare improvement is unconditional. The two conditions are actually related: welfare improves in monopoly when Q^P is large, while networks benefit in competition when Q^{PC} is small. Indeed the conditions are consistent: Proposition 3 says that welfare improvement is more likely for small κ_M or large κ_N or m; Proposition 7 says that networks are more likely to benefit for large κ_M or small κ_N or m.

4 Extensions and robustness

This paper is built on a key insight: proportional fees mitigate the double markup problem and therefore have the *potential* to increase total surplus.[22] This is a property independent of equilibrium characterization, i.e., it holds for any given output or fee level, and is therefore very robust. All the other results are either further development of the discount effect, or the consequence of additional forces working on it. For example, the equilibrium on a monopoly network is determined by a tradeoff between the expansion and concession effects (which are about marginal revenues) that arise directly from the discount effect (which is about the average revenue). When networks compete, the expansion effect is amplified by a business stealing effect, resulting in a strong incentive to adopt proportional fees even if networks' profits may drop in equilibrium.

[21] This is the condition for $n \geq 2$. When $n = 1$, the condition becomes $\frac{\kappa_N}{\kappa_M} < -\frac{m}{m-1}$, which of course cannot hold.

[22] Wang and Wright (2017) offer another explanation: Proportional fees allow a network to price discriminate among heterogenous merchants.

In what follows, we will check the robustness of the discount effect in Bertrand competition among merchants and two-sided market with cross network externality. The analysis is also extended to mixed fees, i.e., merchants and consumers are levied different types of fees.

4.1 Bertrand competition

Consider a monopoly network on which merchants carry out Bertrand competition with differentiated products. The setting is the same as in Section 2 except that all merchants face symmetric final demand $Q_i(p_i, \mathbf{p}_{-i})$ with $\frac{\partial Q_i}{\partial p_i} < 0$. Unique equilibrium is guaranteed by usual assumptions about Bertrand competition, which are omitted for brevity.

Under unit fees, the profit of merchant i is $\pi_i^U = [p_i - t_C - t_M - \kappa_M]Q_i(p_i, \mathbf{p}_{-i})$. The merchant FOC is:

$$p_i - t_C - t_M - \kappa_M + \frac{Q_i}{\frac{\partial Q_i}{\partial p_i}} = 0,$$

which in turn generates the average revenue from merchant i and its customers:

$$t_C + t_M = p_i - \kappa_M + \frac{Q_i}{\frac{\partial Q_i}{\partial p_i}}.$$

Under proportional fees, the profit of merchant i is $\pi_i^P = \left[\frac{1-\tau_M}{1+\tau_C} p_i - \kappa_M\right] Q_i(p_i, \mathbf{p}_{-i})$. The merchant FOC is:

$$\frac{1-\tau_M}{1+\tau_C}\left[Q_i + p_i \frac{\partial Q_i}{\partial p_i}\right] = \kappa_M \frac{\partial Q_i}{\partial p_i}.$$

As a result, the average revenue is:

$$(\tau_C + \tau_M)p_i = p_i - \kappa_M + \frac{Q_i \kappa_M}{Q_i + p_i \frac{\partial Q_i}{\partial p_i}}.$$

The discount effect is

$$(\tau_C + \tau_M)p_i - (t_C + t_M) = \underbrace{-\frac{\tau_C + \tau_M}{1+\tau_C}\frac{Q_i}{\frac{\partial Q_i}{\partial p_i}}}_{\text{Discount Effect } (+)}.$$

Therefore, discount effect still shows up in Bertrand competition as long as merchants have market power, i.e., $\frac{\partial Q_i}{\partial p_i} < \infty$.

4.2 Two-sided markets

The payment card industry is commonly regarded as a two-sided market. There are several definitions in the literature about what constitutes a two-sided market (Weyl, 2010). According to Rochet and Tirole (2003, 2006), the key characterization of a two-sided market is price non-neutrality. In addition, they argue that price neutrality holds for payment cards as long as three mild conditions are satisfied, so the industry should be regarded as a one-sided market despite apparent two-sidedness (Rochet and Tirole, 2006, p.648). Indeed, price neutrality holds in our model. That is why the setting can be interpreted (or framed) as an upstream-downstream vertical relationship.

Another definition for two-sided market is the existence of cross network externality, i.e., the utility derived by users on one side increases with the network size of the other side (Armstrong 2006, Rysman 2009, and others). To investigate how this might affect our results, we endogenize cross network externalities through merchant free entry. More specifically, assume that there are n_C identical consumers, each of whom has the demand $p(Q)$. For any fee scheme, after the network announces its particular choice of fee levels, each potential merchant must pay an entry cost, $H > 0$, in order to operate. Therefore, the number of merchants, m, is endogenized.

Under unit fees, given m, the merchant FOC is identical to that in the main model, (1). In equilibrium, $q_i = \frac{Q}{m}$, and merchant i's profit is

$$\pi_i^U = -n_C p'(Q) \left(\frac{Q}{m} \right)^2 - H.$$

The free entry condition requires $\pi_i^U = 0$, which means that the endogenous number of merchants, $m^U(Q)$, must satisfy:

$$-n_C p'(Q) \left(\frac{Q}{m^U(Q)} \right)^2 = H.$$

Here m^U is treated as a function of Q. Then the network's average revenue from every individual consumer is

$$D^U(Q) = p(Q) - \kappa_M - \frac{m^U(Q)}{Q} H.$$

Under proportional fees, the same procedure is used. Merchant i's profit is

$$\pi_i^P = -n_C \frac{1-\tau_M}{1+\tau_C} p'(Q) \left(\frac{Q}{m} \right)^2 - H.$$

The endogenous number of merchants, $m^P(Q)$, must therefore satisfy:

$$-n_C \frac{1-\tau_M}{1+\tau_C} p'(Q) \left(\frac{Q}{m^P(Q)} \right)^2 = H.$$

The network's average revenue from every individual consumer is

$$D^P(Q) = p(Q) - \kappa_M - \frac{m^P(Q)}{Q} H$$

From the expressions of π_i^U and π_i^P, it is clear that the more consumers, the larger is each merchant's profit, and thus a larger number of merchants. That is, consumers have positive cross network externality to merchants. On the other hand, the larger number of merchants, the larger is equilibrium output. Since consumers surplus, $S = \int_0^Q p(x)dx - p(Q)Q$, increases with the level of output ($\frac{\partial S}{\partial Q} = -p'(Q)Q > 0$), the more merchants, the larger is consumer surplus. Therefore, merchants also have positive cross-side network externality to consumers.

For any given Q, since $-n_C\left[1-\beta(Q)\right]p'(Q)\left(\frac{Q}{m^P}\right)^2 = -n_C p'(Q)\left(\frac{Q}{m^U}\right)^2 = H$, we have

$$m^P = \sqrt{1-\beta(Q)}m^U < m^U.$$

Then

$$\begin{aligned}
D^P(Q) - D^U(Q) &= \underbrace{\frac{H}{Q}\left(m^U - m^P\right)}_{\text{Net effect }(+)} \qquad (>0 \text{ because } m^U > m^P) \\
&= [1-\beta(Q)]p'(Q)\frac{Q}{m^P} - p'(Q)\frac{Q}{m^U} \\
&= \underbrace{-\beta(Q)p'(Q)\frac{Q}{m^P}}_{\text{Discount Effect }(+)} + \underbrace{p'(Q)Q\left(\frac{1}{m^P} - \frac{1}{m^U}\right)}_{\text{Entry/Exit effect }(-)}.
\end{aligned}$$

The intuition is the following. If the number of merchants, m, is the same between the two fee schemes, $D^P(Q) > D^U(Q)$ for any given Q, which is the discount effect established in the main model. Now allow m to be endogenously determined. Given Q, proportional fees will reduce the merchants' markup, which leads to a smaller m. The reduced competition would partially restore merchant markup. This is the entry/exit effect. In the end, the discount effect always dominates the entry/exit effect and, hence, $D^P(Q, m^P(Q)) > D^U(Q, m^U(Q))$. An immediate corollary is that, with free-entry merchants, a monopoly network always earns a higher profit under proportional fees than under unit fees.

4.3 Other considerations

Our model has assumed that merchants have the same marginal cost. If merchants' costs differ, the equilibrium output will not change as long as their average marginal cost remains the same. In that case, proportional fees will generate an additional welfare gain by encouraging efficient merchants to produce more, and less efficient merchants to produce less. This can be seen from $q_j^P - q_i^P = -\frac{c_i - c_j}{(1-\beta)p'(Q)} > -\frac{c_i - c_j}{p'(Q)} = q_j^U - q_i^U$ for $c_i > c_j$. Such effect will further strengthen the advantage of proportional fees.

Consumers have to use the card in the model. What if they can transact with merchants without using any card? What if some constraints prevent the network from hurting merchants? This would be the case, for example, if the number of merchants is endogenized through free entry as analyzed above. More generally, these two questions essentially ask whether the discount effect still holds when the other two parties, i.e., consumers and merchants, must be given some minimum payoff due to outside options. The discount effect implies that when a network switches from unit fees to proportional fees, it is always *feasible* to increase both consumer surplus and the joint profit of the network and its merchants, which leaves room for Kaldor–Hicks improvement (but not necessarily Pareto improvement). If some parties have to earn some minimum payoffs, it would still be feasible for the network to benefit from proportional fees while keeping all those parties equally better off. In other words, the presence of outside options will only change the particular equilibrium choice, but not the profitability of proportional fees for a monopoly network.

In network competition, we have assumed that a merchant belongs to a network exclusively. What happens if a merchant can freely choose which network it wants to join after observing the fees announced by all networks? Fixing all other networks' fee types, if a particular network switches from unit fees to proportional fees, it is feasible for its whole supply chain system (i.e., the network and all merchants that transact on it) to be collectively better off due to the discount and business-stealing effects. Therefore, whatever terms the network offers to its merchants under unit fees, it can make a better offer under proportional fees and at the same time earn more profit. This would increase the network's competitiveness against rival networks in attracting merchants. Therefore, when networks actively compete for merchants, proportional fee should still be the optimal choice for each network individually.

What if a merchant can join several networks? Multi-homing by merchants can be modeled, for example, if a merchant must pay some entry cost in order to join a network. Then the above argument about the effect of outside options will apply, and the discount effect should still be valid as long as merchants have market power. Multi-homing on consumer side is inconsequential, as the model setting does not restrict individual consumers to exclusive networks; each consumer buys only once and simply chooses a particular merchant.

4.4 Mixed fees

In the main analysis a network is assumed to charge the same type of fees to both merchants and consumers. This leads to price neutrality: although the network seems to have two instruments, what really matters is a composite fee, $t_M + t_C$ in the case of unit fee, and $\frac{\tau_M+\tau_C}{1+\tau_C}$ in the case of proportional fees. What will happen if the two sides of the market are charged different types of fees, referred to as mixed fees?

For simplicity, we will consider only monopoly network. Suppose the network charges a unit fee, t_M, to merchants, and a proportional fee, τ_C, to consumers. Then merchant FOC is:

$$\frac{1}{m}p'(Q)Q + p(Q) = (\kappa_M + t_M)(1 + \tau_C).$$

For any given Q, then, there is an inverse relationship between t_M and τ_C. The network's profit is

$$[p(Q) - \kappa_M - \kappa_N] Q + \frac{1}{1+\tau_C}\frac{1}{m}p'(Q)Q^2.$$

After substituting one of the fees, the other fee remain an independent choice for given Q. In other words, price neutrality no longer holds. It can be shown that the network's profit increases with the proportional fee, or equivalently decreases with the unit fee. The optimal choice is $t_M \to -\kappa_M$ and $\tau_C \to +\infty$. Notice that we do not require $\tau_C \leq 1$. If the infinite proportional fee looks a little strange, we only need to point out that the true proportion is $\frac{\tau_C}{1+\tau_C}$, which approaches 1, a much more sensible outcome. In equilibrium, the network achieves monopoly outcome (i.e., Q^*, which maximizes the joint industry profit), and leaves zero profit for the merchants.

If the network charges a unit fee, t_C, to consumers and a proportional fee, τ_M, to merchants, the result is the same. In particular, merchant FOC is:

$$\frac{1}{m}p'(Q)Q + p(Q) = \frac{\kappa_M}{1-\tau_M} + t_C.$$

And the network's profit is

$$[p(Q) - \kappa_M - \kappa_N]\, Q + (1-\tau_M)\, \frac{1}{m}p'(Q)Q^2.$$

The optimal choice is $\tau_M \to 1$ and $t_C = -\frac{\kappa_M}{1-\tau_M}$. Again, the network is able to obtain the monopoly outcome.

Proposition 8 *If a monopoly network uses mixed fees,*

(i) it will subsidize one side (to which it charges the unit fee) and extract maximal surplus from the other side (to which it charges the proportional fee).

(ii) The network earns the monopoly profit, while merchants earn zero profit.

(iii) Social welfare and consumer surplus are larger under mixed fees than under pure fees (i.e., when both sides pay unit fees or both sides pay proportional fees).

Mixed fee is a combination of the two (pure) fee schemes, but the equilibrium outcome under mixed fees is not a midway combination between the two arrangements. Basically, mixed fees give the network two truly independent instruments. It will then use a negative unit fee to increase the size of the pie, and use the remaining tool of proportional fee to extract all the surplus from the other side. In the main model, subsidy to one side does not work because, due to price neutrality, what really matters is the composite fee. Since the network's average revenue comes from the composite fee, it cannot be negative in equilibrium.

5 Applications and discussion

As explained earlier, the setting of payment card is almost identical to other settings such as retailing, taxation and technology licensing. In what follows, we will first discuss how our results compare with those developed for retailing. We then move on to tax competition and system competition, before finally noting a peculiar feature of proportional fees that may dampen merchant innovation.

5.1 Retailing

In the wholesale model, merchants sell their products to the platform, who resells them to final consumers. In the agency model, the platform sets commission rates first, and then merchants set the final retail prices.[23] If a platform is interpreted as our card network, then the agency model corresponds exactly to our proportional fees. As a result, the equilibrium output will be characterized by (8). The wholesale model is similar to our unit fees scheme, except that the wholesale price is set by merchants, whereas the unit fee is set by the card network.

5.1.1 Monopoly platform

To characterize the equilibrium in the wholesale model on a monopoly platform, suppose that merchants' Cournot competition results in a wholesale price, t. Facing t, the network's profit is $[p(Q) - t - \kappa_N]Q$. Treating Q as the choice variable, network FOC leads to $p'(Q)Q + p(Q) = \kappa_N + t$. Anticipating this, the profit of merchant i is $(t - \kappa_M)\, q_i = [p'(Q)Q + p(Q) - \kappa_M - \kappa_N]\, q_i$ if it produces q_i. The resulting merchant FOC, after substituting $q_i = \frac{Q}{m}$, is identical to (4). Therefore,

Lemma 1 *For a monopoly network or platform, units fees (in which the network sets the fee) and the wholesale model (in which the merchants set the fee) lead to exactly the same equilibrium output and social welfare.*

Lemma 1 establishes a role neutrality for a monopoly network or platform such that the wholesale model is equivalent to unit fees. As a result, our conclusions about social welfare directly carry over to the comparison between the two business models. However, each party's profit is incomparable due to the role swapping.

In a setting of bilateral monopoly for the ebook market, Gaudin and White (2014, Proposition 2 for the "non-essential case") find that the agency model leads to a strictly larger output than the wholesale model if the demand is sub-convex. Our model allows competitive merchants and is therefore more general. Our result is also stronger, as (11) and (13) are necessary and sufficient conditions covering parameters beyond the demand convexity, whereas Gaudin and White (2014) only establish the sufficient condition (if the demand is super-convex, they show that the result is reversed for *some* marginal cost).

Johnson (2017) shows that for bilateral monopoly, sub-convex demand is sufficient for the agency model to raise output. We allow for merchant competition, and provide the sufficient and necessary condition.

[23] There can also be a franchise model, in which merchants set shares, and platforms/retailers set retail prices. Johnson (2017) groups the agency and franschise models into one category called revenue sharing, in which one party sets shares and the other party sets retail prices.

5.1.2 Competing platforms

Johnson (2017) models bilateral oligopoly with a conduct-parameter approach, in which the conducts of both first-movers and second-movers are captured by exogenous parameters that are invariant to the business models he is comparing. Such approach precludes horizontal strategic interactions within each layer and seems to forego a major force from the outset. By contrast, our standard oligopoly approach explicitly models strategic interactions among networks. All movers' conducts are endogenous and therefore differ naturally between the two fee schemes. Strategic interaction among networks gives rise to a business stealing effect that drives the welfare outcome. This fundamental difference explains the different or even contradictory conclusions between this paper and Johnson's.

For platform profitability, Johnson (2017) shows that revenue-sharing (equivalent to proportional fees) benefits first-movers (equivalent to our networks) but hurts second-movers (equivalent to merchants) when the demand is either log-linear (which is a subset of sub-convexity) or constant-elastic. We show that the profit comparison is ambiguous. For example, with constant-elastic demand, proportional fees may hurt networks.

For welfare effect, Johnson (2017) finds that revenue-sharing improves welfare if the elasticity of Mills ratio is less than one (equivalent to sub-convex demand). By contrast, we show that welfare improvement is unconditional. This is because our setting allows strategic interaction among networks, which gives rise to an additional business stealing effect such that the quantity expansion effect always outweighs the concession effect.

5.2 Tax competition and system competition

The introduction has discussed how our results compare with those in taxation and licensing in a monopoly setting. Now consider tax competition among regional or national governments. The "standard" model in the literature assumes welfare-maximizing governments that implement specific taxes, and usually finds that tax competition reduces welfare by lowering tax rates and hence public expenditures (George R. Zodrow and Peter Mieszhkowski 1986; John D. Wilson 1986; 1999). The "Leviathan" model challenges such view by arguing that tax competition may improve welfare by taming overblown governments (Wallace E. Oates 1985; Michael Rauscher 1998; John D. Wilson and David E. Wildasin 2004). Our analysis suggests that tax type also matters if double markup is present. A unilateral switch to ad valorem tax by a single government can increase social surplus, making the government more competitive in attracting capital even if a required amount of tax revenue must be collected. Without such requirement, tax competition may indeed lower all governments' provision of public goods (corresponding to the decreased network profits), but the social gain from mitigating double markup (corresponding to increased social welfare) is an additional, positive effect. Finally, competition will further amplify the social gains from the discount effect. Even if welfare is reduced by tax competition using specific taxes, it may still be possible for tax competition to increase welfare if ad valorem tax is used.

Our framework of competing networks may help understand system competition. In real life, many incompatible systems or standards coexist and compete with one another, especially in the computer, telecommunication, electronics and payment card industries. Examples include DOS vs. UNIX; iOS vs. Android; IBM-Microsoft vs. Apple computers; Visa vs. American Express vs. Discover; CDMA vs. TDMA vs. GSM, and so on (Stanley M. Besen and Joseph Farrell1994; Michael L. Katz and Carl Shapiro, 1994). The literature predominantly focuses on compatibility choices and how to win a standards war (Joseph Farrell and Garth Saloner, 1986; Michael L. Katz and Carl Shapiro, 1985, and many others). Our research suggests that a useful competitive strategy is to license technologies based on ad valorem rather than per-unit royalties. A unilateral adoption of such strategy will give a system advantages in competition and, by extension, may even affect the coexistence outcomes. However, simultaneous adoption by competing systems may hurt them all, but will benefit consumers and society.

5.3 Proportional fees and merchant innovation

Proportional fees have a peculiar feature: merchants' profits are non-monotonic in their own costs. To see this, recall that merchant distortion is smaller if κ_M is smaller. In the extreme case of $\kappa_M = 0$, a merchant maximizes $(1-\beta)p(Q)q_i$, which is equivalent to maximizing $p(Q)q_i$. This allows the network to charge the maximal β, leaving zero profit for the merchants, $\Pi_M^P(Q^P) = 0$ when $\kappa_M = 0$.

More generally, the merchants' profits under proportional fees,

$$\Pi_M^P = -\frac{\kappa_M}{p'(Q^P)Q^P + mp(Q^P)} p'(Q^P)(Q^P)^2,$$

is a non-monotonic function of κ_M. When $\kappa_M \to 0$, the equilibrium commission rate $\beta \to 1$, and $\Pi_M^P \to 0$, as explained above. When $\kappa_M \to p(0)$, the merchant production cost is so high that consumers can barely afford the product, leading to $Q^P \to 0$ and again $\Pi_M^P \to 0$. Since Π_M^P is non-negative and continuous in κ_M, the following is established:

Proposition 9 *Under proportional fees, the merchants' equilibrium profit is non-monotonic in their own costs. In particular, when κ_M is sufficiently small, the profit increases with the cost.*

For an example, consider constant-elastic demand $p = \alpha Q^{-\frac{1}{\varepsilon}}$, which leads to:

$$\Pi_M^P(Q) = \frac{Q}{m\varepsilon}\left[\alpha\left(1-\frac{1}{\varepsilon}\right)Q^{-\frac{1}{\varepsilon}} - \kappa_N\right]$$

in equilibrium, with $\frac{\partial \Pi_M^P(Q)}{\partial Q} = \frac{(1-\frac{1}{\varepsilon})^2\alpha Q^{-\frac{1}{\varepsilon}} - \kappa_N}{m\varepsilon}$. Obviously, the merchants' profit is inversely U-shaped in Q, and consequently inversely U-shaped in κ_M. In addition, Π_M^P is maximized at $Q^* = \left[\frac{\alpha\left(1-\frac{1}{\varepsilon}\right)^2}{\kappa_N}\right]^{\varepsilon}$, or equivalently $\kappa_M^* = \frac{m\varepsilon-1}{m\varepsilon(\varepsilon-1)}\kappa_N$. Notice that $\frac{m\varepsilon-1}{m\varepsilon(\varepsilon-1)}$ decreases with the elasticity, ε, which means that a larger demand elasticity will encourage merchants to reduce their costs.

When a firm's profit increases with its own cost, the firm's incentive to reduce the cost through innovation will be discouraged. Since this perversive effect exists only for proportional fees, an important implication is that if merchant innovation is of some concern, proportional fees may have an additional, adverse effect on social welfare as compared to unit fees.

6 Conclusions

Proportional fees encourage output expansion by reducing a merchant's cost of expansion. In situations where double markup leads to under-production, output expansion can be achieved through other methods. For example, an upstream firm can offer quantity discounts to downstream firms. In fact, proportional fees are equivalent to a continuous quantity discount. If downstream firms hold passive ownership of upstream firms, a downstream firm's effective cost is reduced,[24] which would apparently encourage output expansion (Dingwei Gu, et al., 2018).

Two-part pricing is also a commonly observed practice.[25] If this pricing scheme is applied to consumers, then price neutrality implies that a change in marginal consumer pricing will not affect our results qualitatively. The only difference is that consumer surplus is turned into the network's additional profit. On the other hand, if two-part pricing is applied to merchants, then double markup will be solved completely. In that case, merchants will not distort output under either fee schemes, and the advantage of proportional fees disappears. This is similar to mixed fees in the sense that if a network has two independent instruments, there is no need for proportional fees.

Our analysis demonstrates that proportional fees improve welfare when at least two networks compete, but may reduce welfare if the network is a monopolist. Therefore, the easiest way to promote welfare is to encourage competition (among networks and/or among merchants) rather than lowering the fee, which usually involve costly litigations. Competition can be encouraged by regulations that reduce switching cost. For example, Facebook users should be allowed to take their social connections with them when migrating to another network.

The Introduction mentions the 2012 ebook price fixing case. In fact the case is complicated because it involves complements (i.e., ebook reader such as Amazon's Kindle and Apple's iPad or iPhone), substitutes (the sale of an ebook cannibalized the same title's hardcover version, which is more lucrative for the publisher), as well as complementary arrangements such as the

[24] If a downstream firm holds 50% of its upstream firm, then given any wholesale price w, the downstream firm will receive a compensation of $\frac{w}{2}$ for every unit it sells (through the passive ownership), which means its effective wholesale price is only $\frac{w}{2}$.

[25] Bedre-Defolie and Calvano (2013) address two-part tariff in payment card pricing.

Most Favored Nation (MFN) clause that Apple imposed on publishers along with the agency model. The antitrust case focused on whether Apple and the five publishers conspired to raise ebook prices rather than the agency model per se. In fact, Supreme Court explicitly asserted that "both courts emphasized that their decisions cast no doubt on the 'broader legality' of the contract terms at issue, such as the 'agency model and MFNs'" (Supreme Court 2015). The agency model is involved in many other cases mentioned in Introduction because price-fixing is easier when retail prices are set by retailers, especially coupled with MFN.

This research has generated some insights toward the fee types in payment card, retailing, and other industries. Given the policy importance, more research needs to be done to take into account specific economic background of each industries.

7 Appendix

7.1 Proposition 3: Welfare on monopoly network, super-convex demand

For super-convex demand with $\kappa_N > 0$, we have established that $Q^P > Q^U$ under condition (13), which is for any arbitrary Q. If $Q = Q^U$, then condition (4) can be used to show that (13) is equivalent to (11). Since $\widehat{\kappa}_M > 0$, (13) always holds if κ_M is sufficiently small, in particular if $\kappa_M = 0$. Since κ_M is exogenous, for any given Q and $p(Q)$ (hence ε and ρ are fixed), (13) is more likely to hold when κ_M is smaller or m is larger. Finally, since both MRs are downward sloping, a larger κ_N will lead to a smaller Q and consequently a larger $p(Q)$. If ρ is constant, $\frac{(\rho+2)(m\varepsilon-1)^2}{m\varepsilon\{m[\varepsilon(\rho+2)+1]-1\}}$ decreases in ε and consequently decreases in Q (as ε increases in Q for super-convex demands). Therefore, when ρ is constant, (13) is more likely to hold when κ_N is larger.

7.2 Network competition: equilibrium characterization

Unit fees: To guarantee the profit function is well-behaved, we need to assume $\frac{\partial^2 \Pi_j^U}{\partial t_j^2} < 0$. From (15) we have $\frac{n}{m}\left[p''Q_j + p'\right]\frac{\partial Q}{\partial t_j} + \frac{n}{m}p'\frac{\partial Q_j}{\partial t_j} = 1$. From (16) we have $\frac{1}{m}\left[p''Q + (m+1)p'\right]\frac{\partial Q}{\partial t_j} = \frac{1}{n}$, which gives $\frac{\partial Q}{\partial t_j} = \frac{m}{n[p''Q+(m+1)p']}$. Substitute $\frac{\partial Q}{\partial t_j}$ into the above equation to obtain $\frac{\partial Q_j}{\partial t_j} = \frac{m}{n^2 p'}\frac{(n-1)p''Q+(mn+n-m)p'}{p''Q+(m+1)p'}$, which is then plugged into $\frac{\partial \Pi_j^U}{\partial t_j} = Q_j + (t_j - \kappa_N)\frac{\partial Q_j}{\partial t_j} = 0$, or equivalently, $(t_j - \kappa_N) + \frac{Q_j}{\frac{\partial Q_j}{\partial t_j}} = 0$. Finally, substitute $t_j = \frac{1}{m}p'Q + p - \kappa_M$ and $Q_j = \frac{Q}{n}$ to arrive at (17). To summarize, under unit fees,

$$
\begin{aligned}
\frac{\partial Q}{\partial t_j} &= \frac{m}{n\left[p''Q + (m+1)p'\right]} \\
\frac{\partial Q_j}{\partial t_j} &= \frac{m}{n^2 p'}\frac{(n-1)p''Q + (mn+n-m)p'}{p''Q + (m+1)p'} \\
t_j &= \frac{1}{m}p'Q + p - \kappa_M
\end{aligned}
$$

Proportional fee: From (18) we have $\frac{\kappa_M}{(1-\tau_j)^2} = \left[\frac{n}{m}p''Q_j + p'\right]\frac{\partial Q}{\partial \tau_j} + \frac{n}{m}p'\frac{\partial Q_j}{\partial \tau_j}$. From (6) we have $\frac{\kappa_M}{n(1-\tau_j)^2} = \frac{1}{m}\left[p''Q + (m+1)p'\right]\frac{\partial Q}{\partial \tau_j}$, which gives $\frac{\partial Q}{\partial \tau_j} = \frac{m\kappa_M}{n(1-\tau_j)^2[p''Q+(m+1)p']} = \frac{[p'Q+mp]^2}{nm\kappa_M[p''Q+(m+1)p']}$ (given $1-\tau_j = \frac{m\kappa_M}{p'Q+mp}$ in equilibrium). Substitute $\frac{\partial Q}{\partial \tau_j}$ into the above equation to have $\frac{\partial Q_j}{\partial \tau_j} = \frac{[p'Q+mp]^2}{nm\kappa_M[p''Q+(m+1)p']}\frac{[(n-1)p''Q+(mn+n-m)p']}{np'}$. Then, plug them into $\frac{\partial \Pi_j^P}{\partial \tau_j} = \left[p + \tau_j p'\frac{\partial Q}{\partial \tau_j}\right]Q_j + (\tau_j p - \kappa_N)\frac{\partial Q_j}{\partial \tau_j} = 0$, or equivalently, $(\tau_j p - \kappa_N) + \frac{\left[p+\tau_j p'\frac{\partial Q}{\partial \tau_j}\right]Q_j}{\frac{\partial Q_j}{\partial \tau_j}} = 0$, and substitute $\tau_j = 1 - \frac{m\kappa_M}{p'Q+mp}$ and $Q_j = \frac{Q}{n}$, then we have (20). To summarize, under proportional fees,

$$\begin{aligned}
\frac{\partial Q}{\partial \tau_j} &= \frac{[p'Q+mp]^2}{nm\kappa_M\left[p''Q+(m+1)p'\right]} \\
\frac{\partial Q_j}{\partial \tau_j} &= \frac{[p'Q+mp]^2}{nm\kappa_M\left[p''Q+(m+1)p'\right]}\frac{[(n-1)p''Q+(mn+n-m)p']}{np'} \\
\tau_j &= 1 - \frac{m\kappa_M}{p'Q+mp}
\end{aligned}$$

7.3 Proposition 5: Welfare on competitive networks

The decomposition of changes in equilibrium outputs:

$$\begin{aligned}
& sign\left\{Q^{PC} - Q^{UC}\right\} \\
= \; & sign\left\{\frac{\partial \Pi_j^P/\partial \tau_j}{\partial Q_j^U/\partial \tau_j} - \frac{\partial \Pi_j^U/\partial t_j}{\partial Q_j^U/\partial t_j}\right\} = -sign\left\{\frac{\partial \Pi_j^P}{\partial \tau_j} - \frac{\partial \Pi_j^U}{\partial t_j}\frac{\partial Q_j^U/\partial \tau_j}{\partial Q_j^U/\partial t_j}\right\} \\
= \; & -sign\left\{\frac{\partial \Pi_j^P}{\partial \tau_j} - \frac{\partial \Pi_j^U}{\partial t_j}\frac{\partial t_j}{\partial \tau_j}\right\} = -sign\left\{\frac{\partial\left[-\tau_j\frac{n}{m}p'Q_j^2\right]}{\partial \tau_j}\right\} \\
= \; & -sign\left\{-\beta_j\frac{n}{m}\left[p''Q_j\frac{\partial Q}{\partial \tau_j} + 2p'\frac{\partial Q_j}{\partial \tau_j}\right]Q_j - \frac{n}{m}p'Q_j^2\right\} \\
= \; & sign\left\{\underbrace{\beta_j(\rho+2)}_{\text{Expansion Effect }(+)} + \underbrace{\beta_j(\rho+2)\frac{n-1}{n}\left[1+\frac{2(m-1)}{\rho+2}\right]}_{\text{(Additional) Business-stealing Effect }(+)} + \underbrace{\left[-\left(1-\beta_j\right)\frac{\rho+m+1}{m\varepsilon-1}\right]}_{\text{Concession Effect }(-)}\right\}
\end{aligned}$$

The last equality is obtained by substituting $\frac{\partial Q}{\partial \tau_j}$ and $\frac{\partial Q_j}{\partial \tau_j}$ from (19) and (18) as functions of Q, utilizing the symmetry property (i.e., $Q_j = \frac{Q}{n}$), and finally eliminating a common negative term.[26]

[26] $\beta_j\frac{n}{m}\left[p''Q_j\frac{\partial Q}{\partial \tau_j} + 2p'\frac{\partial Q_j}{\partial \tau_j}\right]Q_j - \frac{n}{m}p'Q_j^2 = -\frac{p'Q\left[p'Q+mp\right]^2}{m^2n\kappa_M[p''Q+(m+1)p']} \times Y$, where Y is the expression inside the {} on the last line, and the common term $-\frac{p'Q\left[p'Q+mp\right]^2}{m^2n\kappa_M[p''Q+(m+1)p']} < 0$.

By comparing (17) with (20), we can verify that $Q^{PC} > Q^{UC}$ if and only if

$$\begin{aligned}\kappa_N &> -\frac{p(Q^{UC})}{[(n-1)(\rho+m+1)+1]} \\ &\quad\times\frac{n(\rho+m+1)\left\{\varepsilon\left[n\left(\rho+1\right)+m\left(n-1\right)\right]+1\right\}}{\varepsilon\left\{\left[\left(m\varepsilon-1\right)\left(2n-1\right)+n\right]\left(\rho+m+1\right)-(m-1)(m\varepsilon-1)\right\}}.\end{aligned}$$

Note that

$$\begin{aligned}(\rho+m+1) &> 0 \\ [(n-1)(\rho+m+1)+1] &> 0 \\ \varepsilon\left\{\left[\left(m\varepsilon-1\right)\left(2n-1\right)+n\right]\left(\rho+m+1\right)-(m-1)(m\varepsilon-1)\right\} &> 0\end{aligned}$$

Moreover, given $n \geq 2$, we have $\varepsilon\left[n\left(\rho+1\right)+m\left(n-1\right)\right]+1 = \varepsilon\left[n\left(\rho+2\right)+\left(m-1\right)\left(n-1\right)-1\right]+1 > 0$. The right hand side of (21) is always negative, meaning that it always holds, i.e., $Q^{PC} > Q^{UC}$ unconditionally.

7.4 Network competition: Joint profit and individual incentives

Consider linear demand $p = 1 - Q$. The first example demonstrates that proportional fees may raise the joint profit when $m = n = 2$ (and when merchants' cost is large):

Table 2: $m = n = 2$

	Unit fees				Proportional fees			
Profit	Network	Merchant	Joint	Price	Network	Merchant	Joint	Price
$\kappa_M = 0.2$; $\kappa_N = 0.2$	0.0267	0.0178	0.0444	0.733	0.0324	0.0118	0.0442	0.660
$\kappa_M = 0.4$; $\kappa_N = 0.2$	0.0119	0.0079	0.0198	0.822	0.0134	0.0065	0.0199	0.789

The next two examples demonstrate network's individual incentive to adopt proportional fees. Let $n = 2$ and $m = 4$. Then:

Table 3: Small κ_M

	Network's profit		Merchants' profit	
$\kappa_M = 0.01$; $\kappa_N = 0.2$	Unit fees	Proportional fees	Unit fees	Proportional fees
Unit fees	0.047; 0.047	0.018; 0.082	0.028; 0.028	0.01; 0.005
Proportional fees	0.082; 0.018	0.042; 0.042	0.005; 0.01	0.003; 0.003

Table 4: Large κ_M

	Network's profit		Merchants' profit	
$\kappa_M = 0.2$; $\kappa_N = 0.2$	Unit fees	Proportional fees	Unit fees	Proportional fees
Unit fees	0.027; 0.027	0.019; 0.037	0.016; 0.016	0.0117; 0.0123
Proportional fees	0.037; 0.019	0.028; 0.028	0.0123; 0.0117	0.009; 0.009

In both examples, proportional fee is a dominant strategy. It is Prisoners' Dilemma in Table 3 (when κ_M is small), but not Table 4 (when κ_M is large). Note that merchants' profit increases with their own cost when both networks charge proportional fees ($0.009 > 0.003$), and merchants are worse off when networks switch from unit fees to proportional fees ($0.003 < 0.028$, and $0.009 < 0.016$).

7.5 Network competition: Constant-elastic demand

Given $p(Q) = \alpha Q^{-\frac{1}{\varepsilon}}$ with $\varepsilon \in (1, +\infty)$. Let $h = (n-1)(m\varepsilon - 1) + \varepsilon$, then the equilibrium can be calculated as:[27]

$$
\begin{aligned}
Q^{UC} &= \left[\frac{(m\varepsilon-1)(h-n)}{hm\varepsilon(\kappa_M+\kappa_N)}\right]^{\varepsilon}, \\
\Pi_N^U &= \frac{n(\kappa_M+\kappa_N)}{h-n} Q^{UC}, \text{ and } \Pi_M^U = \frac{h(\kappa_M+\kappa_N)}{(m\varepsilon-1)(h-n)} Q^{UC}. \\
Q^{PC} &= \left[\frac{(m\varepsilon-1)(h-1)}{m\varepsilon[h+(n-1)]\kappa_M + h(m\varepsilon-1)\kappa_N}\right]^{\varepsilon}, \\
\Pi_N^P &= \frac{mn\varepsilon\kappa_M + (m\varepsilon-1)\kappa_N}{(m\varepsilon-1)(h-1)} Q^{PC}, \text{ and } \Pi_M^P = \frac{\kappa_M}{m\varepsilon-1} Q^{PC}.
\end{aligned}
$$

7.5.1 Unit fees: networks' and merchants' profits

Direct calculation leads to

$$\frac{\Pi_N^U}{\Pi_M^U} = \frac{n(m\varepsilon-1)}{(n-1)(m\varepsilon-1)+\varepsilon},$$

which is independent of κ_M or κ_N. The profit ratio decreases in n and increases in ε and m. In fact there is a "cost neutrality" here: the three major variables, Q^{UC}, Π_N^U, and Π_M^U, all depend on $\kappa_M + \kappa_N$ rather than the individual cost.[28]

7.5.2 Proportional fees: networks' and merchants' profits

The merchants' profit:

$$\Pi_M^P = \alpha^{\varepsilon} \frac{\kappa_M}{m\varepsilon-1} \left\{ \frac{(m\varepsilon-1)\left[(n-1)(m\varepsilon-1)+\varepsilon-1\right]}{m\varepsilon[(n-1)m+1]\kappa_M + \left[(n-1)(m\varepsilon-1)+\varepsilon\right](m\varepsilon-1)\kappa_N} \right\}^{\varepsilon},$$

is inversely U-shaped in κ_M, and is maximized at $\kappa_M^* = \frac{(m\varepsilon-1)[(n-1)(m\varepsilon-1)+\varepsilon]}{m\varepsilon^2(\varepsilon-1)[(n-1)m+1]}\kappa_N$. The example after Proposition 9 for monopoly network is a special case of this.

The networks' profit, Π_N^P, always decreases with κ_N. In addition,

$$\frac{\Pi_N^P}{\Pi_M^P} = \frac{1}{\varepsilon\left[(n-1)m+1\right]-n} \left[mn\varepsilon + (m\varepsilon-1)\frac{\kappa_N}{\kappa_M} \right].$$

[27] Shy and Wang (2011) becomes a special case with $n = 1$. Their conclusion is based on direct compartison of the expressions: $Q^P > Q^U$ when $\kappa_N > 0$, and $Q^P = Q^U$ when $\kappa_N = 0$.

[28] This is true for general demand: by substituting $t = \frac{1}{m}p'(Q)Q + p(Q) - \kappa_M$ in (monopoly or symmetric competition) equilibrium, $\frac{\Pi_N^U}{\Pi_M^U} = \frac{(t-\kappa_N)Q}{(p-t-\kappa_M)Q} = \frac{t-\kappa_N}{p-t-\kappa_M} = \frac{\frac{1}{m}p'(Q)Q+p(Q)-\kappa_M-\kappa_N}{-\frac{1}{m}p'(Q)Q}$. Obviously, their profits depend on $\kappa_M+\kappa_N$ rather than the individual cost, which means the equilibrium Q depends only on $\kappa_M+\kappa_N$.

Note that $\frac{\Pi_N^P}{\Pi_M^P}$ increases with $\frac{\kappa_N}{\kappa_M}$: for both parties (merchants and networks), a party's profit relative to the other party's increases with its own cost and decreases with the other party's cost. $\frac{\Pi_N^P}{\Pi_M^P}$ increases with m, and decreases with n and ε.

7.5.3 Networks' profits under the two fees

$$\begin{aligned}\frac{\Pi_N^U}{\Pi_N^P} &= \frac{n(m\varepsilon-1)(\kappa_M+\kappa_N)}{mn\varepsilon\kappa_M+(m\varepsilon-1)\kappa_N}\frac{(h-1)}{(h-n)}\left[\frac{h-n}{h-1}\frac{m\varepsilon[h+(n-1)]\kappa_M+h(m\varepsilon-1)\kappa_N}{hm\varepsilon(\kappa_M+\kappa_N)}\right]^{\varepsilon} \\ &= Z\left[\frac{h-n}{h-1}\frac{m\varepsilon[h+(n-1)]\kappa_M+h(m\varepsilon-1)\kappa_N}{hm\varepsilon(\kappa_M+\kappa_N)}\right]^{\varepsilon-1},\end{aligned}$$

where

$$Z=\frac{n(m\varepsilon-1)}{hm\varepsilon}\frac{m\varepsilon[h+(n-1)]\kappa_M+h(m\varepsilon-1)\kappa_N}{mn\varepsilon\kappa_M+(m\varepsilon-1)\kappa_N}.$$

Since $Q^{PC}>Q^{UC}$, the term in the square brackets is less than 1. Then, a sufficient (but not necessary) condition for $\Pi_N^P>\Pi_N^U$ is $Z<1$, which is equivalent to (22).

Note that $\frac{(m\varepsilon-1)[(n-1)(m\varepsilon-1)-1][(n-1)(m\varepsilon-1)+\varepsilon]}{mn\varepsilon^2}$ increases with ε. If

$$\lim_{\varepsilon\to 1}\frac{mn\varepsilon^2}{(m\varepsilon-1)\left[(n-1)(m\varepsilon-1)-1\right]\left[(n-1)(m\varepsilon-1)+\varepsilon\right]}<\frac{\kappa_N}{\kappa_M},$$

i.e., (23), then $\lim_{\varepsilon\to 1}Z>1$. Since the term in the square brackets approaches 1 as $\varepsilon\to 1$ ($\lim_{\varepsilon\to 1}\left[\frac{h-n}{h-1}\frac{m\varepsilon[h+(n-1)]\kappa_M+h(m\varepsilon-1)\kappa_N}{hm\varepsilon(\kappa_M+\kappa_N)}\right]^{\varepsilon-1}=1$), we have $\lim_{\varepsilon\to 1}\frac{\Pi_N^U}{\Pi_N^P}>1$. That is, as long as condition (23) holds, there is always a range of ε close to 1 under which $\Pi_N^P<\Pi_N^U$.

7.5.4 Conditions for proportional fees to raise network profits

Assume constant elasticity demand with $\varepsilon=2$. Then $\Pi_N^P>\Pi_N^U$ if and only if (i.e., a sufficient and necessary condition)

$$\frac{\kappa_N}{\kappa_M}<\widehat{\kappa}(m,n),$$

where $\widehat{\kappa}(m,n)$ is a function of m and n. As Figure 3 shows, the cutoff $\widehat{\kappa}$ (for $\frac{\kappa_N}{\kappa_M}$) decreases with m and n, which means that $\Pi_N^P>\Pi_N^U$ is more likely if κ_M is larger, or κ_N, m, or n is smaller.

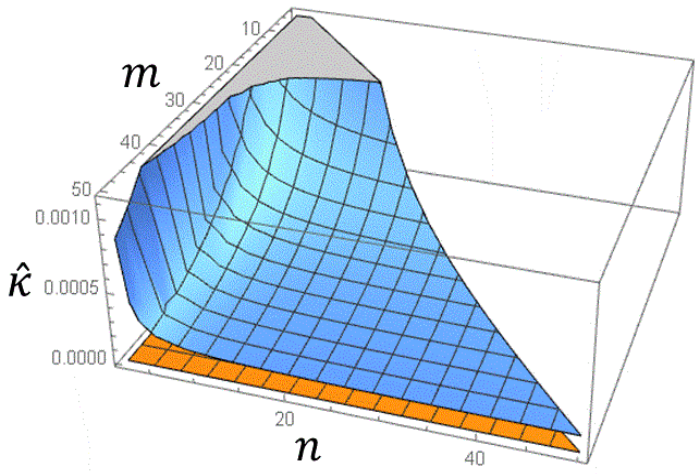

Figure 3: The cutoff $\widehat{\kappa}$ (for $\frac{\kappa_N}{\kappa_M}$) as a function of m and n when $\varepsilon = 2$

References

[1] Anderson, Simon P., Andre de Palma, and Brent Kreider. 2001. "The efficiency of indirect taxes under imperfect competition." *Journal of Public Economics* 81(2): 231–251.

[2] Armstrong, Mark. 2006. "Competition in two-sided markets." *RAND Journal of Economics* 37 (3): 668–91.

[3] Baxter, William F. 1983. "Bank Interchange of Transactional Paper: Legal and Economic Perspectives." *Journal of Law and Economics* 26 (3): 541–88.

[4] Bedre-Defolie, Ozlem and Emilio Calvano. 2013 "Pricing Payment Cards." *American Economic Journal: Microeconomics* 5(3): 206–231.

[5] Besen, Stanley M., and Joseph Farrell. 1994. "Choosing How to Compete: Strategies and Tactics in Standardization." *Journal of Economic Perspectives* 8(2): 117–31.

[6] Bishop, Robert L. 1968. "The effects of specific and ad valorem taxes." *The Quarterly Journal of Economics* 82(2): 198–218.

[7] Bousquet, Alain, Helmuth Cremer, Marc Ivaldi, and Michel Wolkowicz. 1998. "Risk sharing in licensing." *International Journal of Industrial Organization* 16(5): 535–554.

[8] Delipalla, Sofia, and Michael Keen. 1992. "The comparison between ad valorem and specific taxation under imperfect competition." *Journal of Public Economics*, 49(3): 351–367.

[9] Farrell, Joseph, and Garth Saloner. 1986. "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation." *American Economic Review* 76(5): 940–55.

[10] Federal Reserve: The Federal Reserve Payments Study: 2017 Annual Supplement.

[11] Gaudin, Germain, and Alexander White. 2014. "On the Antitrust Economics of the Electronic Books Industry." mimeo.

[12] Getter, Darryl E. 2017. "Regulation of Debit Interchange Fees." *Congress Research Service Report 7-5700.*

[13] Gu, Dingwei, Zhiyong Yao, Wen Zhou, and Rangrang Bai. 2018. "How Can Downstream Firms Benefit from Upstream Collusion?" mimeo.

[14] Gu, Dingwei, Zhiyong Yao, Wen Zhou, and Rangrang Bai. forthcoming. "When is Upstream Collusion Profitable?." *RAND Journal of Economics.*

[15] Hagiu, Andrei., and Julian Wright. 2015. "Marketplace or reseller?." *Management Science* 61(1): 184–203.

[16] Katz, Michael L., and Carl Shapiro. 1985. "Network Externalities, Competition, and Compatibility." *American Economic Review* 75(3): 424–40.

[17] Katz, Michael L., and Carl Shapiro. 1994. "System Competition and Network Effect." *Journal of Economic Perspectives* 8(2): 93–115.

[18] Kokkoris, Ioannis. 2013. "Expedia and Booking.Com: Agent or Distributor." *Competition Policy International.*

[19] Johnson, Justin P. 2017. "The Agency Model and MFN Clauses." *Review of Economic Studies* 84: 1151–1185.

[20] Llobet, Gerard, and Jorge Padilla. 2016, "The Optimal Scope of the Royalty Base in Patent Licensing." *Journal of Law and Economics* 59: 45–73

[21] Manuszak, Mark D., and Krzysztof Wozniak. 2017. "The Impact of Price Controls in Two-Sided Markets: Evidence from US Debit Card Interchange Fee Regulation." FEDS Working Paper No. 2017-074. Available at SSRN: https://ssrn.com/abstract=2999628 or http://dx.doi.org/10.17016/FEDS.2017.074

[22] Mrázová, Monika, and J. Peter Neary. 2013. "Selection effects with heterogeneous firms." *Journal of the European Economic Association.* 1–41.

[23] Mrázová, Monika, and J. Peter Neary. 2017. "Not So Demanding: Demand Structure and Firm Behavior." *American Economic Review* 107(12): 3835–3874.

[24] Oates, Wallace E. 1985. "Searching for Leviathan: an empirical analysis." *American Economic Review* 75, 748–757.

[25] Rauscher, Michael, 1998. "Leviathan and competition among jurisdictions: the case of benefit taxation." *Journal of Urban Economics* 44, 59–67.

[26] Rochet, Jean-Charles, and Jean Tirole. 2002. "Cooperation among competitors: some economics of payment card associations." *RAND Journal of Economics* 33 (4): 549–70.

[27] Rochet, Jean-Charles, and Jean Tirole. 2003. "Platform Competition In Two-Sided Markets." *Journal of the European Economic Association* 1 (4): 990–1029.

[28] Rochet, Jean-Charles, and Jean Tirole. 2006. "Two-sided markets: a progress report." *RAND Journal of Economics* 37 (3): 645–67.

[29] Rysman, Marc. 2009. "The Economics of Two-Sided Markets." *Journal of Economic Perspectives* 23 (3): 125–43.

[30] Suits, D. B., Musgrave, R. A. 1953. "Ad valorem and unit taxes compared." *The Quarterly Journal of Economics* 67(4): 598–604.

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

[32] The Nilson Report: Merchant Fees Report.

[33] U.S. Supreme Court. 2015. "Apple Inc. Petitioner versus United States et al." No. 15-565.

[34] Wang, Zhu. 2010. "Market structure and payment card pricing: What drives the interchange?" *International Journal of Industrial Organization* 28 (1): 86–98.

[35] Wang, Zhu, and Julian Wright. 2017, "Ad valorem platform fees, indirect taxes, and efficient price discrimination." *RAND Journal of Economics* 48(2): 467–484.

[36] Weyl, E. Glen. 2010. "A Price Theory of Multi-Sided Platforms." *American Economic Review* 100 (4):1642–72.

[37] Wilson, John D. 1986. "A theory of interregional tax competition." *Journal of Urban Economics* 19, 296–315.

[38] Wilson, John D. 1999. "Theories of tax competition." *National Tax Journal* 52, 269–304.

[39] Wilson, John D., and David E. Wildasin. 2004. "Capital tax competition: bane or boon." *Journal of Public Economics* 88: 1065–1091.

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

[41] Wright, Julian. 2013. "Why payment cards fees are biased against retailers." *RAND Journal of Economics* 43 (4): 761–80.

[42] Zodrow, George R. and Peter Mieszkowski. 1986. "Pigou, Tiebout, property taxation, and the underprovision of local public goods." *Journal of Urban Economics* 19, 356–370.