WORKING PAPERS

N° 9

2023

“Fintechs and financial inclusion in payment markets with operational risk”

Carola Müller

FIT IN Initiative

Fintechs and financial inclusion in payment markets with operational risk *

Carola Müller[ab]

[a] *Center for Latin American Monetary Studies (CEMLA), Durango 54, 06700 Ciudad de México, Mexico.*

[b] *Halle Institute for Economic Research (IWH), Kleine Märkerstrasse 8, 06108 Halle (Saale), Germany.*

Abstract

Fintechs promise to offer payment services to merchants and consumers that were previously excluded from digital payment markets, especially in emerging economies. Using a two-sided market model, the paper outlines conditions that lead to financial exclusion. Besides from cost inefficiencies and market power, operational risk can cause financial exclusion. While the entry of a fintech can eliminate cost-driven and market-power-driven exclusion, risk-driven financial exclusion cannot be fully banished. In order to improve financial inclusion in payment markets with operational risk, fintechs need a degree of market power.

JEL Classification: D45, L12, G59, G21, O16.

Keywords: Retail payments, debit cards, credit cards, market power, financial inclusion, operational risk, fintech.

*This work received financing from the TSE FIT-IN Initiative sponsored by the Bill and Melinda Gates Foundation. It exists thanks to the discussions with Anahí Rodríguez Martínez, Raul Morales, and José Aurazo. Anahí also provided excellent research assistance. I further thank Milo Bianchi and Matthieu Bouvard for useful comments and suggestions. All errors are my own.

1 Introduction

The landscape of the retail payments market is changing rapidly at the face of wide spreading innovative technologies. The arrival of fintechs in form of new payment service providers (PSPs) comes with the hope of solving some of the prevailing market failures in payments markets: tougher competition should bring down elevated price levels and easy technologies should provide universal access to payments services (Philippon, 2019; IMF, 2020). These hopes are often provoked by fintechs' marketing strategies.[1] In emerging economies, many customers and merchants do not use or accept card payments nor other forms of digital payment methods. This may be due to a lack of demand or due to supply frictions. On the one hand, informal labor and tax evasion foster the use of cash. On the other, financial firms have market power and fees on financial services are high. As a result, financial inclusion remains limited, especially for low-income households or small and medium sized merchants.

The question arises why traditional payment firms excluded customers in the past and why fintechs would serve them now? This paper aims to explain the root causes of lacking financial inclusion and fintechs' business models' potential to address these. The first contribution of the paper is to augment a two-sided market model of the card payments market with heterogeneous clients in order to study financial exclusion. Another crucial innovation is to explore a novel explanation of exclusion: operational risk.

Operational risks are a major concern for payment companies.[2] They comprise malfunctioning risk, data security risk, and risk of fraud which can ultimately result in reputation risks for the payment company. In part, these risks can be amplified through negligent or even fraudulent behaviour by clients. Operational risks also cause costs in form of compliance duties, know-your-customer checks, and on-going controls.

The paper illustrates that financial exclusion can be driven by costs, market power, or operational risk. When fintechs have a cost advantage, the results show that they can eliminate cost-driven exclusion and provide sufficient competition to disrupt market-power driven exclusion. However, when accounting for operational risks, even fintechs optimally exclude the riskiest clients. Furthermore, in order to improve financial inclusion in risky payment markets, fintechs need a degree of market power. Otherwise, fintech entry reduces fees but financial exclusion stays at the same level as when traditional firms operate in the market.

The model is based on the two-sided card payment model in Rochet and Tirole (2002, 2003). While card holders contract with issuers, which are usually commercial banks, merchants interact with acquirers, often non-banks. Card holders pay a fee per transaction to issuers. In reality, these fees are often negative, i.e., customers get benefits when using their payment cards. Merchants pay the so-called discount rate to acquirers. And acquirers pay the interchange fee

[1] For example, Marcos Galperin, CEO and President of Mercado Libre Inc., said: "We continue to revolutionize the democratization of payments and the inclusion of small and medium sized merchants in the financial system" (Mercado Libre Sustainability Report 2019). Adolfo Babatz, CEO of clip, a Mexican payment aggregator, claims that "Clip is the missing piece in financial inclusion" (Interview on 04/01/2018). Markets also place high valuations on fintech PSPs. In 2021, Silicon-valley's highest prized company was the payment aggregator Stripe (Financial Times, 2021).

[2] For example, there were cybersecurity breaches at PayPal in 2020 (Forbes), and 2014 (Washington Post); or cyber attacks on payment systems at convenience stores in the UK in 2021 (BBC News).

to issuers. The merchant-side is often the profit-making segment of the market (Bedre-Defolie and Calvano, 2013). Merchant fees are high, especially in emerging markets[3], not least, because acquirers pass interchange fees through to merchants.

The model in this paper focuses on the exclusion of merchants from the acquiring side of the market. Usually smaller shops and micro-businesses do not accept cards. Exclusion affects the less wealthy, less established, and less formal firms. In order to account for this, the model includes heterogeneous merchants. Merchants differ in quality. Low-quality merchants have lower transaction volumes, hence quality is synonymous with size. However, low-quality merchants also expose acquierers to more operational risk. There are no information asymmetries and acquirers can deny merchants access to the payment system based on their risk. The model renders both, optimal fees and optimal exclusion levels, or the optimal relation between them.

The baseline of the analysis is a perfectly competitive acquiring market without the existence of operational risk. In this setting, acquirers charge marginal costs to merchants and all merchants are admitted in the payment system. Merchants would only be excluded if the acquiring technology is so costly, that merchants prefer incurring the costs of cash. This would be cost-driven exclusion.

However, when operational risk is taken into account, perfect competitors optimally exclude low-quality merchants. This would be risk-driven exclusion. Acquirers provision for expected losses due to operational risks. These increase as they admit merchants with lower quality.In perfect competition, acquirers have to set prices above marginal costs in order to compensate for expected losses. By excluding another merchant, an acquirer reduces expected losses, a safety gain, but forfeits revenue on the transaction volume that the excluded merchant has to offer. Under a zero-profit-condition, the relation between merchant rates and exclusion levels is U-shaped: Initially, when very low-quality merchants were excluded, the safety gain would be high and the revenue loss small. Hence, acquirers could offer lower rates to the remaining merchants and still make zero profits. Later, when high-quality merchants were excluded, the safety gain would be small while the forfeited revenue would be large. Hence, the acquirer would need to increase the merchant rate in order to avoid losses. The perfectly competitive equilibrium features an intermediate level of exclusion at the lowest possible merchant rate. In equilibrium, none of the competitors has an incentive to deviate since neither excluding one more nor one less merchants would be profitable. Consequently, operational risks can be one reason that explains financial exclusion, even in perfectly competitive markets.

Nevertheless, perfect competition in acquiring markets might not be a fitting assumption for emerging economies that have low levels of financial inclusion.[4] The model therefore also analyses the case of a monopolist. Acquierers with market power face a trade-off: when they charge higher fees, they might lose some low-quality clients but they earn higher margins on the remaining ones. The model shows that a monopolist-acquirer optimally excludes a range of low-quality merchants and charges a markup on marginal costs. When operational risks exist, the monopolists charges an additional markup on risk. Hence, market power in the acquiring

[3] See Arango-Arango et al., 2022, for Colombia.

[4] The eight largest acquirers in Latin America account for 78% of Mastercard and Visa transactions acquired by the top 45 acquiring businesses according to an industry report by Nilson.

market is another explanation for financial exclusion of low-quality merchants which worsens when operational risks are also taken into account.

Finally, the paper studies if and how fintech acquirers can increase financial inclusion. By assumption, fintechs have a competitive advantage due to lower marginal costs. With fintech competitors, the model reduces to a Bertrand competition game with asymmetric costs. Having one fintech competitor with a cost advantage is sufficient to eliminate financial exclusion stemming from market power of the traditional acquirer. The fintech can steel all clients from the monopolist by offering a merchant rate marginally below the marginal costs of the monopolist. The same is true when the fintech company competes against a set of traditional acquirers. With such a clear advantage, fintech competition would be disruptive.

Moreover, when there are operational risks, fintechs also have to provision for expected losses. The results show an interesting trade-off between market power and financial inclusion: While a single fintech competitor can reduce, but not eliminate, financial exclusion stemming from operational risks, many fintech competitors that compete among themselves cannot reduce financial exclusion. Since the optimal exclusion level in perfect competition is defined solely by the relation between merchants' risk and merchants' size, it is independent of acquirers' characteristics such as marginal costs. Hence, if exclusion is risk-driven, fintechs' promise to "democratize payments" crucially depends on them having certain market power and competitive advantage.

The paper provides useful implications for policies. The paper shows that fintech entry is mostly beneficial for clients by reducing fees. Fintech entry could even ease financial exclusion. However, in order to improve financial inclusion in the long-run, the fintech serving those neglected clients would need market power to be able to maintain an elevated price level that compensates the fintech for the additional risks. Competition sets incentives to trade-off inclusion against lower prices. Overall, the analysis in the paper can be used as a guide to study empirically which frictions are relevant for exclusion in payment markets in a country-specific setting.

The paper contributes to the literature that analyses retail payment systems based on the theory of two-sided markets pioneered in Rochet and Tirole (2003) by introducing operational risks into the model and studying financial exclusion as an outcome.[5] The model contributes a framework to explain financial exclusion from the perspective of the supply side, i.e., frictions originating from payment service providers. The idea of risk-driven exclusion is novel for payment markets although it has been studied in credit markets (Stiglitz and Weiss, 1981). The model thus also provides a novel justification for high merchant fees, often observed in emerging economies, and contributes to the discussion of why fees in the retail payment market are set at the expense of merchants rather than consumers (Guthrie and Wright, 2007; Wright, 2012; Bedre-Defolie and Calvano, 2013; Ding and Wright, 2017; Liu, Teh, Wright, and Zhou, 2019). The paper complements the work in Aurazo and Vasquez (2019) which studies tax evasion and the informal economy as explanations for demand-side driven exclusion from payment markets.

The paper also contributes to the growing literature analysing the effects of competition from

[5]See Evans and Schmalensee (2013) and Jullien and Sand-Zantman (2021) for reviews on competition policies in any two-sided market and Verdier (2011) and Rysman and Wright (2014) for reviews focused on payment markets.

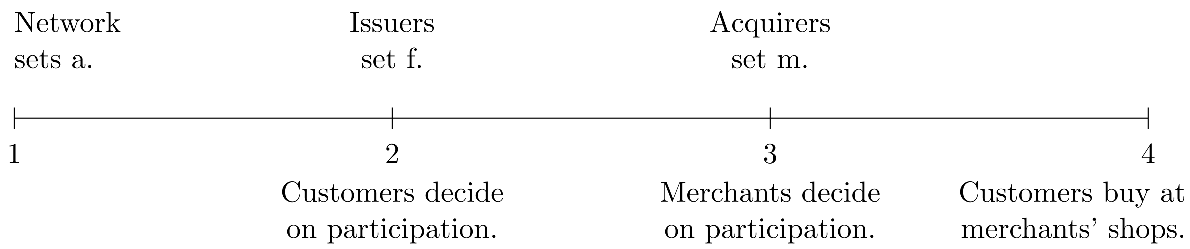

Figure 1: Timeline of the model.

fintech companies. With respect to payment markets, Jun and Yeo (2016) and Aurazo (2023) study the entry of fintech firms whose business models depend on the cooperation of traditional payment service providers. They show that traditional players can generate entry frictions at the expense of clients' welfare. Other theoretical work focuses on interaction between payment and credit markets. Parlour, Rajan, and Zhu (2022) and Ghosh, Vallee, and Zeng (2022) study how fintechs exploit information from payment services for lending decisions. However, these works do not study the effect of fintech entry on financial inclusion. Several empirical papers provide evidence that fintech entry was able to improve financial inclusion. For example, improved inclusion and development due to innovations in payment markets is documented for Kenya (Suri and Jack, 2016), India (Bandi, Moreno, Ngwe, and Xu, 2019), or Singapore (Agarwal, Qian, Ren, Tsai, and Yeung, 2020). Positive effects are also found in developed economies, like the US (Erel and Liebersohn, 2020; Reher and Sokolinski, 2020). In light of this empirical evidence, this work delivers a theoretical underpinning that connects fintech entry and improved financial inclusion with respect to the payment market.

2 The Model

The model is based on a four-party scheme payment network, as described in Rochet and Tirole (2002). The network is operated by a single payment card association which sets the interchange fee a, i.e., a fee associated to each transaction that the acquirer pays to the issuer. After the card network has set the interchange fee, the issuers set the customer fee f based on which consumers decide whether to hold a card or not. Subsequently, the acquirers set the merchant discount rate m and merchants decide whether to accept card payments or not. The network imposes a no-surcharge rule so that merchants have to sell their product at one price independent of the payment method used by the customer. After decisions about the payment network participation are made, real transactions take place when consumers shop at merchants' sites. The timeline is illustrated in Figure 1.

2.1 Merchants

Heterogeneity A crucial element to explain financial exclusion is to introduce heterogeneity. I focus on heterogeneity on the side of merchants. Specifically, I assume that there is a continuum of merchants and that they draw their quality τ from an underlying continuous and differentiable distribution function t with cumulative distribution T.

Merchants with higher quality are also larger merchants in terms of their potential card payment transaction volumes.[6] I denote $D(\tau_j)$ the sales volume of merchant j which is processed with cards and assume that

$$D(\tau_j) \leq D(\tau_{j+1}) \ . \tag{1}$$

Note that the total card volume depends on merchants' quality as well as the willingness of buyers to use cards, denoted as D_B. If more buyers use cards, the probability that any merchants' customers pay by card increases. Therefore $D'_{D_B} > 0$.[7]

The total transaction volume of the card payment market is

$$\int_{-\infty}^{\infty} D(\tau_j, D_B) dT \ . \tag{2}$$

Excluding merchants reduces the total transaction volume. Since the merchants with the lowest quality contribute the lowest transaction volumes, excluding low-quality merchants reduces the total transaction volume less than excluding high-quality merchants. We define the cut-off quality in the following.

Definition 1 (Cut-off quality)**.** *Let $\hat{\tau}$ be the lowest quality of all merchants who participate in the card payment market. The merchant with quality $\hat{\tau}$ is referred to as the marginal merchant.*

The transaction volume of the card payment market when merchants with qualities $\tau_j < \hat{\tau}$ are excluded is then

$$\int_{\hat{\tau}}^{\infty} D(\tau_j, D_B) dT \ . \tag{3}$$

Operational risks Merchant heterogeneity also defines the marginal risk that merchants pose to the payment system. I assume that total operational losses are distributed with a continuous and differentiable distribution function L with an expected loss denoted as EL. Let this distribution and its expected value depend on the quality of the underlying merchant pool. I assume that the marginal contribution of a high-quality merchant to expected losses is lower than the marginal contribution of a low-quality merchant. Formally,

$$EL'_{\tau}(\tau_j) \geq EL'_{\tau}(\tau_{j+1}) \ . \tag{4}$$

As the marginal contribution to risk is continuously decreasing as quality is increasing, expected losses decrease as low-quality merchants are excluded. Thus, the quality of the underlying

[6] A positive relationship between quality and sales would, for example, obtain in a vertically differentiated product market (see Appendix **??** for details).

[7] For the ease of notation, I use Y'_x as a shorthand for $\partial Y / \partial x$ throughout the paper.

merchant pool can be summarized using the cut-off quality $\hat{\tau}$. Expected losses are decreasing and convex in the cut-off quality, i.e.,

$$EL'_{\hat{\tau}} < 0\ , \quad EL''_{\hat{\tau}} > 0\ . \tag{5}$$

Payment market participation I assume that merchants incur costs $c_{s,j}$ per transacted volume in cash[8] and m per transacted volume in cards. The costs m is also called the merchant discount rate which is set by acquirers. I assume there are no further fees, such as an entry fee. Merchants accept cards only if it is at least as profitable as continuing with fully cash-based payment transactions. The payment market participation constraint of merchant j is given by

$$c_{sj} \geq m\ . \tag{6}$$

Merchants only accept cards if the marginal costs of cards is lower than the marginal costs of cash. For most of the following analysis, I assume that merchant heterogeneity also defines the costs of cash such that low-quality merchants have lower costs of cash than high-quality merchants.[9] Formally,

$$c_s(\tau_j) < c_s(\tau_{j+1})\ . \tag{7}$$

In summary, under these assumptions, low-quality merchants are set out to be less attractive for the payment market. They offer less transaction volume, are riskier, and have better outside options.

2.2 Buyers

Buyers purchase goods from merchants. They can do so with cash or card. For simplicity, I assume that buyers do not have to pay a membership fee to hold a card. Therefore, all buyers potentially hold a card. A buyer derives a benefit b_B from being able to pay by card but has to pay a customer fee f per transaction volume. The customer uses the card if and only if

$$\widetilde{b}_B \geq f. \tag{8}$$

where $\widetilde{b}_B$ is a random variable that results from the cumulative distribution function H defined as

$$H\left(b_B\right) = \Pr\left(\widetilde{b}_B \leq b_B\right) \tag{9}$$

with a continuous density function $h\left(b_B\right) = H'\left(b_B\right)$ and a monotone increasing hazard rate which is

$$\frac{h\left(b_B\right)}{1 - H\left(b_B\right)}\ . \tag{10}$$

The net benefit of paying by card is $\widetilde{b}_B - f$. Then, $D_B\left(f\right)$ is the demand for card payments

[8] This is usually described as the unit benefit of accepting a card payment.

[9] The importance of this assumption and the costs of cash are discussed in section **??**.

given that the store accepts cards and is denoted as

$$D_B(f) = \Pr\left(\widetilde{b}_B \geq f\right) = 1 - H(f) \ . \tag{11}$$

2.3 Issuers

The issuer faces a technological cost c_I and receives the interchange fee a per transacted value. The maximisation problem of an issuer is

$$\underset{f}{\text{Max}}\, \Pi_I = (f + a - c_I)\, D_B(f) \tag{12}$$

We assume that issuers operate in monopolistic competition.[10] Then, we get the following -known- result.

Lemma 1. *The optimal buyers' fee f^* for a given interchange fee a is defined in Eq. (??) where f^* is a unique optimum solution.*

$$f^* = (c_I - a) - \frac{D_B(f^*)}{D_B'(f^*)} \tag{13}$$

Proof. To show that Eq. (13) has a unique solution, we have to prove that $f + \frac{D(f)}{D'(f)}$ is monotone in f. We proof that both parts, f^* and $\frac{D(f)}{D'(f)}$ are monotone increasing and hence their sum is as well. With f^* it is straightforward. We know $\frac{D_B(f)}{D_B'(f)}$ is equal to:

$$\frac{D(f^*)}{D'(f^*)} = \frac{1 - H(f^*)}{-h(f^*)} = -\frac{1}{\dfrac{1 - H(f^*)}{h(f^*)}} = -\frac{1}{hazardrate} \tag{14}$$

where Eq. (14) represents the negative reciprocal of hazard rate function which is strictly monotone increasing since we know from Eq. (10) that the hazard rate is monotone increasing. Hence, we conclude that $f + \dfrac{D(f)}{D'(f)}$ is monotone increasing and therefore Eq. (13) has an unique solution. □

2.4 Acquirers

According to the timeline, acquirers set their prices after the interchange fee and buyers' fee have been determined. The acquirer incurs a unit cost of c_A and has to pay the interchange fee a to the payment network (or the issuer). Acquirers ask merchants to pay a unit cost m per transacted value, the merchant discount rate. As said before, there are no membership fees.

The profit function of an acquirer is

$$\Pi_A(\hat{\tau}, m) = \int_{\tau=\hat{\tau}}^{\infty} (m - c_A - a) D(\tau_j)\, dT - EL(\hat{\tau}) \tag{15}$$

[10] In a perfect competition model $(f + a - c_I)\, D_B = 0$, then $f^* = c_I - a$.

where the acquirer uses revenues from card transactions to cover expected losses due to operational risks. The acquirer can choose to exclude merchants from the payment network even though they might be willing to accept cards at price m. This assumption is in line with the practice that acquirers have to do Know-Your-Customer checks before enrolling new merchants.

Non-negative profits together with the necessary participation of merchants as in Eq. (6) imply that the merchant discount rate must be set within the limits of

$$c_A + a \leq m \leq \bar{c}_s \tag{16}$$

where $\bar{c}_s$ is the highest cost of cash that any merchant has. Otherwise, acquirers would not have a viable business model because they would either not be able to cover their marginal costs or would not have any clients. I impose an even stricter condition in order to ensure that acquirers' marginal costs are so low that they would be able to serve all merchants. Henceforth, I assume that

$$c_A + a \leq \underline{c}_s \tag{17}$$

where $\underline{c}_s$ is the lowest cost of cash that any merchant has. Otherwise, trivially merchants with very low costs of cash would be financially excluded because acquiring technology is not cost efficient enough to provide a benefit relative to cash payments.

3 Causes of financial exclusion

3.1 Cost-driven financial exclusion

As a benchmark for the following analysis, I first consider a perfectly competitive market without operational risks. Setting thus $EL = 0$, the zero-profit condition (ZPC) given as

$$\Pi_A = \int_{\tau=\hat{\tau}}^{\infty} (m - c_A - a) D(\tau_j)\, dT \ = 0 \tag{18}$$

which holds if

$$m = c_a + a \tag{19}$$

The merchant discount rate is fully determined by acquirers' marginal costs which are independent of the quality of merchants. Therefore, all merchants would accept cards, as long as merchants' cost of cash are higher than acquirers' marginal costs which is true if condition 17 holds. Consequently, there is full financial inclusion in a perfectly competitive acquiring market without risks.

Corollary 1 (Cost-driven financial exclusion). *If condition 17 does not hold, merchants with very low costs of cash would be excluded because acquiring technology is not cost efficient enough to provide a benefit relative to cash payments.*

In summary, perfectly competitive markets should not exclude any merchants except when merchants have extraordinarily low costs of cash. However, it is rather unrealistic to assume that costs of cash are lower than the cost of cards. Further, the notion underlying this paper is that digital payments are more efficient than handling cash.

3.2 Risk-driven financial exclusion

In the following, I consider a perfectly competitive market in the presence of operational risks. Rearranging the zero-profit-condition gives

$$m^{ZP}(\hat{\tau}) = \underbrace{c_A + a}_{\text{Marginal costs}} + \underbrace{\frac{EL(\hat{\tau})}{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT}}_{\text{Volume-weighted risk-markup}} \tag{20}$$

Acquirers charge a mark-up in addition to marginal costs in order to provision for expected losses. As a consequence, the equilibrium merchant rate depends on the level of exclusion. For now, we simplify the analysis by assuming that merchant participation is still feasible so that

$$m^{ZP}(\hat{\tau}) \leq \underline{c}_s \quad \forall\, \hat{\tau} \,. \tag{21}$$

Consequently, exclusion is solely defined by the acquirers choice to admit a merchant in the network.

The mark-up behaves non-linear with respect to the level of exclusion due to the convexity in EL: While initially excluding very low-quality merchants leads to a more pronounced drop in expected losses with only a moderate loss of transaction volume, excluding high-quality merchants has the opposite effect. The relationship between the optimal merchant rate and optimal level of exclusion that result in zero profits is illustrated in Figure 2 where the zero-profit-curve (ZP) depicts Eq. 20. Acquirers can set high merchant rates either to compensate for high risk at low levels of exclusion and high transaction volumes (point I), or to offset low transaction volumes at high level of exclusion and low risk (point II). Formally,

$$\frac{\partial m^{ZP}}{\partial \hat{\tau}} = \frac{EL'_{\hat{\tau}}}{\int_{\hat{\tau}} D(\tau_j) dT} + \frac{EL(\hat{\tau})\; D(\hat{\tau})}{(\int_{\hat{\tau}} D(\tau_j) dT)^2} \begin{cases} < 0 & \text{if} \quad \frac{-EL'_{\hat{\tau}}}{D(\hat{\tau})} < \frac{EL(\hat{\tau})}{\int_{\hat{\tau}} D(\tau_j) dT} & (i) \\ > 0 & \text{if} \quad \frac{-EL'_{\hat{\tau}}}{D(\hat{\tau})} > \frac{EL(\hat{\tau})}{\int_{\hat{\tau}} D(\tau_j) dT} & (ii) \\ = 0 & \text{if} \quad \frac{-EL'_{\hat{\tau}}}{D(\hat{\tau})} = \frac{EL(\hat{\tau})}{\int_{\hat{\tau}} D(\tau_j) dT} & (iii) \end{cases} . \tag{22}$$

There are many possible $(m, \hat{\tau})$-combinations that result in zero profits, as illustrated by the ZP-curve in Figure 2. However, among these candidate equilibria only one gives competitors no further incentive to deviate. At the local minimum of eq. 20 equivalent to case (iii) in eq. 22, the marginal revenue loss from excluding one more merchant equals the marginal safety gain. Reducing the merchant rate further in order to steel clients from competitors is not possible without making losses. Furthermore, neither excluding nor admitting an additional merchant could increase profits at point $(m^{PC}, \hat{\tau}^{PC})$.

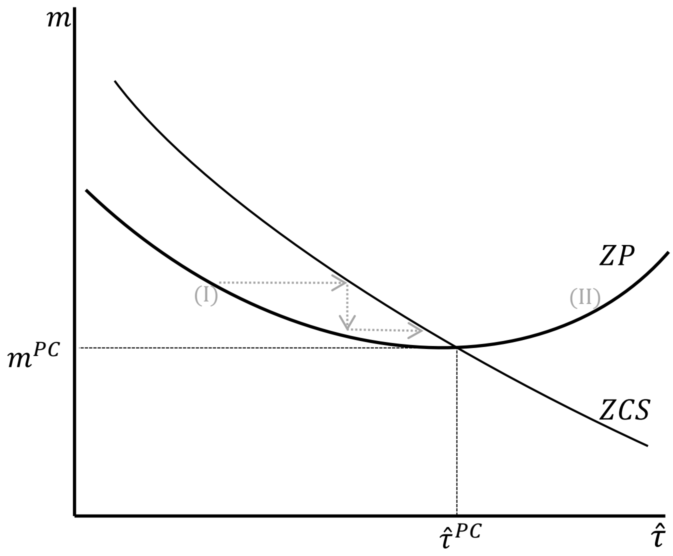

Figure 2: Merchant discount rate (m) and exclusion level ($\hat{\tau}$) in perfect competition.

In detail, the marginal effect of exclusion on acquirers' profit is

$$\Pi'_{\hat{\tau}} = \underbrace{-(m - c_A - a)D(\hat{\tau})}_{\text{Revenue loss}} \underbrace{-EL'_{\hat{\tau}}}_{\text{Safety gain}} \quad . \tag{23}$$

When keeping the merchant rate constant, an acquirer loses the revenue on the transaction volume of the marginal merchant who was excluded but gains safety by excluding this merchant's marginal contribution to risk. For each level of exclusion, one can define the merchant rate m^{ZCS} such that the marginal merchant's revenue covers exactly his marginal contribution to risk. Revenues of all other merchants included by the acquirer also cover their risk contribution, since at m^{ZCS} their transaction volumes are higher and their risks lower than those of the marginal merchant. Hence, there is zero cross-subsidization (ZCS) at m^{ZCS}. Formally,

$$\Pi'_{\hat{\tau}} = 0 \quad \Leftrightarrow \quad m^{ZCS} = c_A + a - \frac{EL'_{\hat{\tau}}}{D(\hat{\tau})} \tag{24}$$

m^{ZCS} is strictly decreasing as the quality of the marginal merchant increases. From eq. 20 and 22, we see that

$$\begin{cases} m^{ZP} < m^{ZCS} & \forall\, \hat{\tau} < \hat{\tau}^{PC} \quad (i) \\ m^{ZP} > m^{ZCS} & \forall\, \hat{\tau} > \hat{\tau}^{PC} \quad (ii) \\ m^{ZP} = m^{ZCS} & \forall\, \hat{\tau} = \hat{\tau}^{PC} \quad (iii)\ . \end{cases} \tag{25}$$

Rates below m^{ZCS} would not adequately price the additional risk that the marginal merchant imposes on the system. However, when $m^{ZP} < m < m^{ZCS}$ (e.g., at point I), revenue from higher-quality merchants can be used to offset expected losses on lower quality merchants. Thus, expected profits are still positive because risk is pooled between more and less risky merchants. Risk pooling takes place at exclusion levels $\hat{\tau} < \hat{\tau}^{PC}$.

Assuming the perfectly competitive equilibrium were in point I in Figure 2, acquirers could make positive profits by excluding the riskiest merchants (at constant m). However, competitors could then steel all clients by lowering the price. Thus, by following this strategy, the competitive equilibrium moves towards $(m^{PC}, \hat{\tau}^{PC})$. Similarly, if the equilibrium were in point II, acquirers could make positive profits by including more merchants, but competitive forces would then lower the price so that the competitive equilibrium moves towards $(m^{PC}, \hat{\tau}^{PC})$. The following proposition summarizes the above.

Proposition 1 (Risk-driven financial exclusion)**.** *If the acquiring market is perfectly competitive and operational risks are increasing as the quality of the marginal merchant falls, acquirers optimally exclude low-quality merchants from the card payment market.*

Proof. Follows from the above. □

3.3 Market power driven financial exclusion

In the following, we consider that the acquirer is a monopolist. The acquirer optimizes expected profit with respect to the merchant rate m. Marginal profits are

$$\frac{\partial \Pi^A}{\partial m} = -\Big((m - c_A - a)D(\hat{\tau})\Big)\frac{\partial \hat{\tau}}{\partial m} + \int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT\ . \tag{26}$$

The optimal merchant discount rate is thus

$$m^{MC} = c_A + a + \frac{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT}{D(\hat{\tau})\hat{\tau}'_m} \tag{27}$$

which is higher than the competitive price due to a markup. Rearranging, we have

$$\underbrace{\Big((m - c_A - a)D(\hat{\tau})\Big)\frac{\partial \hat{\tau}}{\partial m}}_{R(\hat{\tau})\text{: Loss from exclusion}} \quad = \quad \underbrace{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT}_{G(\hat{\tau})\text{: Gain from exclusion}} \tag{28}$$

where the left hand side represents the loss caused by the exclusion of the marginal merchant with quality $(\hat{\tau})$ and the right hand side shows the gain on all remaining merchants caused by an increase of the merchant discount rate.

Exclusion as defined in eq. 28 depends on the heterogeneity of merchants' cost of cash. So far, we assumed that marginal costs of acquirers are low enough to serve even the merchants with the lowest cost of cash (ref. eq. 17). The monopolist considers raising the markup so high that the monopoly price exceeds the cost of cash of some merchants. Otherwise, $\underline{c}_s$ would be the upper limit for the monopoly price with full inclusion. The heterogeneity of merchants' cost of cash is decisive. In principle, we can think of situation where cost of cash also depend on quality. For now, we assume a positive relation between quality and cost of cash, such that low quality merchants with low transaction volumes are also those that have low costs of cash and hence, the firsts to be excluded at higher costs of cards.

Intuitively, whereas losses of exclusion are initially low because very low quality merchants have only limited transaction volumes, gains are high because still a high number of merchants remains in the network on all of whom additional profits can be made. Gradually, the quality and hence transaction volume of the marginal merchant improves while the volume remaining decreases. Therefore, maximum profits are reached at an interim quality level. Formally, we can take the derivative of both sides with respect to the cut-off quality. Starting with the gain on the right hand side, we get

$$\frac{\partial G}{\partial \hat{\tau}} = -D(\hat{\tau}) \leq 0 \quad \forall\, \hat{\tau} \tag{29}$$

since $D(\tau) \geq 0$ for all qualities. From this, one can derive that the gains from exclusion are a monotone decreasing function of the cutoff quality. Taking the derivative on the left hand side of Eq. 28 with respect to the cutoff quality, we get

$$\frac{\partial R}{\partial \hat{\tau}} = (m - c_A - a)\frac{\partial D}{\partial \hat{\tau}}\frac{\partial \hat{\tau}}{\partial m} \geq 0 \quad \forall\, \hat{\tau} \tag{30}$$

since $D'(\hat{\tau}) > 0$ and $\hat{\tau}'(m) > 0$. Here, we see that the loss caused by the exclusion of the marginal merchant increases as the quality of the marginal merchants increases. Losses from exclusion are therefore a monotone increasing function of the cutoff quality.

Let us further consider the gain and loss from excluding a merchant with infinitesimal low quality. The gain would be

$$\lim_{\hat{\tau} \to -\infty} \int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT = \int_{-\infty}^{\infty} D(\tau_j)\, dT \geq 0 \tag{31}$$

and the loss is

$$\lim_{\hat{\tau} \to -\infty} \Big((m - c_A - a)D(\hat{\tau})\Big)\frac{\partial \hat{\tau}}{\partial m} = 0 \tag{32}$$

since the expected sales volume $D_s(\tau)$ becomes infinitesimal small for very low quality merchants. Hence, for extremely low qualities, losses from exclusion are close to zero while gains are positive and at their maximum. We follow that there is a unique cutoff at an interim quality level which defines the optimal level of exclusion.

Proposition 2 (Market power driven financial exclusion)**.** *If the acquiring market is a monopoly and merchants' cost of cash are increasing in quality, it is optimal for the monopolist acquirer to offer a merchant discount rate that excludes low-quality merchants from the card payment market.*

Proof. Follows from the above. □

In the following, we introduce operational risks to the monopolist's problem. The modified profit function then is

$$\Pi^A = \int_{\hat{\tau}}^{\infty} (m - c_A - a)D(\tau_j)\, dT - EL(\hat{\tau})\ . \tag{33}$$

The optimal merchant discount rate of a monopolist is thus

$$m^{MC} = c_A + a + \frac{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT}{D(\hat{\tau})\hat{\tau}'_m} - \frac{EL'(\hat{\tau})}{D(\hat{\tau})} \tag{34}$$

which is higher than the monopolist price without operational risks because the monopolist charges a markup on risk in addition to the competitive markup. Upon further inspection by comparing eq. 24 and 34, we see that the monopolist does not allow for cross-subsidization. Under the monopolistic price, the revenue from the marginal merchant more than compensates for the marginal contribution to risk of the marginal merchant. Regarding the exclusion trade-off under optimal conditions, we get

$$\underbrace{\Big((m - c_A - a)D(\hat{\tau})\Big)\hat{\tau}'_m}_{R(\hat{\tau})\text{: Loss from exclusion}} = \underbrace{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT - EL'(\hat{\tau})\hat{\tau}'_m}_{G(\hat{\tau})\text{: Profit and safety gain from exclusion}} \tag{35}$$

where the gain from exclusion encompasses not only the additional revenue that can be earned on the remaining clients of the acquierer at higher prices but also savings on provisions for expected losses which decrease as the average quality of the remaining clients increases. We summarize this finding in the following.

Corollary 2 (Monopoly and financial exclusion with operational risks)**.** *At a given optimal monopolistic merchant discount rate m^{MC}, the monopolist excludes more merchants when merchants additionally pose operational risks than when they do not pose any risk.*

Proof. By comparing the exclusion trade-off without risk (eq. 28) and with risk (eq. 35), for any given level of m, the loss from exclusion is equal while the gain from exclusion is higher due to the additional safety gain. Hence, optimal exclusion would be higher. □

4 Can Fintechs reduce financial exclusion?

In the following section, I analyse under which circumstances fintech competitors could potentially improve welfare. I assume fintechs (F) have lower marginal costs than the incumbent traditional acquirers (A), i.e.,

$$c_F < c_A \,. \tag{36}$$

4.1 Fintech competition without operational risks

When incumbent acquirers compete in perfect competition and there is no operational risk to take into account, a fintech could improve merchant welfare by offering lower merchant discount rates. In this scenario, absent other frictions, the fintech company should be able to dominate the payment market, giving fintechs certain market power over the incumbents. In fact, the problem can be solved as a Bertrand duopoly between a fintech and a traditional acquirer with asymmetric costs. In equilibrium, the fintech offers either a merchant rate equal to the competitors' marginal costs ($m_F = c_A$) or slightly below that ($m_F = c_A - \epsilon$) making a positive profit of $c_F - c_A - \epsilon$.[11] Thus, for the market to converge to lower prices in the long-term, a critical number of competing fintechs is necessary, all of which then compete at lower marginal costs than the traditional acquirers. In that case, $m_F = c_F$ and merchants' welfare defined as $c_{S,j} - m$ would increase by $c_F - c_A$.

The effect of fintech-entry on financial inclusion depends on the assumption in eq. 17. When eq. 17 holds, the technology of traditional acquirers is already cost efficient enough to serve all merchants. An even better technology would not make a difference. The market would stay fully inclusive. Trivially, if eq. 17 does not hold and some merchants are excluded because $c_{S,j} < c_A$, there could be a mass of merchants for which $c_F < c_{S,j} < c_A$. These merchants could then be served by the fintech acquirer.

In the case of a monopolist, fintech-entry similarly entails a Bertrand duopoly with asymmetric cost. As argued before, the merchant rate would drop just below the marginal costs of the former monopolist. Therefore, the markup that causes financial exclusion is eliminated and thus all merchants are included in the payment market. The fintech establishes a new monopoly at a lower price and with full financial inclusion. Additional fintech-entry would then lower the price to $m_F = c_F$ but cannot increase financial inclusion further.

Proposition 3. *Assuming that fintech acquirers have lower marginal costs than incumbent traditional acquierers ($c_F < c_A$) and traditional acquirers' technology is cost efficient enough to serve all merchants (condition 17 holds), competition from fintechs increases welfare of the merchants admitted in the payment system by lowering the merchant discount rate. In perfect competition this does not affect already full financial inclusion of merchants. However, when the fintech disrupts a monopoly, financial inclusion increases.*

Proof. Follows from the above. □

4.2 Fintech competition with operational risks

When merchants are associated with operational risks, fintechs also have to provision for potential losses from dealing with low-quality merchants. However, at any given level of exclusion, the fintech could outbid the incumbent acquirers by offering a lower discount rate due to its cost advantage. As illustrated in Figure 3, the zero-profit-curve of the fintech is shifted down. Incumbents can react by offering even lower rates but excluding more merchants. The fintech cannot profitably continue business with the low-quality merchants who are left out by the incumbents since this client pool needs the cross-subsidization from higher-quality merchants to be profitable. However, the lowest possible merchant rate that incumbent acquierers can offer is m^{PC}, therefore the fintech would win the market by offering $m^F = m^{PC} - \epsilon$. The shaded area in Figure 3 shows the price-exclusion combinations with which a fintech could profitably do business. The fintech optimizes profits by choosing the highest possible merchant rate along the zero-cross-subsidization-curve. The implied rate at which the marginal merchant's revenue cover the marginal contribution to risk (cf. eq. 35) is lower for the fintech than the incumbents

[11] See Tirole (1988) Ch. 5. I prefer the epsilon-equilibrium. Otherwise, acquirers that are indifferent between earning zero profits or exiting, would exit, which is at odds with the previous analysis.

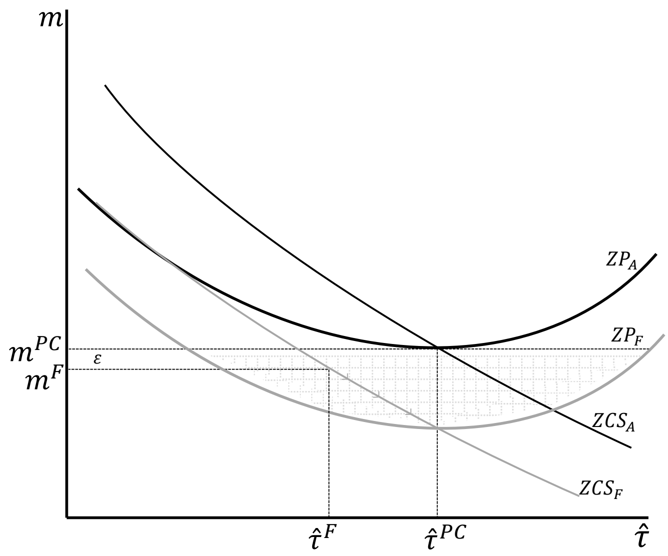

Figure 3: Fintech-entry into a perfectly competitive market.

at any exclusion level. Thus, the fintech includes more merchants at a merchant rate just below the competitive rate of the incumbents, i.e.,

$$\hat{\tau}^F < \hat{\tau}^{PC} \tag{37}$$

only if the fintech has market power. Interestingly, as more and more fintechs with cost advantages enter, competition forces the market equilibrium again towards the lowest possible merchant rate which would then be on the zero-profit-curve of fintechs. In a perfect competition equilibrium of fintechs, the merchant discount rate is

$$m_F = c_F + \frac{EL(\hat{\tau})}{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT} \tag{38}$$

the lowest possible rate the fintech can offer. Yet, for the fintech as for the incumbents, the only exclusion rate which gives no further incentives to deviate is defined by

$$\frac{EL(\hat{\tau})}{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT} = \frac{EL'(\hat{\tau})}{D(\hat{\tau})} \tag{39}$$

which is independent of acquirers' or fintechs' marginal costs.

If one or more fintechs enter into a market dominated by a monopolist, the same results follow since two players are sufficient to restore competition in a Bertrand setting (cf. Bertrand paradox).

Proposition 4. *If one fintech with competitive cost advantage ($c_F < c_A$) enters the market of traditional acquirers and merchants pose operational risks that are disproportionate to their* *quality, the fintech optimally chooses*

$$\begin{aligned} m^F &= m^{PC} - \epsilon \\ \hat{\tau}^F &< \hat{\tau}^{PC} \ . \end{aligned} \tag{40}$$

If many fintechs with competitive cost advantages enter a market of traditional acquierers and compete among themselves, fintechs' perfect competition equilibrium is defined as

$$\begin{aligned} m^F &= c_F + \frac{EL(\hat{\tau})}{\int_{\hat{\tau}}^{\infty} D(\tau_j)\, dT} < m^{PC} \\ \hat{\tau}^F &= \hat{\tau}^{PC} \ . \end{aligned} \tag{41}$$

These results obtain whether the market was previously in perfect competition or not.

Proof. Follows from the above. □

In other words, initial fintech entry can have beneficial effects on financial exclusion by including more and riskier merchants into the payment network. Yet, fintechs finance the additional risk with positive profits. The market is contestable through the traditional acquirers only up to the minimum price they can afford without making losses. At prices below m^{PC}, the fintech gains market power. As fintech entry progresses and fintechs start competing among themselves, the merchant rate falls further, while fintechs start excluding merchants again. Ultimately, the optimal level of exclusion under perfect competition is independent of the competitors' marginal costs giving rise to a trade-off between financial inclusion and market power.

5 Conclusions

Small- and medium-sized merchants often still operate payments purely in cash, especially in emerging economies. For some merchants, cash payments may be very cheap to handle or are associated with other benefits, such as the illicit possibility to avoid taxes. Some merchants might serve customers that do not have debit or credit cards. However, digital payment options, such as mobile money, have increasingly connected low-income individuals to digital payment solutions. Therefore, not being able to accept cards or other forms of digital payments can be a competitive disadvantage for small firms. Further, being connected to digital payment solutions offers small merchants growth opportunities, e.g., in online retail platforms.

The analysis shows that financial exclusion of merchants in the digital payment market can be driven by costs, market power, or risks. If the acquiring technology is too costly compared to cash, merchants might chose not to accept cards. This could be an explanation in remote rural areas where connecting merchants involves high fixed cost investments from acquiring companies Brunnermeier, Limodio, and Spadavecchia (2023). Fintechs that have a competitive cost advantage could then ease the financial exclusion of merchants.

Further, acquirers with market power have incentives to exclude small merchants by charging prices above costs of cash of small merchants. An acquirer with market power thus earns higher revenues on the merchants that continue in the payment network. Any competition, be it from fintechs or other companies, could help easing exclusion due to market power.

Finally, the paper is the first to consider operational risks in the context of payment networks. Many incidences of losses due to fraud, technical difficulties, or cyber attacks in the recent past highlight that operational risks are relevant in payment markets. In principle, each market participant creates potential risks. Merchants, for example, have to be vigilant that the point-of-sale terminals are free from fraudulent manipulation. Thus, acquirers have incentives to exclude merchants when they pose excessive risks relative to the revenue potential. As was shown, acquirers exclude risky merchants even in a perfectly competitive environment. Fintech companies can only improve financial exclusion, when they themselves have market power against the traditional acquirers.

The analysis reveals a complicated assessment of market power in payment markets. While a single firm with market power chooses to exclude merchants for the sake of profit, a fintech with market power uses additional profits to include more merchants. The analysis gives a useful framework to analyse financial inclusion in emerging economies.

References

Agarwal, S., Qian, W., Ren, Y., Tsai, H.-T., and Yeung, B. 2020. The real impact of fintech: Evidence from mobile payment technology. Ssrn working paper.

Aurazo, J. 2023. Interchange fees, access pricing, and sub-acquirers in payment cards. Working papers, FIT-IN TSE.

Aurazo, J. and Vasquez, J. 2019. Merchant's Card Acceptance: An extension of the Tourist Test for Developing Countries. Review of Network Economics, 18(2):109–139.

Bandi, C., Moreno, A., Ngwe, D., and Xu, Z. 2019. The effect of payment choices on online retail: Evidence from the 2016 Indian demonetization. Working Paper 19-123, Harvard Business School, Division of Research.

Bedre-Defolie, Ö. and Calvano, E. 2013. Pricing payment cards. American Economic Journal: Microeconomics, 5(3):206–231.

Brunnermeier, M., Limodio, N., and Spadavecchia, L. 2023. Mobile money, interoperability and financial inclusion. Technical report.

Ding, R. and Wright, J. 2017. Payment Card Interchange Fees and Price Discrimination. Journal of Industrial Economics, 65(1):39–72.

Erel, I. and Liebersohn, J. 2020. Does fintech substitute for banks? evidence from the paycheck protection program. Working Paper 27659, National Bureau of Economic Research.

Evans, D. S. and Schmalensee, R. 2013. The antitrust analysis of multi-sided platform businesses. Working Paper 18783, National Bureau of Economic Research.

Ghosh, P., Vallee, B., and Zeng, Y. 2022. Fintech lending and cashless payments. Working paper.

Guthrie, G. and Wright, J. 2007. Competing payment schemes. The Journal of Industrial Economics, 55(1):37–67.

IMF. 2020. The promise of fintech financial inclusion in the post covid-19 era. Technical Report 20/09, International Monetary Fund. Monetary and Capital Markets Department, Report by Sahay, Ratna and von Allmen, Ulric Eriksson and Lahreche, Amina and Khera, Purva and Ogawa, Sumiko and Bazarbash, Majid and Beaton, Kim.

Jullien, B. and Sand-Zantman, W. 2021. The economics of platforms: A theory guide for competition policy. Information Economics and Policy, 54.

Jun, J. and Yeo, E. 2016. Entry of fintech firms and competition in the retail payments market. Asia-Pacific Journal of Financial Studies, 45:159–184.

Liu, C., Teh, T.-H., Wright, J., and Zhou, J. 2019. Multihoming and oligopolistic platform competition. Working paper.

Parlour, C., Rajan, U., and Zhu, H. 2022. When fintech competes for payment flows. The Review of Financial Studies, 35(11):4985–5024.

Philippon, T. 2019. On fintech and financial inclusion. Working Paper 26330, National Bureau of Economic Research.

Reher, M. and Sokolinski, S. 2020. Does fintech democratize investing? Working paper.

Rochet, J.-C. and Tirole, J. 2002. Cooperation among competitors: Some economics of payment card associations. Rand Journal of Economics, 33(4):549–570.

Rochet, J.-C. and Tirole, J. 2003. An economic analysis of the determination of interchange fees in payment card systems. Review of Network Economics, 2:69–79.

Rysman, M. and Wright, J. 2014. The economics of payment cards. Review of Network Economics, 13(3):303–353.

Stiglitz, J. E. and Weiss, A. 1981. Credit rationing in markets with imperfect information. The American Economic Review, 71(3):393–410.

Suri, T. and Jack, W. 2016. The long-run poverty and gender impacts of mobile money. Science, 354(6317):1288–1292.

Tirole, J. *The theory of industrial organization.* MIT press, 1988.

Verdier, M. 2011. Interchange fees in payment card systems: A survey of the literature. Journal of Economic Surveys, 25:273–297.

Wright, J. 2012. Why payment card fees are biased against retailers. RAND Journal of Economics, 43(4):761–780.