Preliminary version: Please do not circulate*

Competition between payment card networks and rebates†

Vera Lubbersen‡ Wilko Bolt§

February 14, 2025

Abstract

Visa and Mastercard are among the most profitable companies in the world. Despite stricter regulatory scrutiny, their profit margins have only been steadily increasing over the years. Many models on payment competition focus on the interchange fee, while this paper shows that fixed card rebates to card issuers could also play an important role. We have adapted existing platform models to a competitive market for payment cards. It is shown how price distortions evolve as transaction benefits increase. In a monopolistic market, profit-maximizing card rebates are too low compared to the socially optimal level. For low transaction benefits, this is especially the case on the side that shows higher card benefits. For higher transaction benefits, the opposite is true. The effects of card network competition are determined by the "homing structure". Price distortions are highest on the side that uses more than one card network.

Key Words: Rebates; Platform Competition; Payment fees; Card networks; Price distortion
JEL Classification: L12; L13; L14; L20; L21

*This version is still incomplete and preliminary, please do not circulate.

†The views expressed are those of the authors and do not necessarily reflect official positions of De Nederlandsche Bank or of the Eurosystem.

‡Vrije Universiteit Amsterdam and De Nederlandsche Bank, v.r.lubbersen@dnb.nl

§Vrije Universiteit Amsterdam and De Nederlandsche Bank, w.bolt@dnb.nl

I. Introduction

The question addressed in this paper is why card networks give upfront rebates - also called incentive payments - to card-issuing banks.[1] Rebates are usually negotiated between card networks and issuing banks when a bank starts issuing a new type of payment card. These rebates largely consist of cash discounts, but may also include other perks and products (such as additional consultancy services). We analyse the incentives to offer card rebates in a monopoly and duopoly two-sided market structure.

To our knowledge, no one has looked yet at the strategic impact of card rebates on transaction fees and network profit.[2] Figure 1 shows the flow of payment fees in a so-called four-party payment network. As has been well documented, card networks set two types of transaction fees. First, they set the per-transaction interchange fee (IF) that the acquiring banks pay to the issuing banks, and second, they determine the processing and scheme fees that acquiring and issuing banks pay to the card network when a card payment is executed. Given these fees, the issuing and acquiring banks then subsequently determine the cardholder fee and the merchant service charge (MSC) respectively. Most theoretical work focuses on the determination of the optimal IF and its impact on card usage and network profit.

Previous papers often do not make a distinction between card or transaction benefits. Card benefits are the added value consumers and merchants assign to paying or be paid by card compared to cash. Transaction benefits are obtained when the merchant side and the consumer side transact with each other. In payments, transaction benefits are high, as merchant acceptance is key to consumer adoption and vice versa. At the same time, some users prefer cash, which means that their card benefit is heterogeneous and may well be

[1] Both VISA and Mastercard spend about 10 billion US dollar as "upfront and fixed incentive payments" per year, which is about 25 percent of Visa's gross revenues and 30 percent of Mastercard's gross revenues. Interestingly, these rebates are similar in magnitude, while Mastercard's gross revenues are about 75 percent of Visa's gross revenues.

[2] Schwartz and Vincent (2006, 2020) analyse the interplay between rebates and demand in the product market under the "No Surcharge Rule" (see Section V). However, in their analysis "rebates" are equivalent to analysing the IF.

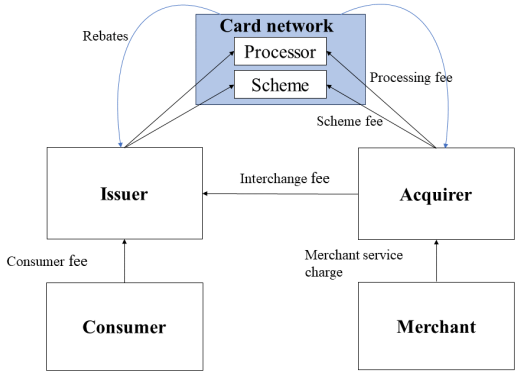


Figure 1: Four-corner model flow of fees

negative. Existing models simplify the analysis by assuming heterogeneity in transaction benefits without card benefits. In such a model, the IF is "the only mechanism through which [card] associations can perform the balancing act" (Rochet and Tirole, 2003a). The IF lowers the total price paid by the consumer side, increasing card usage. These papers have shown that the interchange fee may be too high compared to the socially optimal level leading to 'overusage' of cards. Moreover, previous papers show that competition between card networks may yield "excessive" transaction discounts and therefore do not lead to a welfare increasing outcome.

In contrast to other papers, we show that if a monopolistic card network sets fees based on heterogeneous card benefits and homogeneous transaction benefits, it seems more likely that there exists "underusage" of cards. In our paper, a card network sets transaction fees paid by the issuing and acquiring banks per transaction and card fees paid up front per card issued or card acceptance terminal installed. We define rebates as a negative card fee that reduces the total payment of the issuing and acquiring banks to the card network. Card rebates could be interpreted as a "reduced" mark-up on transaction fees, and may be too low compared to the socially optimal level. The reason for such 'underusage' of cards is because a card network is unable to steer usage. Contracts are negotiated between card networks and banks, based on average card usage per consumer rather than individually

negotiated with end-users. In effect, card networks cannot offer direct transaction discounts to consumers to stimulate card usage since intermediaries "sit in between". The best a card network can do, is to offer rebates to banks to maximize card usage/acceptance given average card usage/card acceptance per bank customer. Our model fits the debit card market where consumers rarely receive anything for making a card payment and rebates are mostly offered to consumer banks when they provide card services to customers.

We add to the literature by examining how this price distortion evolves as transaction benefits increase in both a monopolistic and duopolistic market. In general, homogeneity reduces price distortions, while card benefits have an ambiguous effect. If one of the sides has higher card benefits, this makes most users join even if they receive lower rebates. A "non-distortional shift" arises from user surplus to card network profit. Therefore, for larger transaction benefits, the price distortion is higher on the side that shows lower card benefits. In contrast, in a card network competition context, card networks compete for "singlehoming" consumers with high rebates. This competition effect increases over transaction benefits. Therefore, if card networks compete, the consumer side is "cornered" first which reduces the price distortion on the consumer side.

The rest of this paper is structured as follows. Section II gives an overview of existing literature and links our paper to previous work. Section III introduces the basic monopoly model and explains what determines the level of rebates on the consumer and merchant side. Section IV makes an extension to a duopolistic model where two card networks simultaneously compete for "singlehoming" consumers and "multihoming" merchants. Section V provides a discussion of our critical assumptions and scope for further research. Section VI concludes with some (policy) implications. Formal proofs can be retrieved upon request.

II. Literature Review

The modern payment literature effectively starts with Baxter (1983), who provides the backbone for later economic models centered around the IF. Baxter's analysis goes as follows. A consumer would only be willing to pay by card if the usage fee is lower than the benefit of paying by card compared to cash. Now suppose that the issuing costs are high, while at the same time the joint benefit of the consumer and the merchant exceeds the total resource cost of a card payment. Consequently, the consumer fails to pay by card though it would be socially optimal to do so. Effectively, the consumer exerts a negative usage externality on the merchant which can be internalized by an appropriate level of the IF.

There are some issues with Baxter's simplified analysis.[3] Firstly, Baxter (1983) assumes perfect competition between banks. Later papers show that if issuing banks have market power, then the profit-maximizing IF may be set above the socially optimal level (Bedre-Defolie and Calvano, 2013; Rochet and Tirole, 2011; Wright, 2004). Secondly, consumers and merchants are not homogeneous, as assumed in Baxter's analysis, but heterogeneous. If consumers show more heterogeneity regarding card payments than merchants, then consumer demand is likely to be more sensitive to price variation. The card network may therefore select a higher than socially optimal IF (Wright, 2004). In the extreme case where consumers are heterogeneous and merchants homogeneous, card networks choose the highest IF such that merchants still accept card payments (Rochet and Tirole, 2002). Thirdly, strategic interactions between merchants are not modelled in the original Baxter (1983) analysis. Merchants may pay a higher MSC than their direct card benefit due to competition between merchants. Merchants internalize the card usage externality of consumers (Rochet and Tirole, 2011), which gives rise to "weak merchant resistance" and a higher than socially optimal IF. Fourthly, Baxter (1983) does not take into account the effect of card network competition on the IF. The effect of card network competition depends on what side compe-

[3]See Verdier (2011) for a more extensive literature review.

tition is strongest (Guthrie and Wright, 2007). Lastly, Baxter's analysis does not allow card and usage decisions to be separated. Bedre-Defolie and Calvano (2013) contribute to the payment literature by showing that two-part pricing by card networks could be an optimal strategy. After card membership decisions are made. Usage is determined by consumers. This also results in a higher than socially optimal IF.

Our paper argues there is yet another issue with Baxter's analysis. Baxter (1983) does not differentiate between card and transaction benefits. Others have argued that a model based on heterogeneous card benefits and homogeneous transaction benefits is not suitable for payments. [4] In such a model homogeneous transaction benefits can be incorporated into fixed card fees. Thus, transaction fees play an "irrelevant" role and may be normalized to zero (Armstrong, 2006; Rochet and Tirole, 2006). However, we argue that this is precisely what makes the "Armstrong" model insightful in a debit card context. First, one may argue that users differ more with respect to their preference for cash versus card than their preference for the size of the network. Second, one may argue that card networks are unable to steer card usage since banks "sit in between". Card networks can only steer the amount of cardholders and accepted merchants through contracts with banks. Third, this model allows us to analyse what happens as transaction benefits increase and the card market saturates towards full market coverage. Suppose transaction benefits are high, while some users prefer cash to card. We show in a model with low heterogeneous card benefits and high homogeneous transaction benefits that the IF is irrelevant and rebates are lower than socially optimal. To the best of our knowledge, no one has considered the difference between card and transaction benefits yet. This paper aims to fill this gap.

[4] See the question of Jean-Charles Rochet of the paper presentation "CBDC and Payment Platform Competition" by Edona Reshidi at the Economics of Payments XIII Conference (2024).

III. Monopoly model

In this section, we define rebates and describe how they emerge in a monopoly setting. We briefly review the monopoly outcome of existing models as a benchmark and explain the basic Armstrong (2006) model on which we build. We modify the model to allow for transaction fees and rebates. In the equilibrium outcome, we show that the profit-maximizing rebate is always lower than (or equal to) the socially optimal rebate. Then, we provide the determinants of this price distortion in rebates. We end with a short numerical example before we analyze how card network incentives change with network competition in section IV.

Basic model

The monopolistic card network enables transactions between the consumer side (c) and the merchant side (m) of the market, indexed by $i = c, m$. Each market side is populated by a unit-mass continuum of agents. Each agent derives utility from the card services offered by the network, but also from transacting with agents from the other side of the market who also join the monopolistic card network. [5]

Formally, each side-i agent, $i = c, m$, has a type $\omega_i = (B_i, \alpha_i) \in \Omega_i \subset \mathbb{R}^2$ and by joining the card network derives a gross payoff

$$u_i(\omega_i, n_j) = B_i + \alpha_i n_j \tag{1}$$

from transacting with a mass of agents of size n_j from side j, $j \neq i$.[6] Each agent who does not join the card network obtains its outside option with a payoff normalized to zero

[5] For tractability reasons, we do not model issuers and acquirers directly. This assumption could be confusing, since we are focusing on different price formation between issuers/acquirers and the payment card network. However, a three-party payment network without issuers and acquirers is "mathematically" equivalent to a four-party scheme with a monopolistic issuer and perfectly competitive acquirers. The issuing bank *de facto* controls the network.

[6] We follow model exposition here by Jullien et al. (2021).

– i.e. "paying by cash". Each type ω_i is an independent draw from some distribution G_i and is the agent's private information.[7] The component B_i in the agent's payoff denotes the utility that the agent derives from joining card network and receiving its services, whereas the component α_i denotes the (positive) network effect from transacting with agents from the other side of the market. We refer to B_i as the agent's "card benefit" and to the term α_i as the agent's "transaction benefit".

The gross payoffs in (1) are net of any total payment P_i that the agents make to (or receive from) the card network. Assuming that preferences are quasilinear in payments, the net payoff is equal to $u_i(\omega_i, n_j) - P_i$. The total payment P_i paid by an agent on side i, $i = c, m$, to the card network usually has two components, a transaction fee f_i charged for every transaction with agents from side $j \neq i$ and a card fee $F_i > 0$ paid to the network per cardholder/merchant or a card rebate $R_i > 0$ received from the network as a subsidy per cardholder/merchant.[8] Obviously, $-F_i = R_i$, so when the fixed fee F_i turns negative, it can be seen as a card rebate that is received from the card network (see Definition 1). Assuming that the number of transactions is proportional to the mass of agents on side j, we have that the total payment by each agent joining from side i is equal to $P_i = f_i n_j - R_i$.

Definition 1 *A fixed rebate $R_i > 0$ is a negative card fee that reduces the total payment of an agent on side i, $i = c, m$, to the card network.*

Given the measure of side-j agents n_j that joins the card network, we may then derive demand from side i as the probability that the gross payoff in (1) is higher than the total payment P_i:

$$n_i = D_i(P_i, n_j) = Pr\{\omega_i \in \Omega_i : B_i + \alpha_i n_j \geq P_i\}. \tag{2}$$

The measure of agents n_c and n_m from the two sides of the market is the solution to the

[7] In the rest of this paper G_i refers to the cumulative distribution function (CDF) and g_i refers to the probability density function (PDF).

[8] The assumption of a fixed rebate (independent of transaction volume) might be (partly) justified on the basis that four-party card networks are unable to set transaction rebates to consumers since issuing banks intermediate the cards market. We discuss the implication of this assumption further in section V.

two "demand equations" $n_i = D_i(P_i, n_j)$, $i, j = c, m$, $i \neq j$. This yields $n_c = D_c(P_c, P_m)$ and $n_m = D_m(P_c, P_m)$. These equations reflect that, from side-i's perspective, n_j can be seen as a "quality" dimension affecting side-i's demand that is determined endogenously through prices (Jullien et al., 2021). Total usage demand, or payment usage, is given by $D(P_c, P_m) = D_c(P_c, P_m) \times D_m(P_c, P_m)$.

In this paper, we use the "Armstrong" (2006) model where users on the same side may differ with respect to B_i (preference for cards versus cash), but all users on the same side assign the same value to α_i (preference for network size). [9] At the same time, we allow the card network to set both transaction fees and card fees/rebates (see Assumption 1). Consequently, G_i (and its PDF g_i) is one-dimensional and is only defined on B_i (not on α_i).

Assumption 1 *Users have heterogeneous card benefits B_i (preference for cards versus cash) which is the agent's private information and drawn from some distribution G_i and homogeneous transaction benefits α_i (preference for size of the network) which is identical for all agents on side i.*

Under Assumption 1, transaction fees can be set equal to transaction benefits w.l.o.g. [10] Transaction fees could be seen as a perfect "insulated tariff". Insulated tariffs control for network effects as they eliminate any transaction benefits received from each participating user on the other side (Jullien et al., 2021; Weyl, 2010; White and Weyl, 2016). When the card network charges transaction fees, all participation externalities are eliminated, $u_i(\omega_i, n_j) - T_i = B_i + R_i$, which is a fixed payoff independent of the opposite-side agents' actions.

This also means that the demand function becomes one-dimensional. To see this, note

[9] Our assumption of homogeneous transaction benefits allows us to analyze different equilibrium solutions on different levels of transaction benefits. Allowing for heterogeneous transaction benefits would make our model mathematically more complex to solve, but would not drastically change our results, see Section V.

[10] The crucial difference between transaction fees and card fees is that cross-group externalities are weaker for the former when users pay for the benefit of transacting with the other side. If a user only needs to pay for a transaction once the transaction occurs, it does not matter how many users are on the other side. However, as shown in Armstrong (2006) and Rochet and Tirole (2006), the distinction only matters in a platform competition model.

that if the card network charges transaction fees $f_i = \alpha_i$ and card fees F_i, that the transaction benefits no longer play a role in the demand function. The transaction fee has no impact on the merchant acceptance or consumer adoption decision as the transaction fee is exactly equal to the transaction benefit, i.e. demand becomes:

$$n_i = 1 - G_i(f_i n_j - \alpha_i n_j + R_i) = 1 - G_i(R_i) = 1 - \int_{-\infty}^{-R_i} g_i(R_i) dR_i. \tag{3}$$

In this case, the inverse price elasticity of demand becomes one-dimensional and only defined over F_i. Rochet and Tirole (2003b) call these the inverse elasticities of quasi-demands:

$$\eta_i(R_i) = \frac{D_i(R_i)}{\partial D_i(R_i)/\partial(R_i)} = -\frac{R_i}{\epsilon_i(R_i)} = -\frac{1 - G_i(R_i)}{g_i(R_i)} \tag{4}$$

where $\epsilon_i(R_i)$ denotes the standard side-i price elasticity of quasi-demand with respect to its own price. When we add the price elasticities of quasi-demand on both sides together, we get the total mark-up (inverse price elasticity of demand): $\eta(R_i, R_j) = \eta_i(R_i) + \eta_j(R_j)$. The inverse elasticities of quasi-demands could in case of rebates be interpreted as a *negative* mark-up reduced from total transaction fees earned per additional cardholder/merchant.

The card network's profits are specified as:

$$\Pi = (f_c D_m(R_m) - R_c - C_c) D_c(R_c) + (f_m D_c(R_c) - R_m - C_m) D_m(R_m), \tag{5}$$

where C_i denotes the cost the card network incurs for each side-i agent it brings on board. [11]

Equilibrium outcome

Proposition (1) specifies the FOC's that characterizes the profit maximizing equilibrium with rebates. The incentive to offer rebates on side i increases over total transaction fees that can be earned by an additional side i user, decreases in the cost incurred to bring a side i user on

[11] Transaction costs incurred for each transaction between the two sides are w.l.o.g. normalized to 0, such that the homogeneous transaction benefit should be interpreted as a benefit net of cost.

board, and decreases in the mark-up earned in transaction fees on side i. In our model, costs are incurred by the card network. If the cost would be incurred by the issuing/acquiring banks, then the rebate would increase in costs. The price structure of the transaction fees and/or the difference in transaction benefits is irrelevant, because transaction fees on side i are equal to transaction benefits on side i and both sides need to be on board for a transaction to take place.

If total transaction benefits are high enough ($\alpha_t = \alpha_c + \alpha_m$), a card network may want to set rebates at a level such that all consumers have cards and/or all merchants accept cards. In this case, rebates are set at the boundary of the distribution of card benefits on side i such that all users on side i are on board, i.e. $B_i + R_i > 0$ for all users on side i: $R_i^{Max} = -\min\{B_i^{Min}, 0\}$. [12] In other words, this is the rebate that is necessary to convince the last user on side i with the lowest card benefit B_i^{Min}. As soon as the card network has convinced all users on side i to join, for a monopolistic card network, there is no need to increase rebates further. It is easy to see that if the profit-maximizing rebate is chosen such that $R_i^* = R_i^{Max}$, then $\eta_i(R_i)$ must be at its maximum at this level of rebates (η_i^{Max}). If the total transaction benefits increase further, $\eta_i(R_i)$ becomes undefined and a boundary solution emerges.

Proposition 1 *A profit maximizing monopolistic card network sets rebates R_i^* on side i as follows:*

- *If*

$$\alpha_t < \frac{R_i^{Max} + C_i + \eta_i^{Max}}{n_j^*}, \quad i \neq j,$$

$$R_i^* = (f_i + f_j) n_j^*(R_j^*) - C_i - \eta_i(R_i^*), \quad i \neq j,$$

- *If*

$$\alpha_t \geq \frac{R_i^{Max} + C_i + \eta_i^{Max}}{n_j^*}, \quad i \neq j,$$

[12] For an unbounded distribution B_i^{Min} is where $P(R_i^{Max} < -B_i^{Min}) \leq \epsilon$ for ϵ sufficiently small.

$$R_i^* = R_i^{Max},$$

where $\alpha_t = \alpha_c + \alpha_m$ *is the total level of transaction benefits,* n_j^* *is the profit maximizing level of side* j *demand, and* η_i^{Max} *is the inverse elasticity of side* i *quasi-demand at its maximal level of rebates* R_i^{Max} *necessary to convince (almost) all side* i *users.*

Proof. For the interior solution, the FOC on side i is:

$$\frac{\partial \Pi}{\partial R_c} = D_c(R_c) + \left(f_c D_m(R_m) - R_c - C_c)\right) \frac{dD(R_c)}{dR_c} + f_m D_m(R_m) \frac{dD(R_c)}{dR_c} = 0.$$

Dividing by $\frac{dD(R_c)}{dR_c}$ and using $T_c = f_c n_m^* - R_c^*$ we have the FOC stated in Proposition 1. The boundary solution can be verified by filling the threshold into the FOC and noticing that in this case $R_i^* = R_i^{Max}$. ■

Consider next a social planner maximizing the sum of consumer and merchant utilities and the platform's profit. The socially optimal participation level depends on how social welfare is defined. We take the economic definition consisting of consumer surplus, merchant surplus and card network profit. For any vector of rebates ($P_i = f_i n_j - R_i$), total user surplus (V) is given by both merchant and consumer utility:

$$V(R_i, R_j) = \sum_{i=c,m} V_i(R_i) = \sum_{i=c,m} \int_{R_i}^{R_i^{Max}} D_i(R_i) dR_i,$$

where $i \neq j$. Total welfare is given by:

$$W = V + \Pi, \tag{6}$$

which results in the FOC's specified in Proposition (2) that characterizes the socially optimal equilibrium with rebates.

In a socially optimal solution, a payment card network gives back rebates to side i equal

to the total transaction fees earned by a marginal user on side i minus the marginal cost it faces of servicing an additional user on side i. As with privately optimal fees, boundary solutions emerge as transaction benefits increase. In fact, these emerge much faster than in the case of privately optimal fees since there is no profit mark-up. If the total transaction benefits per side i user are higher than the maximal rebate and costs necessary to convince the last side i user, the rebate is set at the maximal level. In this case, transaction fees only shift from consumer/merchant surplus to the profit of the card network, but do not lead to any welfare distortion. In this boundary solution, participation levels are already at their socially optimal level and the socially optimal rebate can only be defined as a "minimum" rebate. A higher rebate increases card network's profits, but decreases user surplus by the same amount and vice versa. On the side with full-market coverage, there is no market distortion in the classical sense, but only a transfer from consumer surplus to card network profit.

Proposition 2 *A public planner with full control over the card network sets rebates R_i^w on side i as follows:*

- *If*

$$\alpha_t < \frac{R_i^{Max} + C_i}{n_j^w}, \quad i \neq j,$$

$$R_i^w = (f_i + f_j) n_j^w(R_j^w) - C_i, \quad i \neq j,$$

- *If*

$$\alpha_t \geq \frac{R_i^{Max} + C_i}{n_j^w}, \quad i \neq j,$$

$$R_i^w \geq R_i^{Max},$$

where $\alpha_t = \alpha_c + \alpha_m$ is the total level of transaction benefits, n_j^w is the welfare maximizing level of side j demand, and R_i^{Max} is the maximal level of rebates necessary to convince (almost) all side i users.

Proof. Taking the derivative with respect to R_i and using the fact that the derivative of V_i with respect to R_i is equal to N_i, we have the FOC stated in Proposition 2. ■

If we deduct the privately optimal price from the socially optimal rebate, we retrieve the total price distortion. The total price distortion is:

$$R_i^w - R_i^* = \underbrace{\alpha_t(n_j^w - n_j^*)}_{\text{Scale 'distortion'}} + \underbrace{\eta_i(R_i)}_{\text{Mark-up 'distortion'}} \underbrace{- \max\{0, \alpha_t n_j^w - C_i - R_i^{Max}\})}_{\text{Non-distortional shift}}$$

Previous papers argued that the total price distortion is higher in a two-sided market case than in a normal one-sided market, and the size of this distortion increases over transaction benefits. The classical mark-up distortion increases by what Tan and Wright (2021) refers to as the scale distortion.[13] This scale distortion emerges since the number of participants on the other side is lower in the monopolistic outcome compared to the socially optimal outcome. For example, fewer merchants accept cards than is socially optimal, which indirectly reduces the utility on the consumer side. The scale distortion reinforces the markup distortion if the transaction benefit on the other side is higher than the marginal transaction cost, which in the case of payments we may assume to be the case.[14]

However, we find that as transaction benefits increase, there is also a point where price distortions decrease due to a non-distortional shift from user surplus to card network profit. Traditionally a distortion arises in a certain market, as demand is lower than if the price were equal to marginal cost. In the case of network goods such as payment cards, demand effects are reduced by an increase in transaction benefits as the size of the network increases. Even if the total price is well above marginal cost, if on the other side all users have/accept this payment card, most users still want the payment card for the transactional benefits they provide. This is because card usage does not depend on the level of rebates directly, but

[13] The mark-up distortion does not contain the Spence and displacement distortion since transaction benefits are homogeneous and therefore there is no difference between average and marginal users' transaction benefits.

[14] In our model we normalize marginal transaction cost to 0, so transaction benefits should be interpreted as a net transaction benefit.

only indirectly through more users. Thus, there is no direct price distortion in the usage of payment cards, but only in adoption. The socially optimal rebate is set at a level such that all merchants and/or consumers participate but does not increase further above this level even though transaction revenues increase.

Determinants of price structure rebates

The price structure of the profit-maximizing rebates and the evolution of price distortions between consumers and merchants is determined by the distribution of card benefits. The distribution of card benefits between consumers and merchants can vary in two major ways: (1) the total value that consumers/merchants derive from cards compared to cash and (2) the uniformity of preferences of cards compared to cash. In what follows, we assume a uniform distribution for both consumers and merchants, $B_i \sim \text{Uniform}(a_i, b_i)$, which we may try to generalize in a later stage.

First, suppose that merchants have higher card benefits than consumers, but for now assume the variance is similar (see below the effect of a change in variance) and costs are also the same on both sides. In other words, the distribution of card benefits for merchants first-order stochastically dominates the distribution of card benefits for consumers. This has two major implications for our equilibrium solutions. First of all, a profit-maximizing card network gives fewer rebates to merchants than to consumers. The same holds for a benevolent planner, but this difference between the two sides for the socially optimal rebates is smaller for low transaction benefits. For higher transaction benefits, the benevolent planner would like to provide a rebate such that all merchants accept cards, because it also follows that $R_m^{Max} < R_c^{Max}$. In other words, the merchant side is "cornered" first. It is easier to convince the last merchant to join than it is to convince the last consumer to join. For higher transaction benefits, the non-distortional shift reduces the price distortion on the merchant side. This result is summarized in Proposition 3.

Proposition 3 *Suppose that merchants value cards more than consumers. In other words, the distribution of card benefits for merchants first order stochastically dominates the distribution of card benefits for consumers ($G_m(x) \leq G_c(x) \ \forall \ x$), while the variance of the distribution is the same:*

- *A profit maximizing card network gives fewer rebates to merchants than to consumers ($R_c^* > R_m^*$) for all levels of transaction benefits.*
- *A social planner also gives fewer rebates to merchants than to consumers ($R_c^w > R_m^w$) for all levels of transaction benefits.*
- *For low transaction benefits, where $\alpha_t \leq \frac{R_m^{Max}+C_m}{N_c^w}$ is a sufficient and necessary condition, the price distortion is higher on the merchant side. For higher transaction benefits, the price distortion may be higher on the consumer side.*

Second, a similar argument holds for homogeneity. Suppose that on average merchants prefer cards as much as consumers do ($\overline{B_m} = \overline{B_c}$), but consumers are more heterogeneous than merchants (B_c is a mean-preserving spread of B_m). The distribution of card benefits of merchants second-order stochastically dominates the distribution of card benefits of consumers. For low transaction benefits, the card network gives higher rebates to merchants than to consumers, because merchant demand is more elastic to price changes than consumer demand. However, it also follows that for the same average benefit $R_m^{Max} < R_c^{Max}$, i.e. the merchant side is again "cornered" first. Thus, for higher transaction benefits, the card network gives higher rebates to consumers than to merchants. In both cases, the price distortion on the merchant side is lower. If transaction benefits are low, price discrimination and/or homogeneity reduces demand effects. If transaction benefits are high, the non-distortional shift reduces the price distortion further on the merchant side. This result is leads to Proposition 4 below.

Proposition 4 *Suppose that merchants are more homogeneous than consumers. In other words, the distribution of card benefits for merchants second order stochastically dominates*

the distribution of card benefits for consumers while the average benefit of cards compared to cash for merchants is the same as the average benefit for consumers ($\overline{B_m} = \overline{B_c}$):

- *For low transaction benefits, where $\alpha_t < \frac{R_m^{Max}+C_m+\eta_m^{Max}}{N_c^*}$ is a necessary and sufficient condition, the card network gives higher rebates to merchants than to consumers ($R_m^* > R_c^*$).*
- *For high transaction benefits, where $\alpha_t \geq \frac{R_m^{Max}+C_m+\eta_m^{Max}}{N_c^*}$ is a necessary and sufficient condition, the card network gives higher rebates to consumers than to merchants ($R_c^* > R_m^*$).*
- *The price distortion is higher on the consumer side for all levels of transaction benefits.*

Numerical example

In this numerical example card benefits are $B_c \sim \text{Uniform}(-5, 5)$ and $B_m \sim \text{Uniform}(-2, 8)$, while the costs of issuing a new card or accepting a card are $C_c = 1$ and $C_m = 1$ respectively. Homogeneous transaction fees/benefits are in total $f_t = \alpha_c + \alpha_m = 11$ [15] The threshold to make full consumer adoption profit maximizing is $16/0.66 = 24,2$ and the threshold on the merchant side is $13/0,56 = 23,2$. This threshold is lower on the merchant side than on the consumer side - meaning that the merchant side will be "cornered" first. Both thresholds are well above the total transaction benefits. It follows that the profit maximizing equilibrium arises at the interior solution. As shown in figure 2, the interior solution solves for: $R_c^* = 0.63$ and $R_m^* = -1.43$ with $n_c^* = 0.56$ and $n_m^* = 0.66$ and profits $\Pi = 3.43$.

The socially optimal rebate is at full market coverage, i.e. $R_c^w = 5$ and $R_m^w = 2$ with $n_c^w = 1$ and $n_m^w = 1$ (since $11 > \max\{\frac{5+1}{1}, \frac{2+1}{1}\}$). It follows that the total price distortion is 4.37 on the consumer side and 3.43 on the merchant side. This price distortion is build-up as follows. On the consumer side, the mark-up distortion and the scale distortion are 5.63

[15] we set the total level of transaction fees, as the distribution between the two sides in our model does not matter.

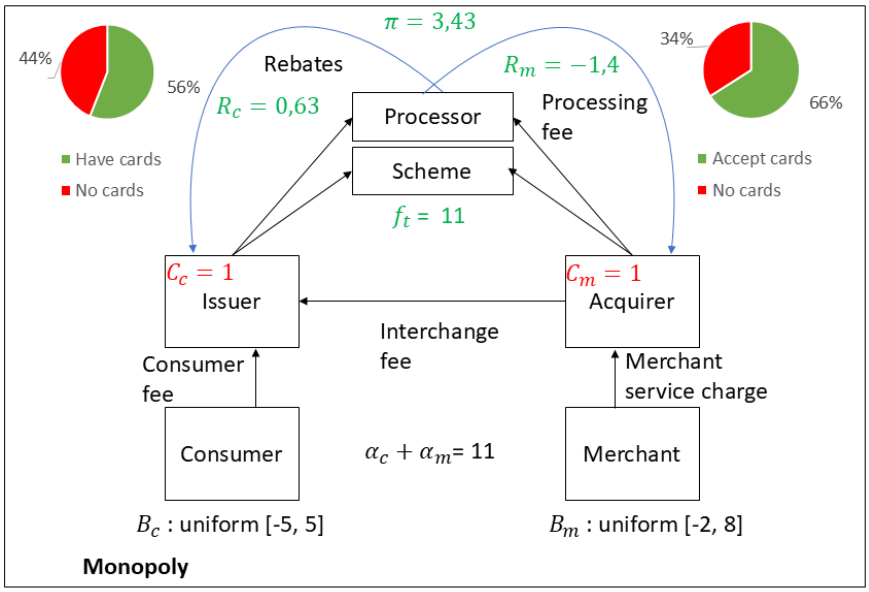


Figure 2: Profit maximizing solution with total transaction fees and benefits set at $f_t = 11$ and with card benefits B_c and B_m that are uniformly distributed $[-5, 5]$ for consumers and $[-2, 8]$ for merchants. The costs of issuing a new card or accepting a card are $C_c = 1$ and $C_m = 1$ respectively.

and 3.74 respectively, which adds up to 9.37, of which 5 is a non-distortional shift from consumer surplus to card network profit. On the merchant side, the mark-up distortion the scale distortion are 6.59 and 4.84 respectively, which adds up to 11.43, of which 8 is a non-distortional shift from merchant surplus to card network profit.

The figures below show how price distortions evolve with increasing transaction benefits over the two determinants discussed (i.e. level of benefits and uniformity of preferences). As shown in Figure 3, if merchants value cards more than consumers (i.e. distribution used above), the price distortion is higher on the consumer side for higher level of transaction benefits, because the maximal rebate required to convince consumers is higher than on the merchant side. Figure 4 shows how the price distortion evolves if consumers are more heterogeneous than merchants. Suppose that $B_m \sim \text{Uniform}(-2.5, 2, 5)$ (the rest the same). As merchants are now more homogeneous and have the same average benefit as consumers, the total price distortion is lower for merchants over all values of transaction benefits.

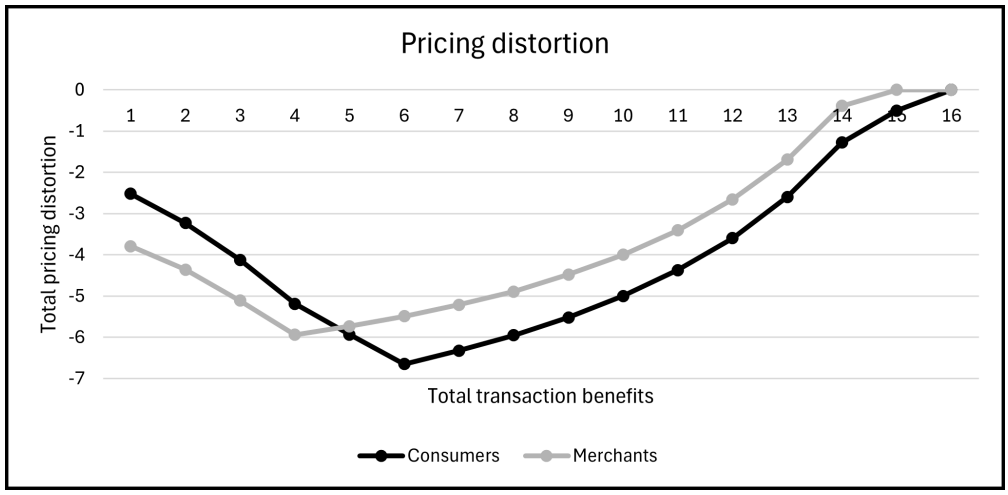


Figure 3: card benefits B_c and B_m are uniformly distributed $[-5, 5]$ for consumers and $[-2, 8]$ for merchants. The costs of issuing a new card or accepting a card are $C_c = 1$ and $C_m = 1$ respectively. There is full market coverage for transaction fees/benefits $f_t > 16$.

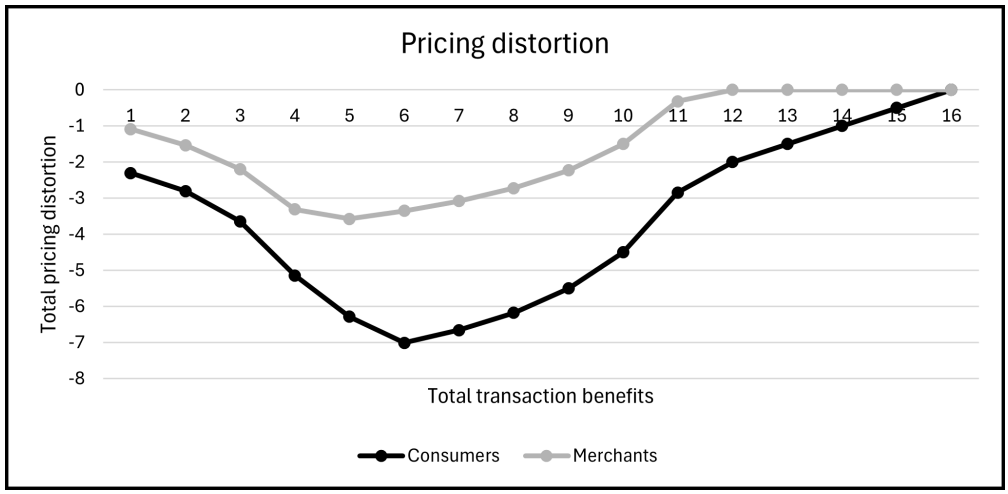


Figure 4: card benefits B_c and B_m are uniformly distributed $[-5, 5]$ for consumers and $[-2.5, 2.5]$ for merchants. The costs of issuing a new card or accepting a card are $C_c = 1$ and $C_m = 1$ respectively. There is full market coverage for transaction fees/benefits $f_t > 16$.

IV. Competitive bottleneck model

Now consider a two-sided market with two card networks, indexed by $k = \{A, B\}$, that compete for consumers and merchants to offer payment services. Competition in a two-sided market is analytically challenging. To keep the two-sided competition game tractable, we impose a strict competitive structure, see section V for a discussion of the assumptions made.

Basic model

First, we assume the same basic structure as in Section III. Transaction benefits are homogeneous for each market side, i.e., $\alpha_{c,k}$ and $\alpha_{m,k}$ are the same for all consumers and merchants respectively (see Assumption 1). Specifically, each side-i agent, $i = c, m$, has a type $\omega_i = (B_{i,A}, B_{i,B})$ and by joining network k, derives a gross payoff

$$u_{i,k}(\omega_i, n_{j,k}) = B_{i,k} + \alpha_{i,k} n_{j,k} \tag{7}$$

from transacting with a mass of agents of size $n_{j,k}$ from side j, $j \neq i$, that also joins network k. Both card networks are perceived as heterogeneous, which means that $B_{i,A}$ may be different from $B_{i,B}$ for the same agent. [16] The payment consists of a transaction fee $f_{i,k}$ paid per transaction with side j on platform k and a rebate $R_{i,k}$ (or card fee $F_{i,k}$) received from (or paid to) card network k by each side i user, as defined in Definition 1. Again, given quasi-linearity, the net payoff is $u_i^k(\omega_i, n_{j,k}) - f_{i,k} n_{j,k} + R_{i,k}$, given to side i by network k. Lastly, each card network incurs a cost $C_{i,k}$ for each side i user it brings on board. [17]

Second, we assume that platforms follow so-called "insulated equilibrium" (IE) strategies (Weyl, 2010; White and Weyl, 2016). In other words, the card networks charge transaction fees equal to transaction benefits. This special two-part tariff makes the participa-

[16] If both platforms would be homogeneous, there exists no market sharing equilibrium, see Jullien et al. (2021).

[17] Marginal transaction costs are w.l.o.g. normalized to 0, such that transaction benefits should be interpreted as net of transaction costs.

tion on each side independent - or "insulated" - of the participation decision on the other side. In the duopoly model with homogeneous transactional benefits, an insulated equilibrium arises when each card network selects an insulated two-part tariff $T_{i,k}$ of the form $T_{i,k} = f_{i,k}n_{j,k} - R_{i,k}$, where $f_{i,k}$ is equal to the homogeneous transaction benefit $\alpha_{i,k}$. All participation externalities are eliminated because

$$u_{i,k}(\omega_i, n_j) - f_{i,k}n_{j,k} + R_{i,k} = B_{i,k} + R_{i,k}$$

which is a fixed payoff independent of the opposite-side agents' actions.

Third, we impose a competitive bottleneck structure (Armstrong, 2006; Guthrie and Wright, 2007). That is, we assume that consumers choose at most one card network – they singlehome – whereas merchants are allowed to sign up for both card networks – they multihome. Given singlehoming on the consumer side, a consumer will only connect to card network k if $B_{c,k} + R_{c,k} \geq B_{c,l} + R_{c,l},\ k \neq l\ and\ B_{c,k} \geq -R_{c,k}$, otherwise she will use the network l (as long as the fixed payoff is larger than the outside option value of zero for cash) or use cash. That is, consumer demand for card network k is given by

$$\begin{aligned} n_{c,k} &= D_{c,k}(R_{c,k}, R_{c,l}) = Pr\{\omega_c \in \Omega_c :\ B_{c,k} \geq B_{c,l} - R_{c,k} + R_{c,l} \wedge B_{c,k} \geq -R_{c,k}\} \\ &= \int_{-R_{c,k}}^{\infty} \int_{-\infty}^{B_{c,k}+R_{c,k}-R_{c,l}} g_c(B_{c,k}, B_{c,l})\, dB_{c,l}\, dB_{c,k}, \quad k \neq l, \quad k, l = A, B. \end{aligned} \tag{8}$$

and corresponding "residual card" demand $n_{c,l}$, where g_c is the joint PDF of consumer card values over card networks A and B. Jullien et al. (2021) assume a fully served consumer market.[18] However, in our case, there may be consumers who do not connect to any card network – "pay by cash" – but would prefer network k to l. Total cash use is given by

$$\begin{aligned} n_{c,C} &= 1 - n_{c,A} - n_{c,B} = Pr\{\omega_c \in \Omega_c :\ B_{c,k} \leq -R_{c,k} \wedge B_{c,l} \leq -R_{c,l}\} \\ &\int_{-\infty}^{-R_{c,A}} \int_{-\infty}^{-R_{c,B}} g_c(B_{c,A}, B_{c,B})\, dB_{c,B}\, dB_{c,A} \end{aligned} \tag{9}$$

[18] In a fully served consumer market, it would simply hold $n_{c,l} = 1 - n_{c,k}$.

On the multihoming merchant side, since there are no business stealing effects, a merchant will connect to card network k if $B_{m,k} + R_{m,k} \geq 0$, which is independent of the rebate given on the other network l. This also means that the CDF G_m (and its PDF g_m) becomes one-dimensional and is only defined over the card benefit on card network k, i.e. $B_{m,k}$ (not $B_{m,l}$, where $k \neq l$). Therefore, merchant demand for card network k is simply given by

$$\begin{aligned} n_{m,k} &= D_m(R_{m,k}) = Pr\{\omega_m \in \Omega_m : \ B_{m,k} \geq -R_{m,k}\} \\ &= 1 - \int_{-\infty}^{-R_{m,k}} g_m(B_{m,k})\, dB_{m,k} = 1 - G_m(-R_{m,k}), \quad k = A, B. \end{aligned} \tag{10}$$

The difference between multihoming and singlehoming becomes clear in figure 5. In panel A, a singlehoming consumer may choose three possible alternatives which are fully disjoint, while in panel B a multihoming merchant may choose any combination of card networks next to cash.

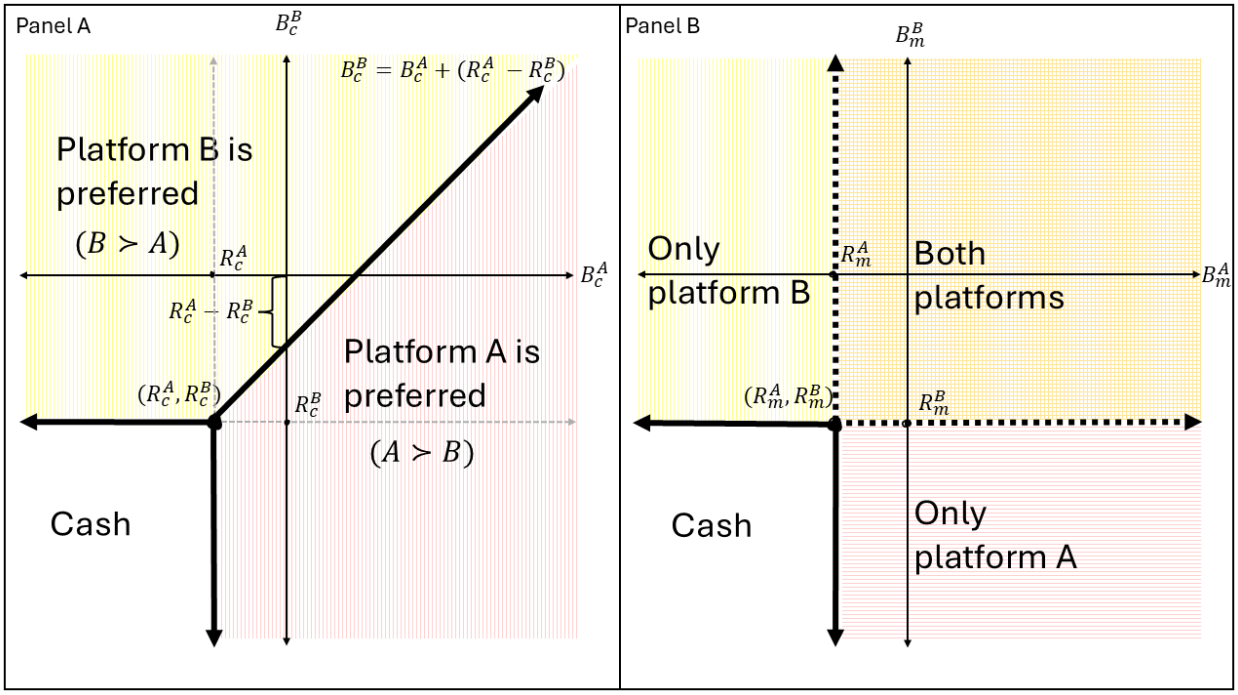


Figure 5: The difference between singlehoming consumers who may choose between card network A, B and cash which are fully disjoint (panel A) and multihoming merchants who may choose any combination of card networks next to cash (panel B).

Equilibrium outcome

Each card network's profits is specified by:

$$\begin{aligned}\Pi_k &= (f_c D_{m,k}(R_{m,k}) - R_{c,k} - C_{c,k}) D_{c,k}(R_{c,A}, R_{c,B}) \quad + \\ &\qquad (f_m D_{c,k}(R_{c,A}, R_{c,B}) - R_{m,k} - C_{m,k}) D_{m,k}(R_{m,k}) \end{aligned} \tag{11}$$

Let us start with the consumer side. In contrast to the monopolistic outcome, if card networks compete, there is no boundary solution with a maximal consumer rebate. Even if all consumers have cards, card networks continue to compete for each other's consumers. Therefore, consumer rebates are relatively easy to solve. The optimal card fees on the singlehoming consumer side are determined by the first-order condition

$$R_{c,k} = (f_c + f_m) n^*_{m,k}(R_{m,k}) - C_{c,k} - \mu_{c,k}(R_{c,A}, R_{c,B}), \tag{12}$$

where $\mu_{c,k}(R_{c,A}, R_{c,B})$ now denotes the inverse price elasticity of the *quasi residual card* demand (relative to network k) on the consumer side:

$$\mu_{c,k}(R_{c,A}, R_{c,B}) = -\frac{D_{c,k}(R_{c,A}, R_{c,B})}{\partial D_{c,k}(R_{c,A}, R_{c,B}))/\partial R_{c,k}} = \frac{1 - G_i(R_{c,A}, R_{c,B})}{g_i(R_{c,A}, R_{c,B})}$$

To see this, note that the FOC for card network k, $k = A, B$, with respect to the consumer side is given by:

$$\begin{aligned}\frac{\partial \Pi^k}{\partial R_c} = D_{c,k}(R_{c,A}, R_{c,B}) + (f_c D_{m,k}(R_{m,k}) - R_{c,k} - C_{c,k}))\, \frac{\partial D_{c,k}(R_{c,A}, R_{c,B})}{\partial R_{c,k}} \\ + f_m D_{m,k}(R_{m,k}) \frac{\partial D_{c,k}(R_{c,A}, R_{c,B})}{\partial R_{c,k}} = 0.\end{aligned}$$

Dividing by $\frac{\partial D(R_{c,A}, R_{c,B})}{\partial R_{c,A}}$ we have the FOC stated in equation (12).

On the merchant side, card networks do not compete and there is a boundary solution

where a card network has convinced all merchants to join. This condition is similar as in the monopolistic market structure. The first-order condition on the multihoming merchant side becomes:

- If

$$\alpha_t < \frac{R_m^{Max} + C_m + \mu_{m,k}^{Max}}{n_{c,k}^*}$$

There is an interior solution ($D_i(R_i) < 1, \quad \forall i$) and merchant rebates are:

$$R_{m,k} = (f_c + f_m) n_{c,k}^*(R_{c,A}, R_{c,B}) - C_{m,k} - \mu_{m,k}(R_{m,k}), \tag{13}$$

where $\mu_{m,k}(R_{m,k})$ denotes the inverse price elasticity of *quasi monopolistic card* demand:

$$\mu_{m,k}(R_{m,k}) = -\frac{D_{m,k}(R_{m,k})}{\partial D_{m,k}(R_{m,k})/\partial R_{m,k}} = \frac{1 - G_m(R_{m,k})}{g_m(R_{m,k})}.$$

- If

$$\alpha_t > \frac{R_{m,k}^{Max} + C_{m,k} + \mu_{m,k}^{Max}}{n_{c,k}^*}$$

There is an asymmetric boundary solution ($D_{m,k}(R_{m,k}) = 1, \quad D_{c,k}(R_{c,k}) < 1$) and rebates are:

$$R_{m,k}^* = R_{m,k}^{Max}$$

Similarly, to prove this, note that the FOC for card network k with respect to the merchant side is given by:

$$\frac{\partial \Pi^k}{\partial R_m} = D_{m,k}(R_{m,k}) + (f_m D_{c,k}(R_{c,A}, R_{c,B}) - R_{m,k} - C_{m,k}) \frac{dD_{m,k}(R_{m,k})}{dR_{m,k}} + f_c D_{c,k}(R_{c,A}, R_{c,B}) \frac{dD_{m,k}(R_{m,k})}{dR_{m,k}} = 0.$$

Dividing by $\frac{dD_m(R_{m,k})}{dR_{m,k}}$ we have the FOC stated in equation (13).

Consider next a social planner maximizing the sum of consumer and merchant utilities and both platforms' profit. For any vector of rebates ($P_{i,k} = f_{i,k}n_{j,k} - R_{i,k}$), consumer surplus ($V_c$) is given by:

$$V_c(R_{c,A}, R_{c,B}) = \int_{-R_{c,A}}^{\infty} \int_{-R_{c,B}}^{\infty} D_{c,A}(R_{c,A}, R_{c,B}) + D_{c,B}(R_{c,A}, R_{c,B}) dR_{c,A} dR_{c,B},$$

and merchant surplus ($V_{m,k}$) on platform k is given by:

$$V_{m,k}(R_{m,k}) = \int_{R_{m,k}}^{R_m^{Max}} D_{m,k}(R_{m,k}) dR_{m,k},$$

where $i \neq j$. Note that merchant surplus is independent of the merchant rebate provided on the other platform, while consumer surplus is dependent on the consumer rebate provided on the other platform.

Total welfare is given by:

$$W = V_c + V_{m,A} + V_{m,B} + \Pi_A + \Pi_B. \tag{14}$$

Taking the derivative with respect to $R_{i,k}$ and using the fact that the derivative of $V_{i,k}$ with respect to $R_{i,k}$ is equal to $n_{i,k}$, it can be shown that the welfare-maximizing rebates $(R^w_{i,k}, R^w_{j,k})$, along with boundary thresholds and the efficient participation level they induce $(n^w_{i,k}, n^w_{j,k})$, are exactly the same as stated in Proposition 2. For sake of shortness, the proposition is not restated. Of course, R^w_i and n^w_i should be replaced by $R^w_{i,k}$ and $n^w_{i,k}$ respectively, indicating the socially optimal level of side i users *on platform* k.

If we deduct the privately optimal price from the socially optimal rebate, we retrieve the total price distortion. The price distortion formula is more complex than in the monopolistic case, because consumer rebates are not maximized. Thus, on the consumer side, for high enough transaction benefits, it may be possible to have a solution where all consumers have joined either card network and consumer rebates increase further leading to a non-distortional

shift in the other direction (from card network profit to consumer surplus).

In general, the total price distortion is:

$$R^w_{i,k} - R^*_{i,k} = \underbrace{\alpha_t(n^w_{j,k} - n^*_{j,k})}_{\text{Scale distortion}} + \underbrace{\mu_{i,k}}_{\text{Mark-up distortion}} - \underbrace{\max\{0, \alpha_t n^w_{j,k} - C_{i,k} - R^{Max}_{i,k}\})}_{\text{Non-distortional shift}},$$

where $i \neq j$, $k \neq l$, $n^w_{i,k}$ is the socially optimal level of side i users on platform k, and $n^*_{m,k}$ is the privately optimal level of merchants of platform k. However, if

$$\alpha_t > \frac{R^{Max}_c + C_c + \mu_{c,k}}{n^w_m},$$

consumers receive more rebates than necessary to convince all of them and the non-distortional shift reverses. In this case, there is no distortion anymore on the consumer side, but the total profit maximizing rebate above the maximum rebate necessary becomes:

$$R^*_{c,k} - R^{Max}_{c,k} = \underbrace{\alpha_t n^*_{m,k} - C_{c,k} - \mu^*_{c,k} - R^{Max}_{c,k}}_{\text{Non-distortional shift}}.$$

It follows that depending on the level of transaction benefits, due to competition consumer rebates may increase beyond the socially optimal level, while merchant rebates fall further from the socially optimal level (due to a higher mark-up distortion). This is because the homing structure has important implications for the level of rebates, leading to Proposition 5. Suppose that card benefits have the same distribution across merchants and consumers. Competition for singlehoming consumers increases consumer rebates. It becomes harder to persuade consumers to use card network k's card, considering that they only use one card and there is a competing card network l. The same is not true for merchants, since they can accept both.

Proposition 5 *Suppose two card network simultaneously set rebates under Assumption 1:*

- *Rebates on the singlehoming side are determined by the FOC in equation (12) and*

rebates on the multihoming side are determined by the FOC in equation (13).

- *If consumers singlehome and merchants multihome while they have the same distribution of card benefits, it follows that in equilibrium consumer rebates are higher and beyond the socially optimal level, while merchant rebates are lower and below the socially optimal level.*

Proof. Given that card benefits have the same distribution across merchants and consumers, by definition the inverse semi-price elasticity of the *quasi residual* demand (relative to network k) on the consumer side is lower than the inverse semi-price elasticity of *quasi monopolistic* demand on the merchant side, i.e. $\eta_{m,k}(R_{m,k}) > \mu_{c,k}(R_{c,A}, R_{c,B})$. It follows in a duopoly with singlehoming consumers that $R^*_{m,k} < R^*_{c,k}$. ■

Numerical example continued

As in the monopolistic setting, suppose that the card benefits are

$$(B_{c,A}, B_{c,B}) \sim \text{Uniform}([-5, 5] \times [-5, 5]),$$

$$(B_{m,A}, B_{m,B}) \sim \text{Uniform}([-2, 8] \times [-2, 8]),$$

while the rest of the assumptions are the same as in section III.

The threshold to make full consumer adoption profit maximizing is $16/0.58 = 27,6$ (recall monopolistic: 24.2) and the threshold on the merchant side is $13/0,43 = 30,23$ (recall monopolistic: 23.2). This threshold is lower on the consumer side than on the merchant side (in contrast to monopolistic case) - meaning that the consumer side will be "cornered" first. Since both thresholds are well above the total transaction benefits, it follows that the profit maximizing equilibrium arises at the interior solution. As shown in figure 6, the interior solution solves for: $R^*_{c,k} = 1.17$ and $R^*_{m,k} = -2.15$ with $n^*_{c,k} = 0.43$ and $n^*_{m,k} = 0.58$

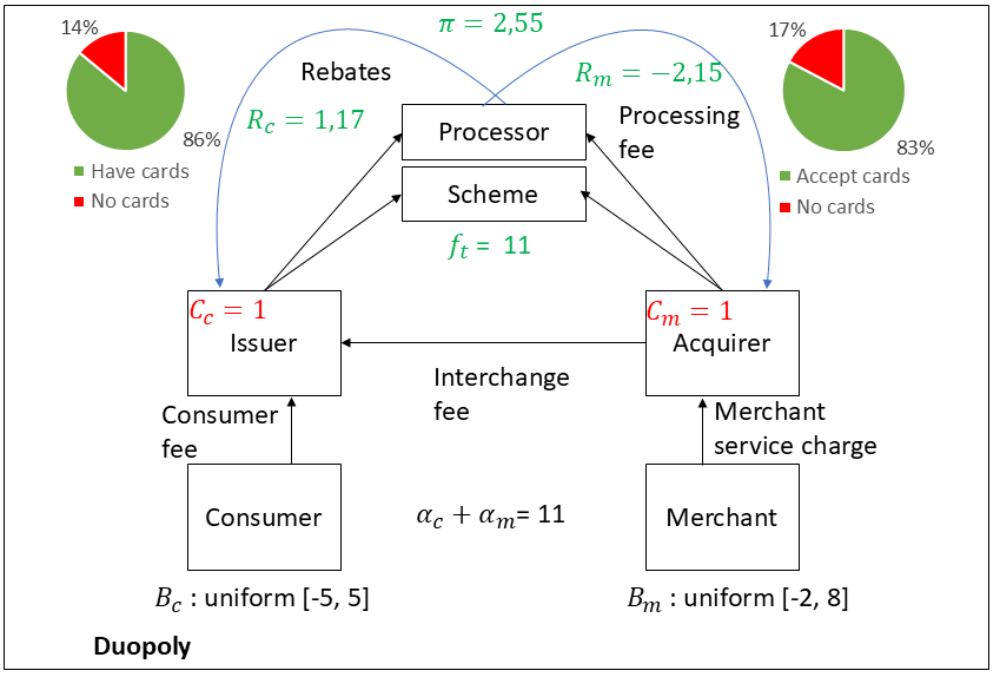


Figure 6: Profit maximizing solution with total transaction fees and benefits set at $f_t = 11$ and with card benefits B_c and B_m that are uniformly distributed $[-5, 5]$ for consumers and $[-2, 8]$ for merchants. The costs of issuing a new card or accepting a card are $C_c = 1$ and $C_m = 1$ respectively.

$(34, 2\%$ of merchants join both) and profits $\Pi_k = 2.49$. In total 86% of consumers and 83% of merchants join one of the two card networks.

The socially optimal rebate is at full market coverage, i.e. $R^w_{c,k} = 5$ and $R^w_{m,k} = 2$ with $n^w_{c,k} = 1$ and $n^w_{m,k} = 1$ (since $11 > \max\{\frac{5+1}{1}, \frac{2+1}{1}\}$). It follows that the total price distortion is 3.83 on the consumer side and 4.15 on the merchant side. This price distortion is build-up as follows. On the consumer side, the mark-up distortion and the scale distortion are 4.26 and 4.57 respectively, which adds up to 8.83, of which 5 is a non-distortional shift from consumer surplus to card network profit. On the merchant side, the mark-up distortion the scale distortion are 5.84 and 0.81 respectively, which adds up to 6.65, of which 2.5 is a non-distortional shift from merchant surplus to card network profit. As shown in Figure 7, the price distortion is higher on the consumer side for higher level of transaction benefits, because the maximal rebate required to convince consumers is higher than on the merchant side.

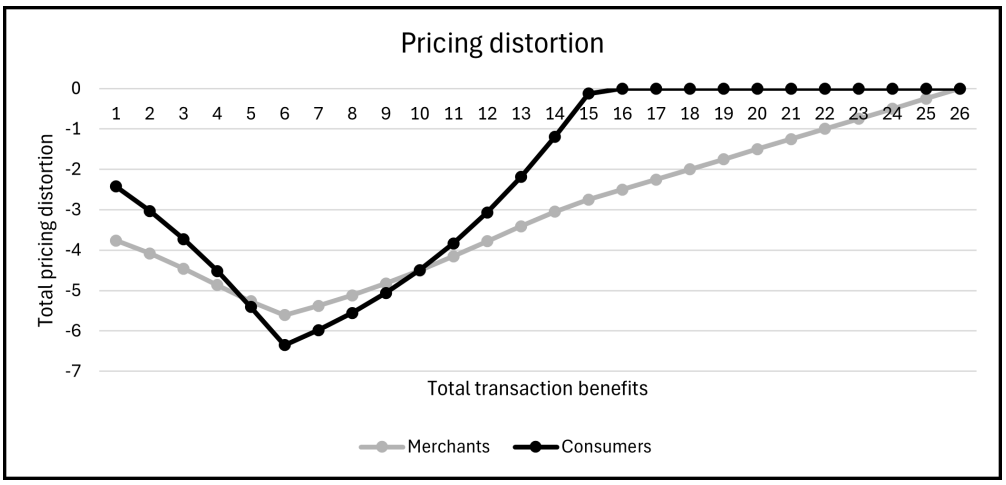


Figure 7: The figure is set for symmetric card networks with card benefits B_c and B_m that are uniformly distributed $[-5, 5]$ for consumers and $[-2, 8]$ for merchants. The costs of issuing a new card or accepting a card are $C_c = 1$ and $C_m = 1$ respectively. There is full market coverage for consumers if transaction fees/benefits $f_t > 16$.

Now suppose transaction fees and benefits increase around 50 percent to $f_t = 16$. The duopolistic rebates increase and solve for: $R_c^{A*} = R_c^{B*} = 6$ with $n_c^{A*} = n_c^{B*} = \frac{1}{2}$ and $R_m^{A*} = R_m^{B*} = -0.5$ with $n_m^{A*} = n_m^{B*} = 0.75$ and profits decrease to $\Pi^A = \Pi^B = 2.13$. The socially optimal rebates are the same as before, resulting in a total price distortion of 0 on the consumer side and 2.50 on the merchant side. This price distortion is build-up as follows. On the consumer side, the mark-up distortion and the scale distortion are 5 (maximal) and 4 respectively, which adds up to 9, all of which is a non-distortional shift from card network profit to consumer surplus. Note that this shift is the other way around, since the profit maximizing rebate is above the maximum rebate necessary to convince all consumers. On the merchant side, the mark-up distortion and the scale distortion are 7.50 and 0 respectively, which adds up to 7.50, of which 5 is a non-distortional shift from merchant surplus to card network profit.

In a monopolistic case, there is no price distortion for this level of transaction benefits, but profits are much higher. The monopolistic rebates solve for $R_c^M = 5$ with $n_c^M = 1$ and $R_m^M = 2$, $n_m^M = 1$ and $\Pi^M = 7$. The intuition behind this result becomes visible in figure 8, which shows how profits develop in a monopoly and duopoly market structure.

In this figure, there is full consumer coverage for transaction fees/benefits $f_t > 11$ in a duopoly market structure (recall that this was $f_t > 16$ in a monopoly market structure). As shown, it follows that transaction benefits only increase equilibrium duopoly profits if the network is smaller (i.e. transaction benefits are lower). Moreover, competition for consumers decreases the total rebates on the merchant side, which leads to a higher price distortion on the merchant side. In contrast, a monopolist does not face competition and as soon as it has persuaded all consumers to have a card, it sets the rebate at $R_c^{Max} = -B_c^{Min}$ for $B_c^{Min} < 0$ and $R_c^{Max} = 0$ otherwise. Thus, for a monopolist the mark-up on transaction fees per consumer increases even more over the transaction fees in the boundary solution compared to the interior solution. Since all consumers have a card, there is no reason to offer additional rebates that would reduce the mark-up in transaction fees.

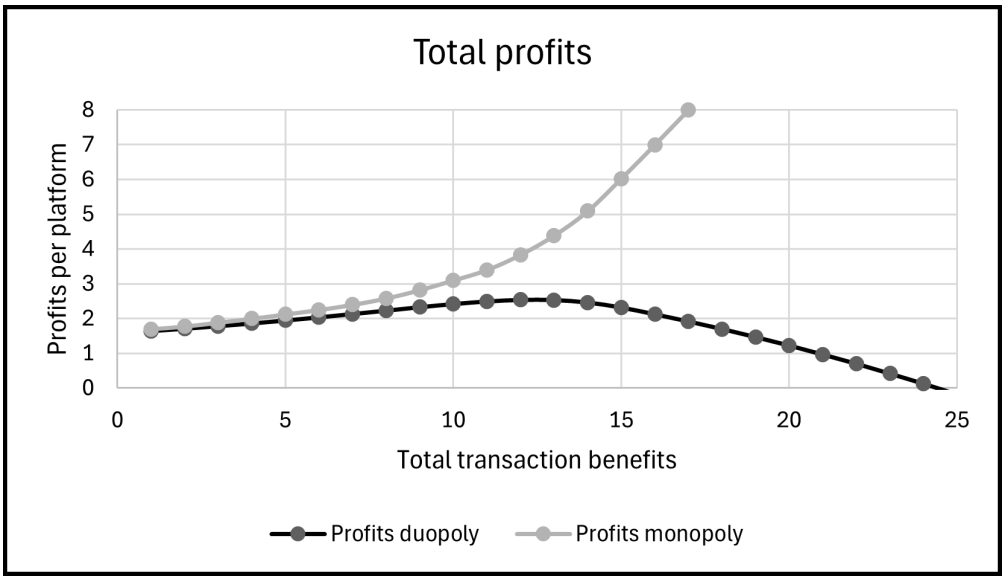


Figure 8: The figure is set for symmetric card networks with card benefits B_c and B_m that are uniformly distributed $[-5, 5]$ for consumers and $[-2, 8]$ for merchants. The costs of issuing a new card or accepting a card are $C_c = 1$ and $C_m = 1$ respectively. All consumers join a card network if transaction fees/benefits $f_t > 11$ in a duopoly market structure (50/50 split) and $f_t > 16$ in a monopoly market structure.

V. Discussion of critical assumptions

A. Insulated tariffs

A key question in a model with competition between platforms is how platforms resolve issues of "coordination failure" between users. As shown for the first time by Katz and Shapiro (1985), the expectation of users about the network size of each platform determines the type of market structure. Different consumer expectations may lead to multiple equilibria. The idea is as follows. If users expect a single network to be large, then they are more likely to join this network leading to monopoly. If users expect two networks to be large, they might join both, leading to a duopoly. Thus, expectations on network sizes lead to a self-fulfilling prophecy. There are two strands of literature that deal differently with the issue of coordination failure.

On the one hand, if users are assumed to resolve this coordination failure in the direction of the same platform, then platforms can make use of "unfavorable beliefs" to "tip" the market and have all users on their platform (Caillaud and Jullien, 2001, 2003; Jullien, 2011). Platforms do this with "divide-and-conquer strategies" where the side with the lowest transaction benefit is subsidized ("divided") so that the transaction benefits on the other side can be extracted ("conquered"). Depending on whether users are "singlehoming" and use one platform, or "multihome", and use both platforms, platforms can use complex tariffs to alter the market structure. In a model with homogeneous platforms that offer intermediation services, Caillaud and Jullien (2001, 2003) show that if users "'singlehome", an incumbent platform sets high transaction fees with negative card fees to deter entry, as the entrant would need to offer a higher subsidy. If users"multihome" platforms set low transaction fees and high card fees to deter entry as the entrant would not be able to earn any profits if transaction fees are low.

On the other hand, there is a large literature that assumes that platforms have a sym-

metric selection criterion (Rochet and Tirole, 2003b; Weyl, 2010; White and Weyl, 2016). Platforms use the so-called "insulated equilibrium" (IE) strategies based on the idea that card network fees can be designed so that participation on each side is independent - or "insulated" - of the participation decisions on the other side.[19] In other words, given these card network fees, each agent's participation decision is determined by a dominant strategy. If both card networks use an IE strategy, they share the market as long as fixed costs $C_{i,k}$ are not too large. As $C_{i,k}$ increases, one of the card networks could set the rebates at a level that makes it unprofitable for the other card network to remain active.

In this paper, we take the approach of the latter strand of literature by focusing on the case where two card networks share the market. This can be (partly) justified by the fact that the two biggest card networks do, in fact, set transaction fees with rebates and thus seem to play IE strategies where $C_{i,k}$ is not too large to let the market tip towards a single card network. As discussed in Armstrong (2006), competition is stronger with fixed card fees instead of variable transaction fees, because with fixed card fees consumer demand is more elastic as consumers care about the amount of merchants. We assume that platforms make use of this fact by pricing the transaction benefits such that these no longer play a role in the participation decision. It would be an interesting avenue for future research to analyze the incentives of "market tipping" in a rebates context.

B. Homogeneous transaction benefits

The assumption of homogeneous transaction benefits and heterogeneous card benefits stated in Assumption 1 has been inspired by the idea that users differ more with respect to their preference for cards compared to cash than the size of the network. This would translate into more heterogeneity in card benefits and less heterogeneity in transaction benefits. Yet, most payment papers focus on heterogeneous transaction benefits (Chakravorti and Roson, 2004; Guthrie and Wright, 2007; Rochet and Tirole, 2002, 2003a, 2011; Wright, 2004). We add to

[19] Under homogeneous interaction benefits and regularity conditions ensuring equilibrium existence and uniqueness, the outcome of Cournot equilibrium and insulated equilibrium coincide, see Jullien et al. (2021).

existing payment literature by looking at heterogeneous card benefits and shows that in case of high homogeneous transaction benefits, card networks charge a mark-up over transaction fees reduced by rebates.

Allowing for heterogeneous transaction benefits would make our model mathematically more complex to solve, but would not drastically change our results. To see this, note that heterogeneous transaction benefits could be extracted with an "insulated fee" with heterogeneous transaction values as solved by Weyl (2010) and White and Weyl (2016). In case of heterogeneous transaction benefits, the best a card network could do is to "insulate" the representative user's valuation of transaction benefits since there is no single transaction fee that gives all users a dominant strategy to join. As explained by Weyl (2010), this representative user is for a private card network the marginal user (instead of the average user), which leads to an extra price distortion. Allowing for heterogeneous transaction benefits would change the price structure of the transaction fees between the merchant and consumer side, but would not change our main results on how the size of the network and price distortions evolve over total transaction benefits. The reason is that level of rebates in our model is determined by the total level of transaction fees and independent of how these are distributed between consumers or merchants.

Note that Bedre-Defolie and Calvano (2013) are among the few who consider both heterogeneous card benefits and heterogeneous transaction benefits and derive surprisingly similar results. In their model, consumers first decide whether they will take a card and then decide on usage of payment cards, while merchants only make a card acceptance decision based on the MSC. The result is that card networks use a two-part tariff. A monopoly card network charges a fixed monopoly mark-up over the fixed consumer fee, while it charges competitive usage fees to both consumers and merchants to maximize usage. In contrast, in our model transaction benefits are homogeneous, while card benefits may be negative. Therefore, transaction fees are set compatible at the maximum level and consumers cannot be incentivised to use cards through lower usage fees. The outcome of our model is surprisingly similar.

Except that in our model the mark-up η_i or $\mu_{i,k}$ does not come from a higher fixed consumer card fee, but operates through higher transaction fees allowing optimal rebates.

C. Fixed rebates independent of transaction volume

In our model, we assume that card networks use rebates to stimulate card issuance assuming that each cardholder will do a constant amount of transactions. Rebates in our paper could be interpreted as a total subsidy, which could also consist of transaction subsidies. Note that in the demand specification (2) above, only the total price level matters. The difference between card and transaction fees is irrelevant. This means that prices only affect "usage" or "participation", but not "usage" conditional on "participation". Existing papers generally do not differentiate between effects of variable transaction fees and fixed card fees on usage demand and card demand (Armstrong, 2006; Caillaud and Jullien, 2001, 2003; Rochet and Tirole, 2003b, 2006; Weyl, 2010). For example, Rochet and Tirole (2003b) only model the usage decision, while Armstrong (2006) only models the membership decision.

One could argue that a large part of rebates is paid based on transaction output to stimulate card usage and next to these rebates issuing banks receive the IF on a per transaction basis. If rebates could be given on a per transaction basis negative card benefits could be offset by the increased "option value" of the card through a usage subsidy. In particular, Bedre-Defolie and Calvano (2013) show that the option value of having a card increases with merchant participation. Therefore, in equilibrium the card benefit (net of the option value) would always be positive and offset any negative card benefits. Our model could be seen as a contrasting benchmark where card networks are unable to set rebates directly to consumers. In practice it may not be feasible for card networks to set usage rebates directly to consumers since issuing banks are "sitting in between". As we show, in this case a card network can only maximise card membership through rebates.

A richer model would allow usage decisions to be taken separately to differentiate between transaction and card rebates. The assumption of homogeneity of transaction benefits comes

at a drawback, as it does not allow to explicitly model payment card usage. The price structure of merchant vis-a-vis consumer transaction fees is unimportant. So rebates can only be interpreted as a total subsidy without differentiating between card usage and card ownership/acceptance. Heterogeneous transaction benefits do explain why transaction fees are higher on the merchant side compared to the consumer side. The card network would distort the network-effect provision passing through the preferences of marginal rather than average consumers or merchants, leading to what e.g. Weyl (2010) and White and Weyl (2016) call the traditional "Spence distortion" and to which Tan and Wright (2018, 2021) later add the "Displacement distortion". So far little progress has been made in differentiating between usage and card subsidies and this is a fruitful topic for future research.

D. Consumer multihoming

Our model does not allow for consumer "multihoming", while an extended model could allow for some consumers to multihome such that some consumers use several cards. One may argue that in a model with "multihoming" consumers, card networks no longer have a monopoly over providing access to this group of consumers via payment services to merchants. Our results in section IV would likely result in lower consumer rebates, as cross-subsidization may be reduced or even disappears when both sides of the card network multihome (Bakos and Halaburda, 2019). Moreover, when consumers multihome, card networks have less market power over merchants, because merchants could "steer" consumers towards their preferred payment method by singlehoming on the cheaper payment method (Rochet and Tirole, 2003b).

Yet, the possibility of merchant steering depends crucially on consumer loyalty (Teh et al., 2023). Our assumption of "singlehoming" consumers may thus also be justified by the fact that although consumers may be multihoming, consumer loyalty means that merchants could lose too much business by turning down preferred cards by consumers (see e.g. Rysman (2007)). In this case each card network still has market power over merchants. card networks

increase consumer loyalty by increasing rebates for higher card usage. It is analytically challenging to allow for such endogenous singlehoming and we also leave it as an interesting question for future research.

E. Inelastic demand on the product market

This paper abstracts from effects of card fees and rebates on consumer demand in the product market. In our model consumer demand at retailers/merchants is given and inelastic to card fees and rebates. Schwartz and Vincent (2006, 2020) show in a model with elastic demand that if rebates increase demand in the product market that card networks have an incentive to raise the spread between the card fee on the merchant side and the rebate on the consumer side. Merchants have no incentive to turn down expensive cards, since they sell more products as consumer demand increases with rebates. It follows that in their model with platform competition, there is no pure-strategy equilibrium. The card network with a lower merchant fee always has an incentive to increase this fee with higher rebates until merchant fees on both card networks are too high for merchants to accept cards. We focus on inelastic consumer demand as one may argue that rebates are sufficiently small such that it cannot increase demand in the product market. We show that in this case a pure strategy equilibrium does exist.

F. No Surcharge Rule

This paper focuses on the case where merchants cannot discriminate in prices between card users and cash users (also known as imposing the "No Surcharge Rule". As first shown by Gans and King (2001) and Rochet and Tirole (2002), lifting the "No Surcharge Rule" would have a neutralizing effect on the IF. Transposing this to our model, the level of rebates would be irrelevant, as consumers would need to pay a higher price for products exactly equal to the rebates received. Thus, there would no longer be an incentive for card network to offer rebates. In most countries, four-party payment schemes use a "No Surcharge Rule"

- justifying our assumption.

G. No competition between merchants

This paper abstracts away from "business stealing effects" on the merchant side. Rochet and Tirole (2011) show that merchants may be willing to accept cards even if the fee they have to pay exceeds their direct benefit for card payments. Card acceptance increases the merchant's quality of service by offering an additional payment method and by doing so a merchant may attract customers from its competitors. Indeed, the fear of losing customers weakens the merchant's position to refuse a card payment, which is then exploited by card networks. This could be an additional reason why card networks do not give rebates to merchants.

VI. Conclusion

This paper shows that the difference between card benefits and transaction benefits matters for card network pricing and profits for two main reasons. First, total transaction fees and profits increase as total transaction benefits increase, while card networks give rebates to push "cash-lovers" towards cards. This is most prevalent in a monopolistic card network setting, where after the card network has set rebates such that all consumers have cards, the platform can earn maximum profit over marginal transaction fees. In a duopolistic market structure, profits stabilise when all consumers have cards of either card network. As card networks compete for each other's cardholders, marginal rebates are set equal to marginal transaction fees.

Second, our analysis shows that the price structure of transaction fees which is determined by the IF is not the only relevant metric to analyse price distortions in the payment market and rebates play an important role. Almost all previous papers in the payments literature assume heterogeneous transaction benefits and homogeneous card benefits normalised to 0

to show price distortions in the price structure of transaction fees. However, if users differ more with respect to their preference of payment method, i.e. cards or cash, instead of their preference for the size of the card network, i.e. consumer adoption or merchant acceptance, then card benefits are more heterogeneous than transaction benefits. We show that in the extreme case where transaction benefits are homogeneous, the price structure of transaction fees is irrelevant. Instead, rebates play an important role to determine the mark-up distortion induced by market power of card networks.

This paper shows that card networks give lower rebates than socially optimal, especially on the merchant side if there is card network competition. Card networks only give rebates to merchants if the total transaction fees earned per consumer are higher than the costs and mark-up per merchant. In reality, transaction fees on the consumer side are mostly negative (through an IF) or non-existent. Low consumer transaction fees indicate a low mark-up on consumer transaction fees per merchant and a high consumer rebate. We provide three main reasons why rebates are offered to the issuing side by card networks. In particular, issuing banks may be offered higher rebates as costs are higher, consumers show more heterogeneity[20], have lower card or transaction benefits than merchants, and/or are more "loyal" to a certain card network. This last effect could also be due to the fact that issuing banks often "choose" the card network for their customers and can only choose one card network for each card they issue.

This is the first paper to analyse the difference between card benefits and transaction benefits on payment pricing, which leaves many open questions for future research discussed in section V.

[20]For example, hotels, restaurants and bars often pay a higher fee as they have a higher willingness to pay, see European Commission (2007). Price discrimination is effectively similar to assuming homogeneous merchants, as card networks can extract all surplus and set the merchant fee at a level such that all merchants participate.

References

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics*, *37*(3), 668–691.

Bakos, Y., & Halaburda, H. (2019). Platform Competition with Multi-homing on Both Sides: Subsidize or Not? *Management Science*, *66*(12), 5599–5607.

Baxter, W. F. (1983). Bank Interchange of Transactional Paper: Legal and Economic Perspectives. *Journal of Law & Economics*, *26*(3), 541–588.

Bedre-Defolie, O., & Calvano, E. (2013). Pricing Payment Cards. *American Economic Journal: Microeconomics*, *5*(3), 206–231.

Caillaud, B., & Jullien, B. (2001). Competing cybermediaries. *European Economic Review*, *45*(4-6), 797–808.

Caillaud, B., & Jullien, B. (2003). Chicken & Egg: Competition among Intermediation Service Providers. *The RAND Journal of Economics*, *34*(2), 309–328.

Chakravorti, S., & Roson, R. (2004). Platform Competition in Two-Sided Markets: The Case of Payment Networks. *SSRN Electronic Journal*.

European Commission. (2007). *Report on the retail banking sector inquiry* (tech. rep.). Directorate-General for Competition.

Gans, J. S., & King, S. P. (2001). The Neutrality of Interchange Fees in Payment Systems. *Topics in Economic Analysis and Policy*, *3*(1), 1069–1069.

Guthrie, G., & Wright, J. (2007). Competing Payment Schemes. *The Journal of Industrial Economics*, *55*(1), 37–67.

Jullien, B. (2011). Competition in Multi-sided Markets: Divide and Conquer. *American Economic Journal: Microeconomics*, *3*(4), 186–220.

Jullien, B., Pavan, A., & Rysman, M. (2021). Two-sided markets, pricing, and network effects. In K. Ho, A. Hortaçsu, & A. Lizzeri (Eds.), *Handbook of Industrial Organization* (pp. 485–592). Elsevier.

Katz, M. L., & Shapiro, C. (1985). Network Externalities, Competition, and Compatibility. *The American Economic Review*, *75*(3), 424–440.

Rochet, J.-C., & Tirole, J. (2002). Cooperation among Competitors: Some Economics of Payment Card Associations. *The RAND Journal of Economics*, *33*(4), 549.

Rochet, J.-C., & Tirole, J. (2003a). An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems. *Review of Network Economics*, *2*(2).

Rochet, J.-C., & Tirole, J. (2003b). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, *1*(4), 990–1029.

Rochet, J.-C., & Tirole, J. (2006). Two-sided markets: A progress report. *The RAND Journal of Economics*, *37*(3), 645–667.

Rochet, J.-C., & Tirole, J. (2011). Must-take Cards: Merchant Discounts and Avoided Costs. *Journal of the European Economic Association*, *9*(3), 462–495.

Rysman, M. (2007). An Empirical Analysis of Payment Card Usage. *The Journal of Industrial Economics*, *55*(1), 1–36.

Schwartz, M., & Vincent, D. R. (2006). The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network. *Review of Network Economics*, *5*(1).

Schwartz, M., & Vincent, D. R. (2020). Platform Competition With Cash-Back Rebates Under No Surcharge Rules. *The Journal of Industrial Economics*, *68*(2), 298–327.

Tan, H., & Wright, J. (2018). A Price Theory of Multi-Sided Platforms: Comment. *American Economic Review*, *108*(9), 2758–2760.

Tan, H., & Wright, J. (2021). Pricing distortions in multi-sided platforms. *International Journal of Industrial Organization*, *79*, 102732.

Teh, T.-H., Liu, C., Wright, J., & Zhou, J. (2023). Multihoming and Oligopolistic Platform Competition. *American Economic Journal: Microeconomics*, *15*(4), 68–113.

Verdier, M. (2011). Interchange Fees in Payment Card Systems: A Survey of the Literature. *Journal of Economic Surveys*, *25*(2), 273–297.

Weyl, E. G. (2010). A Price Theory of Multi-Sided Platforms. *The American Economic Review*, *100*(4), 1642–1672.

White, A., & Weyl, E. G. (2016). Insulated Platform Competition.

Wright, J. (2004). The Determinants of Optimal Interchange Fees in Payment Systems. *The Journal of Industrial Economics*, *52*(1), 1–26.