THE GRADUATE INSTITUTE GENEVA

INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

Graduate Institute of International and Development Studies
International Economics Department
Working Paper Series

Working Paper No. HEIDWP19-2022

A Mixed Duopoly in Interbank Payment Services

Carlos A. Arango-Arango
Banco de la República de Colombia

Yanneth Rocío Betancourt-García
Banco de la República de Colombia

August 2022

Chemin Eugène-Rigot 2
P.O. Box 136
CH - 1211 Geneva 21
Switzerland

A Mixed Duopoly in Interbank Payment Services

Carlos A. Arango-Arango Yanneth Rocío Betancourt-García[*]

Banco de la República de Colombia

Abstract

In this paper, we analyze theoretically the coexistence of two means of payment, such as cash and digital or electronic payments, introducing some distortions in the payments markets to understand the widespread use of cash, specially in emerging countries. Lagos and Wright (2005) theoretical approach allows us to model explicitly the frictions in the exchange process considering money as essential. We introduce in this framework theft and informality (measured by tax evasion) as factors affecting cash usage and competition with a private digital payment platform. Considering heterogeneity in the seller's side by assuming different levels of productivity we find the factors that explain the use of cash or digital payments. If a public provider enters the market with a less expensive platform the fees charged by the private provider have to be adjusted to the cost level of the public platform, decreasing the use of cash in the economy.

Keywords: Cash, means of payments, payments services, digital payments, instant payments.

JEL Classification: E40, E41,E42,E44

[*]Principal Researcher and Advisor, Department of Payment Systems and Banking Operations, Banco de la República de Colombia. The authors thank Cyril Monnet from the University of Bern and the Study Centre Gerzensee for the academic supervision of this paper, as well as Cedric Tille and Jean Charles Rochet for their valuable comments and suggestions. This research took place through the visiting program under the Bilateral Assistance and Capacity Building for Central Banks (BCC), financed by SECO, and the Graduate Institute in Geneva. The views expressed in this paper are solely those of the authors and do not necessarily reflect those of the Banco de la República de Colombia.

Introduction

Rapid improvements in information technology and communications have led to a growing transformation in financial services, especially payment systems have undergone substantial innovation over the years. New payment methods, platforms, and channels have been introduced, and for retail payments, innovations have addressed the entire value chain closing service gaps and increasing the speed of payments.

The payments industry is experiencing a shift driven by the need for faster, cheaper, and more convenient means of making payments. In many countries, central banks and monetary authorities have recently promoted faster payment systems (FPS) to provide a real-time, 24/7 fund-transfer facility (BIS, 2016). Over 60 jurisdictions have already launched FPS and a number of countries are planning to implement such systems in the coming years (CPMI-FPS, 2021).

In the recent past years, digitalization in payments has increased rapidly, and the Covid-19 pandemic accelerated this shift, especially in the retail sector. Although strong growth in FPS volumes was observed in 2020 in some Latin-American countries, such as Mexico, Brazil, Colombia, Chile, and Perú, some problems still prevail in the provision in terms of access, use, and acceptance of electronic instruments (FIS, 2020).

Despite these advancements, cash continues to dominate as the default mode of payment in the low-value retail segment, due mainly to a lack of financial inclusion or the absence of alternative, secure, low-cost payment methods. Around the world, 1.7 billion of adults are tied to cash as their only means of payment, as they do not have a transaction account (World Bank, 2017). In Colombia, around 75% of individuals prefer cash to other payment instruments (Banco de la República, 2022), and the demand for cash (real pesos per capita) has grown annually above 5 % in the last 15 years (Arango-Arango et al, 2020).

The existence of market failures in the provision of digital payments, the lack of interoperability between platforms and segmentation in the markets, have maintained the dominance of cash. All this, added to the benefits of using cash, such as those related to tax evasion, could lead cash to continue being preferred in transactions mainly in the retail sector, despite the existing of certain costs like theft. These distortions and inefficiencies in the payments market can lead to a mix of payments that does not correspond to the socially optimal provision.

As payment services are essential to the well-functioning of the economy, governments (and central banks) face key policy questions. In order to guarantee a safe, efficient, and relevant payment system they have to weigh their role in the regulation of the payments ecosystem against their involvement in the provision of payment infrastructures. In this sense, they are asking if public provision of new payments technologies may fill up the gaps left by the private sector, to bring the economy closer to the public policy objectives. Indeed, a public provider may promote competition, innovation, and the entry of new players in the payment's ecosystem.

Many emerging economies have opted for the public provision of FPS as a bet to leverage on interoperability, network effects and scale economies at the same time as solving coordination failures in the market for faster payments. This is the case of Mexico with its FPS PIX, which in only one year, was able to bring onboard two thirds of its adult population (Duarte et al., 2022). In other jurisdictions the private sector has developed instant payment solutions based on closed networks that create a fragmented market of walled mobile payment gardens. This has been the case in Colombia in the resent past, although, in 2018, the private sector launched an FPS Transfiya. However, this system, as many private FPS in other economies, has a limited number of payment service providers as participants, lacks interoperability with other payment rails, is limited to few use cases such as person to person and hardly make inroads into other payment flows such as person to merchants, excluding a large share of retail transactions (Arango-Arango, Ramirez-Pineda y Restrepo-Bernal, 2021).

The case for public provision of an FPS infrastructure is relevant as a policy mechanism to induce de private sector to enhance its fast payment solutions. The competition of private and public infrastructures it is not foreign to the provision of payment services. For large value payment systems, in Europe the public utility TARGET competes with EURO1, the private one, and in the US Fedwire (public) competes with CHIPS (private). In the case of retail payment systems, the Fed has announced the implementation of an FPS (FedNow) that would compete with RTP, the private FPS launched by the Clearing House. This is also the case in Colombia, where CENIT, the public ACH owned by the central bank, competes with ACH Colombia, the private provider of interbank transfers in batch.

This paper gives a theoretical basis and key insights to the discussions regarding public provision of new payment services, when the market is already served by private suppliers. It also makes an important contribution to the understanding of how public provision shall improve the payments ecosystem in terms of safety and efficiency, increasing the welfare of final users by lifting the existing inefficiencies.

General Framework

An accurate framework to study these topics is the New Monetarism Theory, which is a new macroeconomic approach that considers micro-foundations for institutions that facilitate the process of exchange, like money, banks, financial intermediaries, etc. This theory makes explicit the frictions in trade that makes money essential. A relatively new sub-branch of this theory studies the economics of payments, involving the study of payments systems, particularly among financial institutions, giving the theoretical tools to analyze frictions in payments markets.

Following this approach, we introduce in this paper another means of payment, in addition to cash, which could be interpreted as a digital or electronic payment. The coexistence of two or more means of payments depends on the costs and benefits that each one has for the different agents in the economy. In this way, agents (buyers or sellers) may prefer one means of payment to another depending on their net benefits, which determines the level of transactions carried out by cash or by electronic payments.

In this sense, agent's payment decisions are affected by some factors like theft and informality (measured by tax evasion), that are involved in cash transactions, and affect the general equilibrium outcome and social welfare. In this paper, theft is introduced as a cost for sellers to accept cash payments because they can be robbed with some probability. Tax evasion could be another important issue for cash payments, especially in emerging economies, where some merchants could prefer to remain in the informal sector to evade taxes. In this context, tax evasion could be interpreted as a measure of the informality level in the economy.

Heterogeneity between agents is also important to understand why people choose one means of payment or another to make their payments, depending on their own characteristics or preferences. In this paper, we consider heterogeneity on the sellers' side, by assuming that they have different levels of productivity according to their size. Small merchants, who could have a low level of productivity, may prefer to accept cash payments because they are less expensive or because of the possibility of evading taxes. Considering all these factors and assumptions in the Lagos and Wright setup, we can identify the inefficiencies and distortions in the payments market by comparing the first-best solution with the equilibrium solution in each market (one for cash and another for electronic payments).

Finally, when there is no charge to the sellers or buyers for using the digital platform, it could be interpreted as the case when a public provider offers these services as a free public good. By comparing these equilibrium conditions with the case when there is a private monopoly that maximizes its profits, given a cost of provision of the platform and a fee charged to the sellers, we find the conditions that must be satisfied to have a mixed duopoly in the provision of payment systems.

Basic Setup

The benchmark model has the following main assumptions, which are shown in Figure 0.1. Time is discrete and continues forever $t = 0, 1, ...$ and all agents discount the future at the rate β.

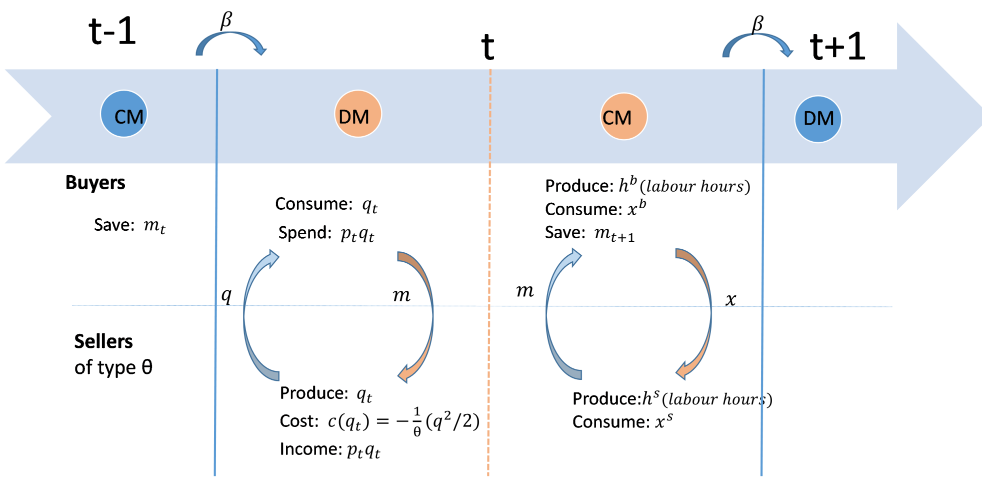

Fig. 0.1: Basic Setup

There is measure one of both buyers and sellers. Each period is divided into two subperiods (day and night) and different activities take place in each one. During the day, trade occurs in a decentralized market (DM). Some agents can produce but do not want to consume (known as sellers). They produce a good q, which is specialized and non-storable, at a cost $c(q;\theta)$, where θ is a measure of their productivity: The higher θ the less costly it is to produce the DM good. Also, it is increasing and convex: $c'(q;\theta) > 0, c''(q;\theta) > 0$ and $c(0;\theta) = 0$. For example, $c(q;\theta) = \frac{1}{\theta} q^2/2$. We assume θ is uniformly distributed over the interval $[0, 1]$. Other agents in this market want to consume but cannot produce (known as buyers). The utility they get by consuming the good is $u(q)$. The utility function is increasing and concave, $u'(q) > 0$, $u''(q) < 0$ and $u(q) = 0$. In each period, the DM price p_t is determined in a Walrasian market.

During the night, there is a centralized market (CM) where both sellers and buyers can produce and consume. Agents produce a general good, which is not storable x. They have a linear utility function $U(x, h) = x - h$. Agents have endowed with labour h and the production function is linear, so one unit of labour produces one unit of good $h = x$. Agent's utility functions are separable across subperiods.

The central bank provides money, which is intrinsically useless, divisible and storable, in a quantity $M \in \mathbb{R}_+$. We denote by ϕ the value of money in terms of the good q. So that, the real value of money in each period is given by: $\phi_t M_t$. Also, when the nominal price of the DM good paid by buyers is p_t, the real price is given by $\phi_t p_t$. The central bank adjusts the money supply M_t by making transfers to buyers in the CM. The money supply grows at a rate γ, so that in a stationary equilibrium the amount of real balances is constant $\phi_t M_t = \phi_{t+1} M_{t+1}$, and inflation is given by $\phi_t / \phi_{t+1} = \gamma$.

First Best solution

The first best solution is obtained by solving the following optimization problem for the society. Since preferences are lineary in the CM for all agents, all the redistribution effects do not have any impact on aggregate social welfare. So the planner's problem is given by:

$$\max_{\{q(\theta)\}} u\left(\int q(\theta) d\theta\right) - \int \left[\frac{1}{\theta} \frac{q\left(\theta\right)^2}{2}\right] d\theta.$$

From the first order conditions with respect to an arbitrary $q(\theta)$, we obtain,

$$\theta u'\left(q^*\right) = q^*\left(\theta\right)$$

where $q^* = \int q^*(\theta) d\theta$. Given that θ has a uniform distribution $\int_0^1 \theta d\theta = \frac{1}{2}$ and integrating the first order condition over all θ, we find the optimal quantity q^* from the following equation:

$$\frac{u'\left(q^*\right)}{2q^*} = 1 \tag{0.1}$$

which implies that the marginal utility for all the buyers have to be equal to the average marginal cost of the sellers (given that they do not have the same cost). So that, the disutility generated by the production of the good must be equal to the utility that it gives to other agents in the society.

The Model: Two means of payments with frictions

We first consider the benchmark case where the digital means of payment and cash are provided by a public authority at no cost. Then we consider the private provision of the digital means of payment by a monopolistic firm. Finally, we discuss the mixed duopoly case where a private firm would compete with a public provider who would price its service at marginal cost.

1 First Setup: Public provision

We assume buyers are anonymous and lack commitment. Therefore sellers requires a means of payment in the DM. In the basic setup, we introduce two means of payment, one is cash that is issued by the central bank and the other could be interpreted as an electronic or digital payment. The digital payment works like a debit card: sellers have to incur a cost in order to acquire the terminal while buyers can get a card for free. A buyer who meets a seller with a terminal in the DM can swipe the card for the amount of the purchase. Then the digital card company can enforce the repayment of the buyer in the following CM. Therefore while buyers are constrained by their cash holding when they use cash (they face a cash in advance constraint), they are not constrained when they pay digitally.

For simplicity, we assume that there is a market for each means of payment in which the other payment instrument is not accepted. This means that, e.g. a buyer holding cash cannot trade in the market where only the digital payment instrument is accepted. However, in each market, buyers and sellers maximize their value functions to choose the optimal quantity of goods they want to buy or sell given the amount of means of payment they hold. We assume some technological constraints on each means of payment: Cash payments are subject to theft, unlike digital payments. However, cash payments can be hidden and therefore it can be used to evade taxes, while agents cannot hide digital payments. (see Figure 1.1).

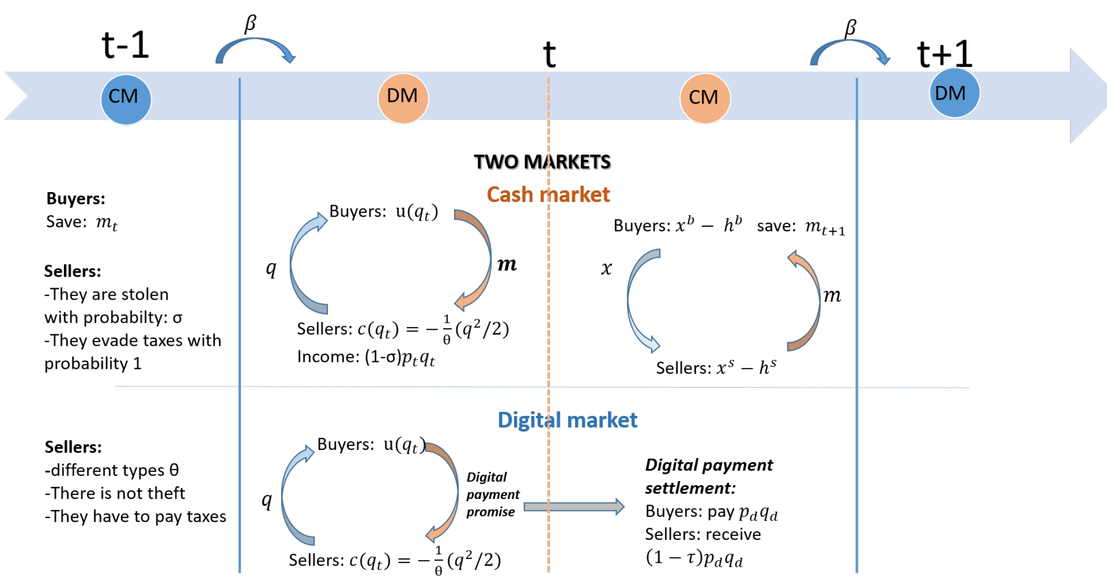

Fig. 1.1: Two means of payments with frictions

To be precise, we assume that cash can be stolen from sellers with a probability σ at the end of the day. Therefore they enter the decentralized market in the next period with a fraction $(1-\sigma)$ of the money they obtained in the previous period.

We model tax evasion as a benefit of cash transactions, assuming that with probability ε sellers can evade and not pay taxes (without being caught) when they accept cash. Otherwise, by accepting digital payments, sellers have to pay taxes τ with probability 1.

We suppose buyers have no cost of accessing the digital payment. Therefore, buyers will always hold the digital payment instrument. Sellers however, can only accept one means of payment in the DM, and they choose which one to accept. We guess and verify that all sellers with productivity level below a threshold $\bar{\theta}$ prefer to accept only cash, while all sellers with a higher productivity prefer to accept digital payments. This translates in an endogenous market segmentation for payments: A fraction $\pi = \bar{\theta}$ of sellers accept cash while the remaining $1-\bar{\theta}$ only accept digital payments. $\bar{\theta}$ is given by the productivity level that leaves the seller with that productivity level indifferent between accepting cash or digital payments, so that the profits for this seller is the same in both markets.

It will be useful to define the mean productivity in both cash and digital payment markets. In the cash only market, the mean productivity is $\int_0^{\bar{\theta}} \theta d\theta = \frac{\bar{\theta}^2}{2}$ while in the market with digital payments only, it is $\int_{\bar{\theta}}^1 \theta d\theta = \frac{\left(1-\bar{\theta}^2\right)}{2}$.

The cash market:

- Buyers' problem is such that they choose the optimal amount of money to maximize the value function in the centralized market $W(\tilde{m})$, considering the value function in the decentralized market, where they can trade with probability π in the cash market and with probability $(1-\pi)$ in the digital market.

$$W(\tilde{m}) = \max_{x,h,m} x - h + \beta \left[\pi V_{cash}(m) + (1-\pi) V_{digital}(m)\right]$$

s.t.

$$x - h = \phi\tilde{m} - \phi m$$

where:

$V_{digital}(m)$ is the value function in DM next period when they trade with digital payments:

$V_{digital}(m) = \max_{q_d} \quad u(q_d) - \phi p_d q_d + W(m)$

$V_{cash}(m)$ is the value of cash in DM next period and is given by:

$V_{cash}(m) = \max_{q} \quad u(q) + W(m - pq)$

subject to $\phi pq \leq \phi m$

Using quasi-linearity, we can write: $W(m - pq) = -\phi pq + W(m)$.

And from the first order condition with respect to q we know that:

$$u'(q) - (1+\lambda)\phi p_{t+1} = 0 \tag{1.1}$$

Substituting these expressions in $W(\tilde{m})$, we can solve the problem for m and get the following first order condition,

$$-\phi_t + \beta\left(\pi\lambda\phi_{t+1} + W'(m)\right) = 0$$

where the marginal value of unspent cash is $W'(m) = \phi_{t+1}$. Substituting λ from 1.1, we get the Euler-equation in the cash market:

$$\frac{\phi_t}{\beta\phi_{t+1}} = 1 + \pi\left(\frac{u'(q_{t+1})}{\phi_{t+1}p_{t+1}} - 1\right) \tag{1.2}$$

- Sellers' problem: given that they face theft with a probability σ when they accept cash, they only receive a fraction of what buyers paid $(1-\sigma)\phi pq$. They can also evade taxes with probability ε if they receive cash. So that, they receive a fraction $(1-\tau(1-\varepsilon))$ of the cash payment. The seller's problem is the same no matter which event comes first. Given the real price in the DM is ϕp, a seller of type θ supplies q to maximize the following expression:
$$\max_q (1-\sigma)(1-\tau(1-\varepsilon))\phi pq - c(q;\theta)$$

Using $c(q) = \frac{1}{\theta}q^2/2$, we get the optimal supply:

$$q(\theta) = \theta(1-\sigma)(1-\tau(1-\varepsilon))\phi p \tag{1.3}$$

Therefore more productive sellers (high values of θ) are also "larger" in the sense that they have more sales.

The total supply in this market is given for the aggregate quantities $q(\theta)$ offered by the sellers of low type $\theta \leq \bar{\theta}$:

$$\int_0^{\bar{\theta}} q(\theta)\,d\theta = (1-\sigma)(1-\tau(1-\varepsilon))\phi p \int_0^{\bar{\theta}} \theta d\theta = \frac{\bar{\theta}^2(1-\sigma)(1-\tau(1-\varepsilon))\phi p}{2}$$

- Market clearing requires that the above expression equals total demand q_t and solving for the (real) price we obtain:

$$\phi_t p_t = \frac{2q_t}{\bar{\theta}^2(1-\sigma)(1-\tau(1-\varepsilon))} \tag{1.4}$$

- Substituting this equation in the Euler equation (1.2) we get the equilibrium allocation q in the cash market:

$$\frac{\gamma}{\beta} = 1 + \bar{\theta}\left[\frac{u'(q)}{2q}\bar{\theta}^2(1-\sigma)(1-\tau(1-\varepsilon)) - 1\right] \tag{1.5}$$

where we have used $\pi = \bar{\theta}$ and in steady-state we have: $q_t = q_{t+1} = q$ and $\gamma = \frac{\phi_t}{\phi_{t+1}}$.

Rewritten this expression we have:

$$\frac{u'(q)}{2q} = \frac{\gamma - \beta\left(1-\bar{\theta}\right)}{\beta\bar{\theta}^3(1-\sigma)(1-\tau(1-\varepsilon))} \tag{1.6}$$

The digital market:

- In the digital market, buyers' maximizes their value function in DM to optimize the quantity of the good demanded in this market q_d.

$$\max_{q_d} u(q_d) - \phi p_d q_d$$

with first order condition,

$$u'(q_d) = \phi p_d \tag{1.7}$$

- Sellers in the digital market do not face theft $\sigma = 0$, but they cannot evade taxes $\varepsilon = 0$: They pay taxes τ for sure to the government and they receive only a fraction of the payment $(1-\tau)\phi p_d q_d$. In the DM with digital payment a seller with productivity θ maximizes the following objective function :

$$\max_{q_d}(1-\tau)\phi p_d q_d - \frac{1}{2\theta}q_d^2$$

from which we obtain the following first order condition

$$q_d(\theta) = \theta(1-\tau)\phi p_d \tag{1.8}$$

Therefore, the total supply in the digital market from the sellers with high productivity (high type $\theta \geq \bar{\theta}$):

$$\int_{\bar{\theta}}^{1} q_d(\theta)\, d\theta = (1-\tau)\phi p_d \int_{\bar{\theta}}^{1} \theta d\theta = (1-\tau)\phi p_d \frac{\left(1-\bar{\theta}^2\right)}{2}$$

- Using market clearing and equalizing this expression with the total demand q we obtain

$$\phi_t p_{d_t} = \frac{2q_{d_t}}{\left(1-\bar{\theta}^2\right)(1-\tau)} \tag{1.9}$$

- Substituting this equation in the Euler equation we get the equilibrium allocation q_d in the digital market:

$$\frac{u'(q_d)}{2q_d} = \frac{1}{\left(1-\bar{\theta}^2\right)(1-\tau)} \tag{1.10}$$

Finding $\bar{\theta}$:

To find the level $\bar{\theta}$ where sellers are indifferent between accepting cash or digital payments we equalize the sellers' profits in both markets:

- Profits in cash market: $(1-\sigma)(1-\tau(1-\varepsilon))\phi pq - \frac{1}{2\theta}q^2$
- Profits in digital market: $(1-\tau)\phi p_d q_d - \frac{1}{2\theta}q^2$

substituting the optimal quantities q and q_d in each market from equations (1.5) and (1.9) and ϕp, ϕp_d from equations (1.4) and (1.8), we obtain[1]:

$$\bar{\theta} = \sqrt{\frac{q}{q+q_d}}. \tag{1.11}$$

The equilibrium conditions in this setup with a digital platform without cost, heterogeneity in the sellers' side, theft of cash and tax evasion are given by:

1. Stationary equilibrium allocation in the cash market (equation 1.6):

$$\frac{u'(q)}{2q} = \frac{\gamma - \beta\left(1-\bar{\theta}\right)}{\beta\bar{\theta}^3(1-\sigma)(1-\tau(1-\varepsilon))}$$

2. Stationary equilibrium allocation in the digital market (equation 1.10):

$$\frac{u'(q_d)}{2q_d} = \frac{1}{\left(1-\bar{\theta}^2\right)(1-\tau)}$$

3. Optimal threshold of sellers in each market (equation 1.11):

$$\bar{\theta} = \sqrt{\frac{q}{q+q_d}}$$

Comparing these results with the optimal condition (equation 0.1) we find that the frictions introduced in our model creates some distortions in the equilibrium allocations for both markets. In the cash market, the distortions are given by inflation γ, theft σ and tax evasion ε; while in the digital market only taxes have a direct impact τ on the equilibrium allocations. Moreover, given that there is segmentation in the seller's side, these frictions have an indirect impact in the traded quantities in both markets via the level of indifferent sellers between the two markets $\bar{\theta}$.

The equilibrium allocation in the cash market q is affected positively by the amount of taxes τ and the probability of evasion ε, so that if taxes increase and the probability of evasion is high, the sellers prefer to accept cash and the quantity traded in this market increases. The impact of an increase in inflation γ and theft σ on the equilibrium allocation in the cash market is negative because

[1] So that: profits in cash are given by $\frac{2\bar{\theta}q^2}{\theta^4}$, and profits in digital market are: $\frac{2\bar{\theta}q_d^2}{\left(1-\bar{\theta}^2\right)^2}$

these are costs for the sellers of accepting cash. Finally, if the level of sellers who are indifferent between the two means of payment is high, so that more sellers prefer to accept cash than the digital payment, then the quantity traded in the cash market increases.

In the digital market, the equilibrium allocation is affected negatively by the level of taxes and the amount of sellers that accept cash. It means that if taxes are high less sellers will accept digital payments and the quantity traded decreases. The same happens if more sellers prefer to accept cash than the digital mean.

The impact of inflation and theft on the level of sellers that are indifferent between the two markets $\bar{\theta}$ is negative. It means that less sellers accept cash when the costs of inflation and theft are high. In contrast, when the level of taxes and the probability of evasion are high, sellers prefer to accept cash, then the quantity of sellers accepting cash increases.

2 Second Setup: Private monopoly provision

Next, we suppose that there is only one private provider of the digital payment platform, who charges a positive fee $\psi > 0$ directly to final users. We assume that the buyers do not face any cost to use the digital platform, so that all of them have access to both means of payments (cash and digital), which implies that there is not distortion in the buyers' side. In this case, only sellers who want to access to this platform are charged with the fee. Now, the existence segmentation in the payments market is affected by the platform's fee, so that the amount of sellers who accept cash depends on this fee $\bar{\theta}(\psi)$. For simplicity, we assume that the digital platform's fee does not depend on the quantity that is traded in this market q_d, so it is fixed and maximizes the monopolist's profits.

The problem for buyers and sellers in the cash and the digital market in terms of the quantities traded does not change. So the equilibrium conditions in both markets are maintained (equations 1.6 and 1.10), only the number of sellers in each market changes $\bar{\theta}$, because the sellers' profits in the digital market are affected directly by the fee and both market's profits are affected indirectly by $\bar{\theta}(\psi)$.

The profit in the cash market is

$$(1-\sigma)(1-\tau(1-\varepsilon))\phi pq - \frac{1}{2\theta}q^2 = \frac{2\bar{\theta}(\psi)q^2}{\bar{\theta}(\psi)^4}$$

while the profit in the digital market is

$$(1-\tau)\phi p_d q_d - \frac{1}{2\theta}q_d^2 - \psi = \frac{2\bar{\theta}(\psi)q_d^2}{\left(1-\bar{\theta}(\psi)^2\right)^2} - \psi$$

Equalizing both expressions we obtain:

$$\psi = \frac{2\bar{\theta}(\psi) q_d^2}{\left(1-\bar{\theta}(\psi)^2\right)^2} - \frac{2\bar{\theta}(\psi) q^2}{\bar{\theta}(\psi)^4}$$

From the Euler's equation in both markets we substitute q and q_d to get:

$$\psi = \frac{\bar{\theta}(\psi)}{2}\left[(1-\tau) u'(q_d) - \frac{\bar{\theta}(\psi)\beta(1-\sigma)(1-\tau(1-\varepsilon)) u'(q)}{\gamma - \beta\left(1+\bar{\theta}(\psi)\right)}\right]^2 \tag{2.1}$$

Monopoly problem:

The private platform chooses a fee ψ to charge to sellers of type $\theta \geq \bar{\theta}$, such that it maximizes its profit. The platform faces some costs related to the provision of the infrastructure and all the services needed to make digital payments (as the provision of terminals and some operating costs). From equation (2.1), we have an expression for ψ in function of $\bar{\theta}$, so that solving the following problem for $\bar{\theta}$ is equivalent to solve for ψ:

$$\max_{\bar{\theta}} \left(1-\bar{\theta}\right)(\psi - cost)$$

$$s.t. \psi = \frac{\bar{\theta}(\psi)}{2}\left[(1-\tau) u'(q_d) - \frac{\bar{\theta}(\psi)\beta(1-\sigma)(1-\tau(1-\varepsilon)) u'(q)}{\gamma - \beta\left(1+\bar{\theta}(\psi)\right)}\right]^2$$

From the first order condition with respect to $\bar{\theta}$ and equation 2.1. we find an expression for the optimal fee $\hat{\psi}$, which is a polynomial in $\bar{\theta}$ and depends on all the parameters of the model. By making some assumptions for the parameters[2] and taking a CRRA utility function[3], we solve this expression for $\hat{\psi}$. We find an inverse relationship between $\hat{\psi}$ and the maximum level of sellers that accept cash $\bar{\theta}$, as expected. It means that if the private monopoly platform charges a high fee $\hat{\psi} \uparrow$, less sellers will accept digital payments ($1-\bar{\theta}$ decreases). Figure 2.1. shows that if the level of taxes is high (τ increases) sellers will prefer to use cash ($\bar{\theta}$ increases). From figure 2.2. we observe that if inflation increases more sellers will prefer digital payments instead of cash ($\bar{\theta}$ decreases).

[2] The assumptions we made are: $\rho = 0.8$, $\tau = 0.2$, $\beta = 0.9$, $\varepsilon = 0.2$, $\sigma = 0.5, \gamma = 1.02$ and $cost = 0.5$.

[3] $u(q) = \frac{q^{1-\rho}}{1-\rho}$ where ρ is the coefficient of relative risk aversion.

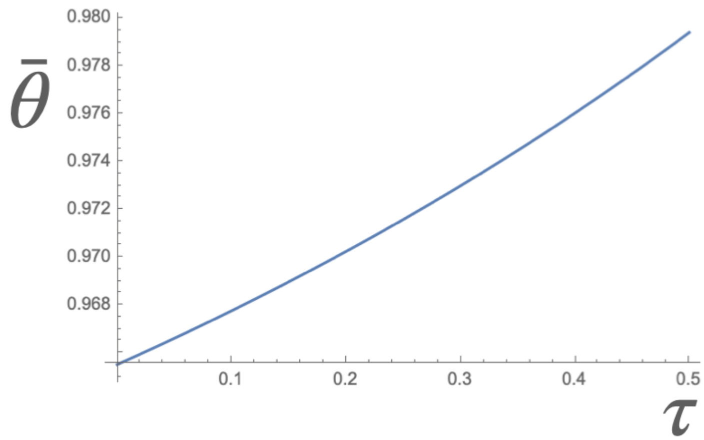

Fig. 2.1: Threshold of sellers and taxes

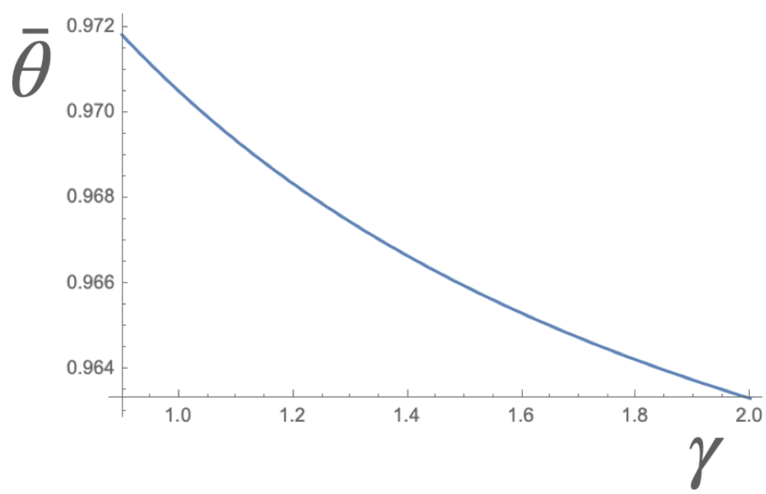

Fig. 2.2: Threshold of sellers and inflation

These results imply that if the digital platform is provided by a private monopoly, which charges a fee to the sellers, the equilibrium allocations continue being suboptimal like in the public provision case, given that the distortions in both markets are the same. However, in the case of a private provider the equilibrium quantities are distorted by the fee, via the amount of sellers who are indifferent between the two markets ($\bar{\theta}$), because sellers face another cost if they want to access to the digital platform. Now, more sellers prefer to accept cash ($\bar{\theta}$ increases) given that the private digital platform is more costly than the public which is free. It means that if the public provider does not charge any fee to the sellers for using the digital platform it is more efficient for the economy to have only the public platform for digital payments. However, given that the private provider faces some costs to provide the digital platform it has no way to compete with the free public platform at zero cost.

3 Third Setup: A Mixed Duopoly

In the mixed duopoly setup, the private provider competes with the public (government) provider. We assume that the government provider sets the fee ψ_g to the public cost $cost_g$ of providing the payment service. Since there is no reason to believe that the public cost ($cost_g$) equals the private cost ($cost$) , we have to consider several possible scenarios.

When the public technology is expensive, $\psi_g = cost_g > \psi$, the government cannot provide a viable payment service and the solution of the mixed duopoly is the same as in the previous section: the private monopoly.

When the public technology is in a medium range, $\psi_g \in [cost, \psi]$, the government can contest the market with the private payment service. Since $\psi_g \leq \psi$, the private provider has no choice but to match the public fee as otherwise the private provider would have no demand. If the private provider matches the fee, we assume that it obtains half the demand for digital payment service: $\left(1 - \bar{\theta}(\psi_g)\right)/2$, where $\bar{\theta}(\psi_g)$ is the threshold of merchant's type when the fee is ψ_g. However, the private provider can do better by setting its fee below but close to ψ_g, $\psi \uparrow \psi_g$: the private provider then captures the entire market with profit (almost) equal to $\left(1 - \bar{\theta}(\psi_g)\right)(\psi_g - cost)$. In this case, the public provider while usefully contesting the market only gets some residual demand, if any. The market is split between the two providers only in the knife-edge case where $cost_g = cost$.

Finally, if the government provider uses a better technology than the private provider ($cost_g < cost$), the government can become the sole provider of digital payments. It is however difficult to envisage that this could be the case.[4]

To conclude this section, a mixed duopoly in the digital payments market will be possible only if both platforms (the public and the private) charges the same fee to sellers. It could be possible because the public also incurs some costs to provide the platform, although the costs of both providers could be different. The most efficient result is one in which the fee charged to the sellers for accessing to one or another platform is the same and it has to be equal to the lower cost of provision.

[4] If this was possible, the government could always "give" its technology to the private provider, while competing with it at the same time.

Conclusions

This paper aims to provide some theoretical tools that help the payment systems analysis and that will serve for actual and future debates about the participation of central banks in these markets. Digital payments, in general, can have significant impacts on the equilibrium when there exist theft and tax evasion. Segmentation in payments systems limits the benefits of the existence of digital payments given the incentives that some merchants have regarding the use of cash, mainly because the existence of some distortions in this market, as tax evasion. Although, theft of cash could limit these incentives, it also imposes distortions on the economy of payments, that affect the optimal equilibrium.

Policy instruments including direct policies on theft and tax evasion are important to solve these market failures. Moreover, the existence of a public provider in the digital payments market could enforce the private providers to improve these services and increase the coverage of digital payments at a lower cost. However, ceteris paribus, the lack of interoperability between the public and private platforms may lead to the prevalence of market segmentation and the persistence of cash with the distortions that it imposes in the economy.. For future research, it could be interesting to analyze how interoperability between platforms could enhance social outcomes.

References

Arango-Arango, C., A. Ramírez-Pineda y M. Restrepo-Bernal, (2021). Person-to-business instant payments: Could they work in Colombia? Journal of Payments Strategy & Systems, Volume 15 / Number 4 / Winter 2021-22, pp. 376-397(22).

Arango-Arango, C, F., Arias, N. Rodríguez, N. Suárez y H. Zárate (2020). “Efectivo y pagos electrónicos”. Ensayos sobre Política Económica No. 93, Banco de la República.

Arango-Arango,C., Y. R. Betancourt-García, M. Restrepo-Bernal y G. Zuluaga-Giraldo, (2021). “Pagos electrónicos y uso del efectivo en los comercios colombianos 2020”. Borradores de Economía, No. 1180, Banco de la Republica, Colombia.

Aruoba, S. Borağan, 2021. "Institutions, tax evasion, and optimal policy," Journal of Monetary Economics, Elsevier, vol. 118(C), pages 212-229.

Bech, M., U. Faruqui, F. Ougaard, C. Picillo (2018). “Payments are a-changin’ but cash still rules”. BIS Quarterly Review, March 2018.

BIS. (2016). “Fast payments – Enhancing the speed and availability of retail payments”. November, Committee on Payments and Market Infrastructures.

CPMI-FPS. (2021). Developments in retail fast payments and implications for RTGS systems.

Duarte, A., J. Frost, L. Gambacorta, P. K. Wilkens and H. S. Shin, (2022). “Central banks, the monetary system and public payment infrastructures: lessons from Brazil's Pix". BIS bulletin, No. 52.

FIS. (2020). "Flavors of Fast".

Graeme, G. and J. Wright, (2007). "Competing Payment Schemes", Journal of Industrial Economics, Wiley Blackwell, vol. 55(1), pages 37-67, March.

Gu, C., Mattesini, F., Monnet, C., & Wright, R. (2013). "Banking: A New Monetarist Approach". The Review of Economic Studies, 80(2 (283)), 636–662.

Lagos, R., & Rocheteau, G. (2009). "Liquidity in Asset Markets with Search Frictions". Econometrica, 77(2), 403–426.

Lagos, Ricardo, Guillaume Rocheteau, and Randall Wright. 2017. "Liquidity: A New Monetarist Perspective." Journal of Economic Literature, 55 (2): 371-440.

Lagos, R., & Wright, R. (2005). "A Unified Framework for Monetary Theory and Policy Analysis". Journal of Political Economy, 113(3), 463–484.

Mohammed, A.L. (2020). "Informality, Frictional Markets and Optimal Policy". Job Market Paper.

Nosal, Ed. & Rocheteau, G. (2011). "Money, Payments, and Liquidity," MIT Press Books, The MIT Press, edition 1, volume 1, number 0262016281, December.

Reporte de Infraestructura Financiera. (2022). Banco de la República, Colombia.

Weinberg, J. A., (2002). "Imperfect competition and the pricing of interbank payment services", Economic Quarterly, Federal Reserve Bank of Richmond, issue Win, pages 51-66.

World Bank, (2017). "Fast payment systems: Preliminary Analysis of Global Developments". December, 2017.