C F P

5TO CALL FOR PAPERS

Contribuyendo al desarrollo
del sistema financiero

MEMORIAS CALL FOR PAPERS 2021

ASOBANCARIA

Presentación

Hernando José Gómez, Presidente Asobancaria

Desde sus inicios, la Asociación Bancaria y de Entidades Financieras de Colombia - Asobancaria ha propendido por ampliar el conocimiento e incentivar la investigación encaminada al desarrollo del sector financiero, conscientes de su papel protagónico en la intermediación y asignación de recursos productivos en la economía. La banca siempre ha encontrado en la academia un gran aliado en su desarrollo, pues es gracias a la investigación que se fomenta el pensamiento crítico para que con ideas innovadoras se logre una mayor productividad y eficiencia del sector, contribuyendo así a un crecimiento económico sostenible y mejor calidad de vida de la población.

Con el fin de continuar estrechando y fomentando los vínculos entre el sector financiero y la academia, por quinto año consecutivo brindamos un espacio para premiar a las mejores investigaciones del país en nuestros concursos *Call For Papers y Mejor Tesis de Maestría en Economía y Finanzas,* en el marco de la 32° edición del Simposio de Mercado de Capitales. De esta manera, buscamos visibilizar y divulgar trabajos de altísima calidad e interés para que, en su quehacer diario, las autoridades regulatorias, Gobierno Nacional, y funcionarios de las entidades puedan estar a la vanguardia de la producción académica y científica, orientada a la construcción y promoción de un sector bancario confiable, sostenible y competitivo.

En esta entrega, nos complace compartir con los lectores las memorias de nuestros concursos, donde encontrarán los mejores trabajos presentados en esta edición 2021, como una manera de premiar el esfuerzo de nuestros investigadores. Queremos, además, extender nuestros agradecimientos a todos los participantes que con sus hallazgos contribuyen de manera activa a la producción académica del país, así como al selecto grupo de jurados por apoyarnos con esta importante iniciativa. Es un honor haber contado con la participación de extraordinarios profesionales y académicos que con su labor constituyen una parte fundamental para seguir ampliando el conocimiento y entendimiento sobre el sector financiero.

Estamos convencidos del importante rol que tiene la academia en el desarrollo del sector y nuestro país, por lo que reiteramos nuestra cordial invitación a participar en las futuras ediciones de este proyecto, fortaleciendo la conexión del tejido académico con el empresarial. Gracias a sus aportes seguiremos avanzando en la consolidación de una banca cada vez más competitiva, sólida y eficiente que siga apoyando la materialización de proyectos de millones de familias y empresas de nuestro país.

Índice

Money Matters: Global Banks, Safe Assets and Monetary Autonomy*

April 5, 2021

Sergio Florez-Orrego

*A previous version of this document was presented as my master's degree thesis for the MA in Economics at Universidad de los Andes. I'm extremely grateful with David Perez Reyna and Andrés Murcia Pabón for their support and guidance. I would also like to thank Miguel Urrutia, Juan David Herreño, Nicolás de Roux, Oscar Becerra, Adriana Camacho, José Ignacio López, Daniel Wills, Hernando Zuleta, Marc Hofstetter, Humberto Martínez, Andrés Octavio Dávila, Laura Gómez, Santiago Torres, Daniel Zarama, Lucía Mendoza, Felipe Perilla, Matilde Angarita, Camilo Gómez, Daniel Espinosa, Marcela de Castro and all seminar participants at Universidad de los Andes and Universidad del Rosario, whose comments have been paramount in the development of this work. All errors are my own.

† Universidad de los Andes (e-mail: sa.florez11@uniandes.edu.co)

Abstract

This paper depicts an often neglected channel of transmission of monetary policy, namely international safety appetite, as an important source of production and risk-taking international monetary spillovers. The model features a local economy with exogenous financial frictions that lead firms to need both local and foreign financing to pay for their factors of production. Global and local risk-averse banks supply firms with risky loans while buying safe assets to governments to hedge themselves against equity shocks. Monetary policy shocks of a hegemon currency issuer affect returns obtained by banks for the risky loans they concede, altering these agents' risk pricing and balance sheet composition. Main results outline that global monetary policy tightening reduces the returns of risky global loans, inducing global banks to reduce risky loan creation, ultimately decreasing both production and consumption volatility internationally. Two more secondary results arise. First, local monetary authorities may counteract global monetary policy spillovers, but this will entail a trade off between boosting production and reducing consumption volatility. Second, both global and local expansive monetary policy increase the demand for global safe assets, relaxing the budget constraint of monopolistic global safe asset issuers. Understanding the international safety appetite mechanism of transmission appears to be of critical importance as it may impact the effectiveness of monetary policy in open economies as well as its optimal design.

JEL Codes: E42, E44, E52, E63, F42, F44

Keywords: global currencies, monetary policy spillovers, exorbitant privilege.

I. Introduction

Debates around non-traditional monetary policy transmission channels have been in the rise for the last three decades. Work as early as Gertler and Gilchrist (1993), Bernanke and Gertler (1995), Bernanke et al. (1999) as well as recent work from Gertler and Kiyotaki (2010) highlight that, by omitting banking and financial intermediaries in conventional monetary policy models, neoclassical transmission mechanisms have often overlooked the important relation between monetary policy, credit cycles and business cycles.

More recently, an important strand of this literature has highlighted 1) that monetary policy plays an important role in risk-taking behaviour of banks, with considerable consequences over aggregate macroeconomic risks (Borio and Zhu, 2012; Borio et al., 2019; Coimbra and Rey, 2017); 2) that there exists a Global Financial Cycle, closely intertwined with the widespread use of the U.S. dollar as the hegemon global currency, that raises questions on international monetary policy autonomy as stated by traditional Mundellian arguments (Miranda-Agrippino and Rey, 2015; Gerko and Rey, 2017;Rey, 2015; Jordà et al., 2018;Rey, 2016); as well as 3) the fact that U.S. debt plays a special role in safe assets' provision that confers an 'exorbitant privilege' to their treasury bonds in the form of lower debt costs (Gourinchas and Rey, 2007; Gourinchas et al., 2010, 2012). However, less work has been done on how these three subjects may interact. As these phenomena may change the nature in which monetary policy transmission channels operate, it is of paramount importance to have a theoretical structure that can integrate these concerns.

The purpose of this paper is to address the issues raised by this emerging macrofinance evidence, studying how do global funding needs by risk-averse global banks affect conventional knowledge on monetary policy transmission channels as well as its repercussions in open economies with free capital flows. For this end, I propose a simple theoretical model that places an often neglected channel of transmission of monetary policy –international safety appetite– as an important source of capital flows and international monetary policy spillovers.

The model highlights four results: 1) Production spillovers of monetary policy: Monetary policy tightening by global currency issuers –referred as global monetary policy shocks– are associated with a decrease in global bank's incentives to create global loans. In a setting where loans are an essential element of production, negative credit shocks triggered by contractive global monetary shocks reduce production in other 'local' economies.

2) Risk-taking spillovers of monetary policy: Changes in loan allocations imply a recomposition of banker's portfolio holdings between safe and risky assets. In this sense, monetary shocks also induce changes in aggregate risks. The model suggests that global monetary expansions increase portfolio risk and aggregate consumption volatility of agents internationally.

3) Responses of local monetary authorities to global monetary spillovers entail a trade-off between boosting production and reducing consumption volatility: Monetary authorities in 'local' countries may take actions that offset negative welfare effects of global monetary policy shocks. Following a global monetary policy contraction that reduces local production and aggregate welfare, local monetary authorities may respond by relaxing their monetary policy, increasing loan creation, production and aggregate welfare. However, as loans as risky, rise in loan creation also generates increased local consumption volatility.

4) Monetary policy conditions alter the budget constraint of the monopolist issuer of global safe assets: Monetary policy determines global bank's portfolio allocation, and by this way, global safe assets demand. Global monetary policy contractions increase the demand for global safe assets, decreasing the price that global investors charge the safe asset issuer for its debt and increasing this agent's consumption. Local monetary policy expansions have the opposite effect, decreasing the demand for global safe assets as well as consumption levels of the global safe asset issuer. This last point may constitute a source of local monetary policy spillovers over the economy of the global safe asset issuer.

What do we learn: The contributions of this paper are threefold. First, it provides a rationalization for a commonly ignored mechanism by which monetary policy may be determining credit cycles as well as business cycles, namely safety appetite. It appears surprising to learn that monetary policy has been historically entwined with the effects of liquidity provision over risk perceptions of financial intermediaries (Warburg, 1930; Gorton, 1984; Miron, 1986; Gorton and Huang, 2006)[1], while not sufficient work has focused on the possible consequences of this fact over business cycles dynamics.

Second, it proposes a simple model that allows to rethink conventional Mundellian wisdom and the relation between capital flows and monetary policy autonomy. Traditional open economy macro arguments have consistently highlighted that floating exchange rates may be enough to isolate an open economy from foreign shocks, including international monetary policy shocks (Mundell, 1963; Fleming, 1962; Galí and Monacelli, 2005). Incorporating monetary policy effects over safe asset markets with risk-averse investors may raise questions on this widespread tenet, as recent empirical macro finance literature suggests.[2] Furthermore, this work places local monetary policy reactions to global monetary spillovers at the center of the discussion on monetary autonomy.

Third, this model re-addresses the policy debate regarding the desirability of international monetary policy cooperation as well as current discussions developed at the heart of the International Monetary Fund on the adequacy of an integrated policy framework for open economies (Basu et al., 2020). Under the presumption that exchange rates flexibility is sufficient to isolate economies from foreign shocks, it arises as a natural conclusion that gains of monetary policy cooperation are negligible (Rogoff, 1985). In this model, the existence of both foreign and local monetary policy spillovers enrich this debate and add new nuances to the traditional deliberation. In the set- ting hereby proposed, risk aversion of global investors gives rise to arguments favouring the idea that there may be gains of international monetary policy cooperation, just as suggested by Keynes(1936),Caballero et al.(2015) and Ocampo(2017).

[1] As a matter of fact, the Aldrich-Vreeland Act, which finally lead to the creation of the Federal Reserve and set the road for modern central banking way-of-doing-things, was a response to the 1907 panic that allowed banking associations –a primitive private-form of Federal Reserve Board– to increase previously banned note issuance in emergency situations, just when financial markets liquidity is low and risk perceptions intensify (Laughlin, 1908).

[2] Gopinath et al.(2020) have addressed a different facet of the relation between foreign shocks isolation and floating exchange rates by studying how does dominant currency pricing in international trade affects the adjustment mechanism of the terms of trade highlighted by traditional open economy models.

Model: The model here proposed hinges on monetary policy effects over safety preferences of risk-averse global investors to integrate international monetary policy spillovers in a unified general equilibrium open economy model. It features one local or non- hegemon economy where firms have working capital requirements that force them to use both local and foreign loans to pay for their factors of production,[3] as well as one foreign or hegemon economy that, lacking firms, gets all its consumption from credit services provided to the local economy. In this milieu, credit creation is at the core of production.

Global as well as local commercial banks supply local firms with the credit needed to pay for factors. Credit creation is a risky endeavour, as it requires banks to hold assets–namely equity holdings used as funding for loans– that are subject to an idiosyncratic shock over their value. I refer to these assets as risky bank equity holdings. As banks are risk-averse, they hedge this risk by buying safe debt to governments, which I refer to as safe assets. In this sense, banks choose their optimal portfolio holdings, between risky bank equity holdings and safe government debt purchases, by weighting expected returns with perceived risks associated with these assets. Monetary policy determines the funding cost of banks which, as a result of the implied loan pricing structure within the model, affects the returns on risky equity holdings. Results suggest that, as monetary conditions tighten, funding costs rise and risky equity returns diminish, creating incentives for banks to reduce their risky bank equity holdings as well as loan creation while increasing their safe debt purchases. In turn, this entails a reduction both in production and in aggregate consumption volatility. This is the way in which the safety-appetite mechanism of transmission works.

While global monetary policy controls funding costs of global banks, local monetary policy controls funding costs of local banks.

While global monetary policy controls funding costs of global banks, local monetary policy controls funding costs of local banks. Global and local loan markets are segmented such that, due to imperfect substitution between these two sources of firm financing,[4] global loan shocks triggered by a global monetary tightening are able to reduce factor contracting by the local firm. This element is the key cornerstone underlying the existence of monetary spillovers of global currency issuers within the local economy.

In the previous setting, monetary policy conditions determine the demand of safe government debt. By modeling the debt issuance problem of the hegemon economy's treasury –which acts as a monopolist supplier of global safe assets–, this paper follows the work of Farhi and Maggiori (2017) and studies how does monetary policy affect what previous literature has called 'the exorbitant privilege'. The pompous term was originally devised by former French finance minister and president Valéry Giscard d'Estaing around the 1960s (Eichengreen, 2010), being used in recent times to identify the liquidity-safety premium enjoyed by the debt of countries that issue hegemon inter- national currencies –in particular U.S. Tresuries' safety premium–. Within the model, the hegemon government enjoys a monopoly premium that changes with monetary policy conditions. In particular, monetary policy tightening increase the demand of global safe assets, reducing the price that investors –in this case, global banks– charge the hegemon government for its debt. This reduction in the cost of the hegemon's debt generates a rise in its monopoly premium as well as its consumption.

[3] Recent work that integrates financial concerns to macro models through working capital require- ments may be found in Borio et al. (2019), Hill and Perez–Reyna (2017) and Jiang et al. (2020a).

[4] Sheng Shen (2019) places specialization in abilities to solve information asymmetries by global and local banks at the core of this segmentation.

Literature: Initially, this paper builds upon recent contributions on real monetary policy effects over funding costs of banks (Drechsler et al., 2017) and risk-premia[5]. In the framework here proposed, I suppose ex-ante that monetary policy has direct control over real funding costs of banks to see how changes in these costs alter the equilibrium risk-liquidity premia enjoyed by the hegemon government in the form of the exorbitant privilege. My approach is then very similar in spirit to recent work by Kekre and Lenel (2020a), where U.S. debt enjoys an inherent premium associated with its liquidity properties that changes with monetary policy shocks. The model framework hereby presented pressuposes real effects of monetary policy, so that this work is also related to the traditional discussions on monetary neutrality originally set by Lucas (1972), Barro (1976) and Sargent and Wallace (1975), which are at the heart of monetary economics.

Secondly, it approaches a new set of macro models that study monetary policy effects over risk-taking behavior of financial intermediaries –the risk-taking channel of monetary policy–. There are two differences between the model here proposed and previous work on this research strand. 1) While previous work has mainly targeted monetary policy effects on risk-taking by financial intermediaries in closed economy settings (Borio and Zhu, 2012; Coimbra and Rey, 2017; Borio et al., 2019), this work extends the argument to open economies with free capital flows. 2) Within less numerous work that has studied this subject in open economies (Bruno and Shin, 2015), there are no theoretical papers that shed light on the mechanisms that underlie banks' international risk-taking.

Thirdly, it builds on a wide range of works highlighting the paramount importance of global currencies in trade and international finance,[6] which appears to be the reason underpinning the emergence of monetary policy spillovers (Rey, 2015), as well as the exorbitant privilege.[7] While extensive

[5] Kekre and Lenel (2020b) study how premiums vary as nominal policy rates change heterogeneous marginal propensities of households to take risks.Lenel (2017) also establishes variations in the availability of safe assets that occur in episodes of non-conventional monetary policy as potential drivers risk-premia.
[6] See Gopinath (2015), Goldberg and Tille (2009), Gourinchas et al. (2019), Gourinchas (2019), Maggiori et al. (2019), Maggiori et al. (2020), and Gabaix and Maggiori (2015).
[7] See Gourinchas et al. (2010), Gourinchas and Rey (2007), Du et al. (2018), Krishnamurthy and Lustig (2019), Jiang et al. (2020b) and Farhi and Maggiori (2017).

empirical research has been done on each individual issue, less theoretical work has been developed on the key elements that knit them together.

Fourthly, the cornerstone mechanism here proposed draws on the latest works concerning the macroeconomic implications of safe asset shortages. Low supply of safe assets relative to their demand gives rise to risk premiums that alter the connection between monetary policy and economic activity (Caballero, 2006; Caballero et al., 2016; Gourinchas and Jeanne, 2012). Caballero and Farhi (2017) refer to this phenomenon as the Safety Trap. The model presented in this paper may be thought of as modeling an economy which is already at the Safety Trap. It is in this sense that monetary policy may affect safe asset holdings of investors as well as equilibrium safe return rates.

The recent papers of Gourinchas et al. (2020), Jiang et al. (2020a) and Kekre and Lenel (2020a) are the most closely related to this work, proposing a model where global investors are willing to pay currency and bond risk-premiums in order to explain how monetary policy may transmit internationally via safe asset markets. However, none of them tackle the possible interactions between global and local monetary policy as this paper does. As previously explained, the results of this paper suggest that, when responding to hegemon monetary policy spillovers, local monetary authorities face a trade-off between boosting production and reducing consumption volatility. To the best of my knowledge, this work is the first paper that includes this element as a theoretical possibility that underlies the international financial architecture.

Outline: The rest of the paper will be organized as follows. Section 2 describes the model. Section 3 presents the main results. Section 4 concludes.

2. Model

The paper presents a general equilibrium two period model composed by a unique local firm with working capital constraints –so that the firm needs loans to pay for its factors of production–, global and local risk-averse bankers that supply loans to these firms, households that supply labor to firms and savings to banks in the form of deposits, and governments that issue safe debt. There is an idiosyncratic shock on equity value of banks, in such way that these agents will hedge themselves by buying safe assets to governments. By altering the returns on bank loans, monetary policy alters risk pricing of banks as well as their asset composition. Local economy variables and agents will be identified by h, while foreign ones will be identified by f.

2.1 Households

The representative local household is characterized by a representative worker with CRRA preferences that lives both periods. In the first period, this agent finances its purchases with labor income obtained by supplying labor hours to

firms. One may think the first period as the moment in which this individual is young and can work. In the second period, he does not work but consumes the returns on its first period savings. Taking the previous analogy, one may think the second period as the moment when the representative worker is too old to work. There is only one financial instrument available for this agent to save for old-age, namely local banks deposits. The paper only models the behaviour of local households, while supposing that there exists a foreign household that elastically supplies deposits to global banks.

Denote the local household by superscript *wh*. Local household's problem may be written as follows

$$\max_{\{d_h, N\}} \quad \mathcal{W}^{wh} = \left[\frac{C_1^{wh\,1-\sigma}}{1-\sigma} + \rho \frac{C_2^{wh\,1-\sigma}}{1-\sigma} - \phi \frac{N^{1+\eta}}{1+\eta} \right] \tag{2.1}$$

$$\text{s.t.} \quad C_1^{wh} = WN - d_h$$

$$C_2^{wh} = R_h^d d_h$$

where W^{wh} denotes the welfare of worker w in local economy h, C^{wh} identifies this agent's consumption in period t, d_h local deposits, N local labor hours, W is the real wage per unit of labor and R_h^d is the real return on deposits. σ and η are consumption and labor substitution parameters that determine the curvature of the utility of consumption and the disutility of labor. ρ characterizes the discount factor while φ is a labor disutility parameter.

2.2 Firms

There is one unique standard representative local firm with Cobb-Douglas technology, that uses both capital and labor as inputs to produce. This firm pays for its factors of production upfront, in the first period of the economy. However, it receives its final produce only at the beginning of the second period. In this setting, working capital requirements arise such that the firm needs loans to pay for its factors of production. As will be clear when defining the concept of equilibrium used within the model, first period consumption goods come from banks' wealth endowments and from wholesale funding supply given by central banks.

The model assumes that each type of bank –local or global– only finances one specific factor of production. Concretely, local banks only finance the local factor –labor N– while global banks only finance the foreign factor –capital K–. This is assumption is equivalent to segmentation between global and local financial markets. In this model, I construct over an extreme case segmentation where neither local banks lend for the foreign factor, nor global banks lend for the local factor. In the real world, one may expect an intermediate scenario where there may be some substitution between local and global loans when financing either locally-produced factors or foreign-produced ones.

Denote the firm by superscript F. Firm's problem is

$$\begin{aligned} \max_{\{N,K\}} \quad & \Pi_2^F = F(K,N) - R_h^L W N - R_f^L K \\ \text{s.t.} \quad & F(K,N) = Z K^{\alpha} N^{1-\alpha} \end{aligned} \tag{2.2}$$

where Π_2^F indicates firm's profits received in the second period, K is capital, R_h^L denotes real return on local loans and R_f^L the real return on foreign loans. Z and α denote total factor productivity and capital share parameters. Local loans L_h demanded by the firm should be equal to total value of labor contracting WN, while total foreign loans L_f demanded by the firm should be equal to total value of capital contracting K.

2.3 Bankers

There exist two types of risk-averse commercial bankers, one local and one global, each of whom owns a bank that allows them to generate income. Each bank intermediates deposits. Global banks intermediate elastically supplied global deposits, while local banks intermediate deposits supplied by local households. Banks face regular leverage constraints[8] that force them to raise internal funds in order to get deposit funding.

Banks may also access wholesale money markets, that provide funding at a spread ψ from deposits. Global banks make use of global currency wholesale funding markets, while local banks use local currency wholesale funding markets. As in the case of deposits, wholesale funding is subject to capital requirements that limit the amount that these agents may borrow.

Finally, bankers face a shock that reduces the value of their equity holdings between period 1 and 2. As these agents are risk-averse, they hedge themselves against the shock by buying safe low-return assets to governments. The model assumes that global banks only consider foreign or hegemon government debt as safe, while local banks only consider non-hegemon or local government debt safe, and that the reason behind this preference lies in the currency of issuance of each debt. This is consistent with empirical evidence set by Maggiori et al. (2020), according to which local investors' security holdings are biased towards safe debt denominated in their own currency, while global investors are towards safe debt denominated in global currency.[9]

[8] See Hill and Perez–Reyna (2016).

[9] Numerous work including Krishnamurthy and Vissing Jorgensen (2012), Greenwood et al. (2015), as well as Nagel (2016) place high liquidity services of the hegemon government debt –the U.S.– in international financial markets at the center of this phenomenon. Borio et al. (2016) have also shown how post-2008 financial regulations that aim to improve global banks' foreign exchange exposures have put pressures on the demand of safe assets denominated in dollars, amplifying this currency denomination bias.

Each banker (local or global) is characterized by an agent with mean-variance preferences that obtains income from its asset purchases. When the economy begins, in period 1, these agents are endowed with wealth resources that they distribute between high-risk/high-yielding assets –namely equity capital destined to fund their banks – and low-risk/low-yielding government debt. Let subscript $j \in \{h, f\}$ denote domestic *(h)* and foreign *(f)* prices and quantities, and superscript $bj \in \{bh, bf\}$ denote domestic *(bh)* and foreign *(bf)* bankers. Bankers' problem may be written as

$$\max_{\{s_j, b_j\}} \quad \mathcal{W}^{bj} - \mathrm{E}\left[C_2^{bj}\right] - \frac{\chi}{2}\mathrm{Var}\left[C_2^{bj}\right] \tag{2.3}$$

$$\text{s.t.} \quad \omega = s_j + b_j$$

$$\Lambda_{j,2} = \pi_j + R_j b_j - \Omega\delta s_j$$

$$C_2^{bj} = \Lambda_{j,2}$$

where W^{bj} denotes the welfare of the banker in economy *j (b in j)* and C^{bj}_2 indicates this agent's consumption which takes place in period 2. First restriction indicates the budget constraint of bankers in period 1, where initial wealth endowments ω should be distributed between equity holdings s_j –directed to fund their banks– and government debt purchases b_j –where global (local) bankers only buy global (local) government debt–. $\Lambda_{j,2}$ is the real value of bankers' portfolio in the second period. π_j are the profits that the bankers obtain from their bank, R_j identifies the return rate that bankers charge the government of each economy for their debt, Ω is a random dichotomic variable that assumes the value of 1 if the equity shock realizes, $\delta \in (0, 1)$ is the loss of equity value in the case that the shock occurs, while χ is a risk-aversion parameter.

Bank profits π_j are determined by the following profit maximization problem

$$\max_{\{L_j, d_j, l_j\}} \pi_j = R_j^L L_j - R_j^d d_j - R_j^d(1 - \psi_j) l_j \tag{2.4}$$

$$\text{s.t.} \quad L_j = d_j + l_j + s_j$$

$$\frac{d_j}{s_j} \leq \lambda_1$$

$$\frac{l_j}{s_j} \leq \lambda_2.$$

Bank profits π_j are equal to the net return on given loans L_j of price R^L_j, after subtracting funding costs of deposits d_j of price R^d_j, as well as costs of wholesale funding l_j of cost $R^d_j(1-\psi_j)$. The model supposes that that central banks act as deep-pocketed institutions that elastically supply all wholesale funding demanded by commercial banks at prevalent interest rates.[10] Variable $\psi_j \in (0, 1)$ denotes the spread of deposits costs over wholesale funding costs, and will be assumed to be controlled directly by central banks. In particular, expansive global (local) monetary policy –or cheaper global (local) currency wholesale funding– will be associated with increases in the spread $\psi_f(\psi_h)$.

First restriction in problem (2.4) indicates banks' balance sheet constraint, where loans should be financed either by deposits, by raising wholesale funds, or by raising internal funds s_j –equity holdings of banks–. Second restriction denotes the exogenous leverage constraint that limits the amount of deposits that banks may borrow from households. According to this constraint, the ratio between deposits and bank equity is capped at λ_1. Similarly, banks must fulfill capital requirements in order to access to wholesale funds. Capital requirements are identified by the third restriction, so that the ratio between wholesale funds and bank equity is capped at λ_2.

2.4 Governments

The paper supposes the existence of one foreign and one local government that issue public debt to banks in order to finance public expenditure. Their debt is safe. Governments have a technology of public investments with fixed returns that determines their equilibrium consumption levels. Constructing on Farhi and Maggiori (2017), the exorbitant privilege is introduced as a monopoly rent of which the foreign or hegemon government enjoys as a monopolist issuer of global safe assets. This allows for a spread to appear between the marginal benefit and the cost of global safe debt issuance, that will in turn be modified as monetary policy conditions change.

Each government issues debt in period 1, allocating the entirety of raised funds to a public investment technology with fixed returns, that lasts 1 period to produce final output. Both governments receive the proceeds of their investments and consume in period 2. As before, let subscript $j \in \{h, f\}$ denote domestic *(h)* and foreign *(f)* prices and quantities, and superscript $gj \in \{gh, gf\}$ denote domestic *(gh)* and foreign *(gf)* governments. Governments' problem may be characterized as

[10] See Coimbra and Rey (2017).

$$\max_{\{b_j\}} \quad \mathcal{W}^{gj} - C_2^{gj} \qquad (2.5)$$
$$= \mathcal{F}(b_j) - R_j(b_j) \cdot b_j \quad : \quad \mathcal{F}(b_j) = \gamma b_j$$

where W^{gj} characterizes the welfare of the government of economy *j (g in j)*, C_2^{gj} denotes this agent's consumption –which takes place in period 2–, F identifies public investment technology with constant return γ, b_j accounts for the level of public debt issuances of the government in j and $R_j(b_j)$ identifies the debt cost function for this agent. The local government is supposed to issue its debt on competitive markets, so that it will take debt costs R_b as given. By contrast, the foreign government is assumed to be a monopolist issuer of global safe debt such that it internalizes the effects of its debt issuance over its costs. In this sense, debt cost function $R_f(b_f)$ is equivalent to the inverse demand function of foreign public debt. Monopoly power will give rise to a monopoly-risk premium on foreign government's debt.

2.5 Equilibrium

Definition 1 presents the general concept of equilibrium used in this work.

Definition 1. Equilibrium: *Given exogenous monetary conditions set by central banks* $\psi = \{\psi_f, \psi_h\}$, *an equilibrium in this economy will be identified by allocations* $x^{wh} \equiv \{d_h^{wh}, N^{wh}\}$, $x^F = \{K^F, N^F\}$, $x^{bj}(\psi) = \{s_j^{bj}(\psi), b^{bj}(\psi), L_j^{bj}(\psi), d_j^{bj}(\psi), l_j^{bj}(\psi)\}$, $x^{gj} = \{b_j^{gj}\}$, *and prices* $p \equiv \{W, R_h^d, R_f^d, R_h^L, R_f^L, R_h, R_f\}$ *such that*

1. *given* p, x^{wh} *is a solution to* (2.1);
2. *given* p, x^F *is a solution to* (2.2);
3. *given* p, $x^{bj}(\psi)$ *is a solution to* (2.3) *and* (2.4) *for each* $j \in \{h, f\}$;
4. *given* p, x^{gj} *is a solution to* (2.5);
5. *All markets clear:*

(a) labor: $N^{wh} = N^F$;

(b) capital or global loans: $L_f^{bf}(\psi) = K^F$;

(c) local loans: $L_h^{bh}(\psi) = WN$;

(d) global deposits: $d_f^{bf}(\psi)$ *(elastically supplied to global bankers);*

(e) local deposits: $d_h^{wh} = d_h^{bh}(\psi)$;

(f) safe assets: $b_j^{gj} = b_j^{bj}(\psi)$ *for each* $j \in \{h, f\}$;

(g) wholesale funding: $l_j^{bj}(\psi)$ *for each* $j \in \{h, f\}$ *(elastically supplied to global and local banks);*

(h) goods:

$$t = 1: \qquad \omega + l_h = \left[C_1^{wh} + b_h + K \right] - \left[\omega + l_f + d_f - b_f \right],$$

$$t = 2: \qquad F(K, N) + \mathcal{F}(b_h) - \Omega\delta s_h = \left[C_2^{wh} + C_2^{bh} + C_2^{gh} + R_h^d(1 - \psi_h) l_h \right] + \left[R_f^d d_f + C_2^{bf} + C_2^{gf} + R_f^d(1 - \psi_f) l_f - (\mathcal{F}(b_f) - \Omega\delta s_f) \right].$$

In definition 1, left hand side of goods markets' clearing equations for both periods depict local supply of goods. In this first period, this amounts to local banker's wealth endowmens ω plus the wholesale funds given by the deep-pocketed central bank. In the second period, this amounts to the sum of private and public local production, less the loss in product in the case the shock Ω occurs. First term of the right hand side of both equations shows domestic absorption, while the second term shows the trade balance. In the first period, there are only imports as local firms need capital that can only be payed with foreign inflows. In the second period, outflows –or exports– are used to compensate foreign agents for their loan services. Appendix A demonstrates that general equilibrium in goods markets holds for both periods 1 and 2.

3. Results

This section presents the main results of the paper. First, I characterize the equilibrium to explain the economics behind the dynamics portrayed in the model: optimal asset allocation of banks, optimal debt issuance by governments and aggregate consumption volatility. Thereupon, I define monetary policy shocks and solve the model numerically to draw conclusions on general equilibrium.

3.1 Characterizing the Equilibrium

Optimal asset allocation: I proceed to characterize banks' optimal asset allocation. I will uniquely focus on the interesting case where $R^L_j > R^d_j$, in which case both leverage and capital requirements –constraints 2 and 3 in equation (2.4)– hold with equality as stated in lemma1.

Lemma 1. *For banks to demand positive deposits and wholesale funds, it must happen that* $R^L_j \geq R^d_j$. *In the interesting case where* $R^L_j > R^d_j$, *both leverage and capital requirement constraints are binding.*

$$\frac{d_j}{s_j} = \lambda_1 \qquad and \qquad \frac{l_j}{s_j} = \lambda_2.$$

Proof: See appendix B.

Using results from lemma 1, in conjunction with bank's balance sheet constraint and banks profits in equation (2.4), one may obtain expressions for L_j, d_j, l_j and π_j that only depend on bank capital s_j and prices, such that *bank* allocations –this is, demanded loans and wholesale funding, and supplied loans– are completely determined by bankers optimal risky equity allocation s_j.

$$d_j(s_j) = \lambda_1 s_j \tag{3.1}$$

$$l_j(s_j) = \lambda_2 s_j \tag{3.2}$$

$$L_j(s_j) - (1 + \lambda_1 + \lambda_2)s_j \tag{3.3}$$

$$\pi_j(s_j) = \left(R^L_j - R^d_j\right)\lambda_1 s_j + \left[R^L_j - R^d_j(1-\psi_j)\right]\lambda_2 s_j + R^L_j s_j. \tag{3.4}$$

Note that loans L_j are increasing in the holdings of S_j, such that global/local loans will rise as risky global/local equity holdings by banks increase. In this sense, leverage constraints and capital requirements bind loan creation to equity allocations. This last assertion is formally stated in proposition 1.

Proposition 1. *Leverage and capital requirement constraints bind loan creation to internal-funding equity allocation of banks. Without constraints on funding sources, total loan levels would not be determined by internal-funding equity allocations, but on supply of external-funding of deposits and wholesale funds.*

Proof : See appendix C.

Proposition (1) highlights the importance that capital regulations play in determining asset allocation dynamics of investors, what will ultimately enhance the safety- appetite transmission mechanism presented in this work. In fact, this element has been a persistent feature in recent literature highlighting how changes in banking regulations after the Great Financial Crisis may have altered the nature of monetary policy transmission channels.[11]

By replacing equation (3.4) in the second constraint of banker's problem –equation (2.3)–, one may easily solve the maximization problem, obtaining the following optimality condition underlying bank's optimal asset allocation[12]

$$\pi_j{}'(s_j) = R_j + \delta \mathrm{E}(\Omega) + s_j \chi \delta^2 \mathrm{Var}(\Omega). \qquad (3.5)$$

Left hand side of equation (3.5) corresponds to the marginal benefit that bankers earn by allocating an extra unit of equity capital into their banks, while right hand side identifies the marginal costs of these investments. The first term of the marginal cost corresponds to renounced returns of safe debt, the second one to the marginal expected loss of equity capital associated with an increase in risky equity holdings, while the third one corresponds to the increase in portfolio volatility caused by the increase in holdings of these risky assets.

Equation (3.5) depicts the trade-offs that bankers face when deciding how much equity to allocate as internal funding to their banks, while this ultimately determines equilibrium supply of global loans as stated by equation (3.3). Main intuition may be stated as follows: contractive monetary policy is associated with reductions in the spread between deposit funding and wholesale funding ψ_j, so that the marginal benefit of equity investments $\pi_j'(s_j)$ decreases. This is evident by recalling that equation (3.4) decreases as ψ_j declines. Ceteris paribus, as the marginal benefit of equity investments falls, investors will prefer to allocate their wealth into safe assets of return R_f, which have no expected losses and are not associated with portfolio volatility. In the previous scenario, official liquidity costs set by global central banks ψ_f are a key determinant of risk pricing of global investors.[13]

[11] See Borio and Zhu (2012) for a discussion on their role on the *risk-taking* channel. Fornaro and Romei (2019) discuss their effects on the eventual rise of liquidity-safety traps with recessionary effects, as suggested by Keynes (1936) original paradox of thrift.

[12] See appendix D for complete derivation of bankers' first order condition.

[13] This characteristic has been recently highlighted by Cohen et al. (2017), who suggest that the ability of market participants to raise global currency cash is a key determinant of their perceived risks. In this sense, by setting the cost of global liquidity in wholesale funding markets, hegemon currency issuers are in a convenient position to 'lean against the wind' of global financial markets safety perceptions, and in this way, influence risk-appetite and balance sheets of lenders internationally (Avdjiev et al., 2016).

Optimal safe asset issuance is determined by the market power of governments when issuing safe debt.

Optimal safe asset issuance: Optimal safe asset issuance is determined by the market power of governments when issuing safe debt. The local government issues its debt in competitive debt markets, such that it takes debt costs as given. Hence, optimality condition for local debt issuance is

$$R_b = \gamma.$$

This comes from the fact that local government problem is linear in debt issuance b_b, such that the marginal cost of debt should equal marginal productivity of public investments. Recall that this condition results in fixed local government consumption (equal to zero).

$$\begin{aligned} C_2^{wh} &= \gamma\, b_h - R_h\, b_h \qquad (3.6) \\ &= 0. \end{aligned}$$

By contrast, the foreign government will be a monopolist issuer of global safe debt such that it internalizes the effect of its debt issuance over the demand of safe assets and over debt costs. This gives rise to a monopoly-risk premium on foreign government debt. First order condition for this agent is then

$$\gamma = R_f'(b_f) \cdot b_f + R_f(b_f), \qquad (3.7)$$

where left hand side of equation (3.7) is the marginal return that the foreign government obtains from its investments, while the right hand side denotes the marginal costs that this agent faces when deciding its optimal issuance. Note that there exists a spread equal to $R_f'(b_f) \cdot b_f$ between returns obtained by the hegemon government for its investments γ and the returns payed by the foreign government to global banks $R_f(b_f)$. This term is positive as it comes from the inverse demand for global safe assets. Recall that increases in the supply of an asset –in this case, government bonds– should be accompanied, ceteris paribus, by an increase in the return of that asset that foster investors to accept the excess supply.[14] The positive spread $R_f'(b_f) \cdot b_f$ characterizes then the monopoly-safety rent of the global safe asset issuer, namely the exorbitant privilege.

Aggregate Volatility: Let C_2^h be total consumption of the local economy's agents in period 2, C_2^h may be expressed as

[14] It is possible to see this formally by finding the demand of global safe assets (from global banker's solution), isolating Rf and derivating with respect to asset issuance b_f.

$$C_2^h = C_2^{wh} + C_2^{bh} + C_2^{gh}$$
$$= R_h^d\, d_h + \Lambda_{h,2}, \tag{3.8}$$

where local government consumption C^{gb}_2 is equal to zero –see equation (3.6)–. Departing from equation (3.8), and recalling that there is no uncertainty on loans, deposits nor on local safe asset return rates $\{R^L_g, R^d_b, R_b\}$ one may define aggregate local consumption volatility as

$$\mathrm{Var}(C_2^h) = \mathrm{Var}(\Lambda_{h,2})$$
$$= s_h^2\, \delta^2\, \theta\, (1-\theta). \tag{3.9}$$

Local consumption volatility is then equal to local banker's portfolio volatility. This is in fact a consequence of the modeling strategy that assumes that all risk in the model is bore by banks. As basic economic intuition may suggest, aggregate volatility is positively correlated with the levels of risky equity holdings of local banks s_b.

3.2 Monetary Policy Shocks

In what follows, I proceed to present the numerical results of the paper, by analyzing how global and local monetary policy shocks affect the equilibrium of the model. Monetary policy determines the spread between deposit and wholesale funding, as stated in definition 2.

Definition 2. Monetary Policy: *Monetary policy of hegemon currency issuers will set the spread ψ_f between global currency wholesale funding and global deposits funding. Global monetary expansions will be associated with increases in the spread ψ_f while contractions will cause reductions in this spread. Hegemon currency issuers will supply all wholesale funding demanded by global banks at prevalent prices. Monetary policy of local currency issuers will set the spread ψ_b between local currency wholesale funding and local deposits funding. Local monetary expansions will be associated with increases in the spread ψ_b while contractions will cause reductions in this spread. Local currency issuers will supply all wholesale funding demanded by local banks at prevalent prices.*

In definition 1, monetary policy conditions $\psi = \{\psi_f, \psi_b\}$ were defined as exogenous variables within the equilibrium. I start by analyzing the effect of a reduction in ψ_f, while leaving local monetary policy spread ψ_b fixed at an arbitrary level. I refer to this scenario as a global monetary contraction. I subsequently depart from the global monetary contraction scenario and characterize a stylized local monetary policy rule to study the effects of local monetary policy responses to global monetary contractions. I refer to this scenario as the optimal local monetary response. The global monetary contraction scenario may be thought of as a first stage within the timing of the model, and the local monetary response as a second stage. For the following exercises, I suppose Ω is Bernoulli-distributed with parameter $\theta = 0.5$. Table 1 shows predetermined parameters used in the numerical analysis.[15]

3.2 Monetary Policy Shocks

Figure 1 depicts the markets of the model before and after the contractive global monetary policy shock. Panel (a) identifies the direct effect of this shock over conceded loans by global banks. As global wholesale funding becomes more expensive, global loan supply drops causing a raise in global loan price R_f^L.

Panel (b) represents the indirect effects of global monetary policy over the local factor market –labor–. Labor demand diminishes due to complementarity between factors of production. Less capital contracted leads to excess capacity so that labor contracting is less appealing to firms. In general equilibrium, there will also be a reduction in labor supply. In the aggregate, both effects generate a fall in effective labor contracted as well as in wages.

[15] In global monetary contraction exercise, $\psi_b = 0.5$, while ψ_f decreases from 0.5 to 0.3. For local monetary response exercise, ψ_f stays at 0.3 while ψ_b goes from 0.5 to it's optimal value, according to a stylized monetary policy rule presented in section 3.2.2. Interest rates on global deposits R_f^d should also be defined as a parameter. This is not problematic as the paper assumes that global deposits are supplied elastically at any deposit rate.

Table 1.

Predetermined parameters

Parameter	Value	Description
ρ	0.995	Household discount factor
σ	0.7	Intertemporal consumption substitution parameter
η	1.1	Labor substitution parameter
ϕ	0.8	Labor disutility parameter
α	0.53	Capital participation
Z	4	Total factor productivity
χ	20	Banker's risk-aversion parameter
ω	1.2	Banker's initial wealth
δ	0.5	Equity loss in case of shock realization
θ	0.5	Bernoulli parameter of shock Ω
λ_1	6	Leverage requirements
λ_2	0.4	Capital requirements
γ	1.06	Public investment productivity
R_f^d	$1/\rho = 1.005$	Global deposits return rate

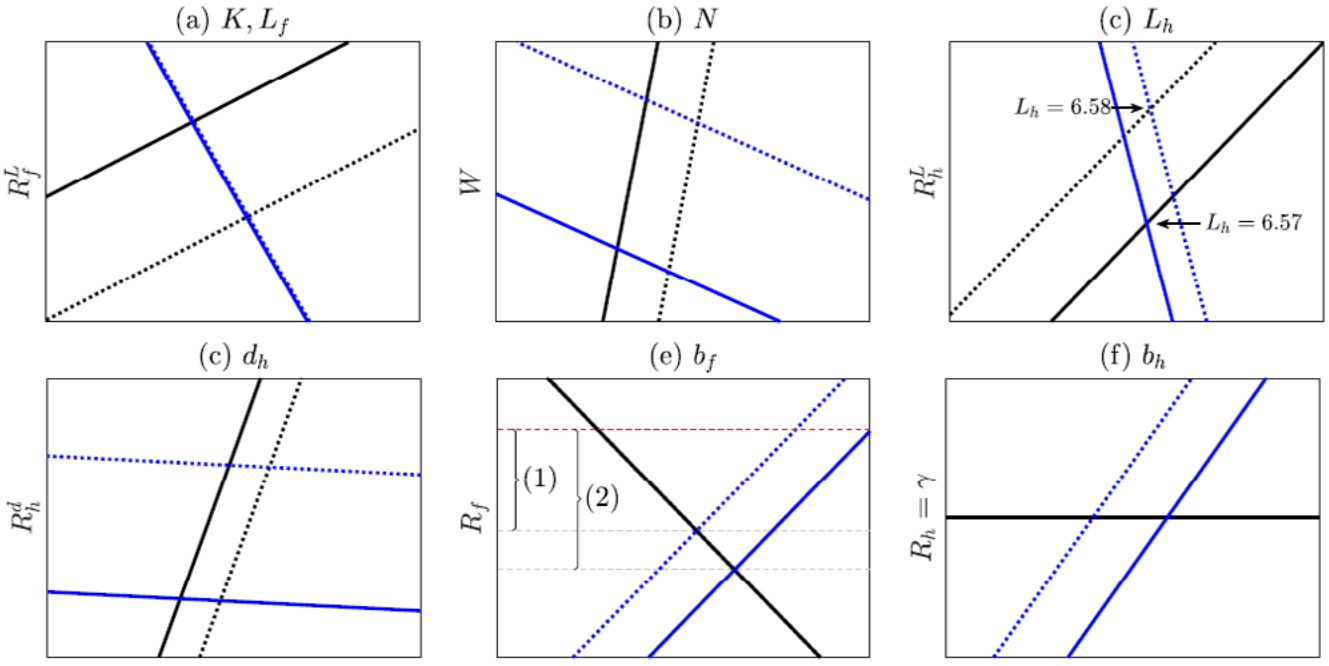
Panel (c) shows the local loans market before and after the shock. Local loans market acts as the payments counterpart to the labor market just explained above. As wages and labor levels fall, labor payments WN decrease and less local loans are demanded by firms. This is consistent with a left movement of the local loans demand that pushes local loans price R_H^L downwards. It also appears that, in equilibrium, local loans supply increases. This is in fact a consequence of the curvature of deposits demand function around selected equilibria. To see this, note that drops in local loans demand should also imply less funding needs by local banks, what entails less local deposits demand. As panel (d) shows, this results in a reduction in deposit costs R_b^d. To sum up, both loan returns and deposit funding costs decrease following the drop in labor demand. The fact that local loan supply in panel (c) increases after the global monetary policy shock suggests that, in equilibrium, marginal deposit costs are decreasing faster than marginal loan returns, what increases local banks incentives to create local loans. However, this minor particularity would not be fundamental to the main results of the paper. Thus, further discussion on the issue is sidelined.

Panel (e) illustrates global monetary policy effects over the global safe asset market/hegemon government debt market. The red dashed line identifies the value of the fixed marginal return on public investments γ, while brackets (1) and (2) identify the premium of public investment returns over equilibrium debt costs for the foreign government –or the exorbitant privilege–, which widens as a result of the global monetary tightening. As will be seen in figure 2, this also leads to a rise in foreign government consumption levels. Global contractive monetary policy caused a reduction in risky

global loan returns –recall equation (3.4)–, inducing global bankers to modify their portfolio holdings by increasing their purchases of safe assets and displacing the demand for this asset rightwards. After the global monetary shock, bankers are willing to accept higher holdings of safe hegemon debt, even if it comes at the price of lower returns on its debt.

Figure 1.

Global monetary policy shock: Effects of contraction in funding spread ψ_f over markets

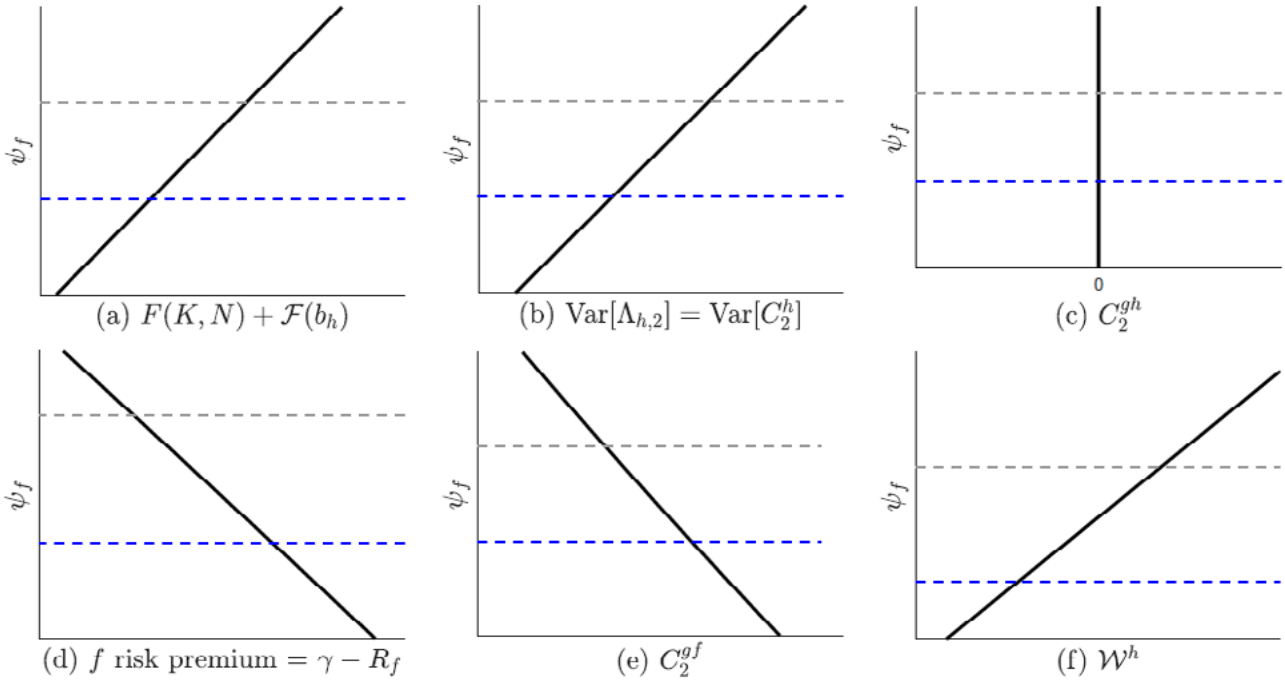


Note: Figure shows market plots. Black curves identify supplies. Blue curves identify demands. Dotted lines identify equilibrium before hegemon monetary shock while solid lines identify equilibrium after the shock. The red dashed line in panel (e) identifies public investment return γ.

Panel (f) shows the dynamic of the local safe asset market. Global monetary contractions causes an increase in this asset's demand. This the counterpart of the reduction in local loans due to the cutback of labor payments depicted in panels (b) and (c). Panel (f) also evidences the fact that local safe debt is elastically supplied at price $R_h = \gamma$. This the main difference between hegemon and non-hegemon processes of debt issuance, and is the main reason underlying the fact that foreign government consumption is positive (and dependent on monetary policy conditions) while local government consumption is always zero, independently of its debt issuance level.

Figure 2.

Global monetary policy shock: Effects of contraction in funding spread ψ_f over interesting variables



Note: Black solid lines identify general equilibrium relations between referenced variables. Grey dashed lines identify global currency wholesale funding spread ψ_f before the shock while blue dashed lines identify the spread after the shock.

Figure 2 illustrates equilibrium relations between global monetary policy levels and interest variables of the model, where the grey dashed lines indicate the funding spread ψ_f before the monetary shock while blue dashed lines indicate the value of this parameter after the shock. Panel (a) states that global monetary contractions cause a reduction in total local production, as less capital and less labor is contracted –as shown in panels (a) and (b) of figure1–. These are the *production* spillovers of monetary policy of global currency issuers.

Panel (b) shows that these contractions are also associated with reductions in local banker's portfolio volatility. This comes from the fact that there is an increase in holdings of safe local debt –as shown in panel (f) of figure 1– with a corresponding reduction in risky local equity holdings by banks. From equation (3.9), local consumption volatility rises hand in hand with local banker's portfolio volatility. These are the *risk-taking* spillovers.

Panel (c) identifies local government's consumption as global monetary conditions change. As shown in equation (3.6), this agent's consumption is always constant and equal to zero, as the local government issues its debt in competitive local safe asset markets so that its debt costs always equal the returns obtained by reinvesting the debt in the public investment technology.

Panel (d) and (e) are extensions of what was shown in panel (e) of figure 1, portraying the way in which the exorbitant privilege and global monetary conditions interplay. As global monetary conditions tighten, the safety premium on hegemon's government debt as well as its consumption increase.

Finally, panel (f) shows the effect of global monetary policy contractions over aggregate welfare in the local economy W^b. Intuitively, as global monetary conditions tighten and local production falls, so does available consumption goods for agents, thus reducing aggregate local welfare.

3.2.2 Local Monetary Policy Response

I proceed to define a stylized monetary policy rule under which local monetary authorities respond to global monetary policy contractions –that reduce local production and welfare– with local monetary expansions –that once again increase both local production and welfare–. This rule is introduced in definition 3. Then, I present numerical results when this monetary policy rule is included.

Definition 3. Local Monetary Policy Rule: *Let $\{\psi_f, \psi_b\}$ be the vector of exogenous prevalent monetary policy conditions, $W^{wb}(\psi_f, \psi_b)$ local household's welfare, $W^{bb}(\psi_f, \psi_b)$ local banker's welfare, $W^{gb}(\psi_f, \psi_b)$ local government's welfare and $W^b(\psi_f, \psi_b)$ the aggregate welfare of the local economy associated with monetary conditions $\{\psi_f, \psi_b\}$.*

Given an arbitrary state 0 with monetary policy conditions $\{\psi_{f0}, \psi_{b0}\}$, and a subsequent exogenous global monetary policy shock $\psi_{f1} \neq \psi_{f0}$, local monetary authorities will set local monetary conditions $\tilde{\psi}_{b1}$ according to the following problem

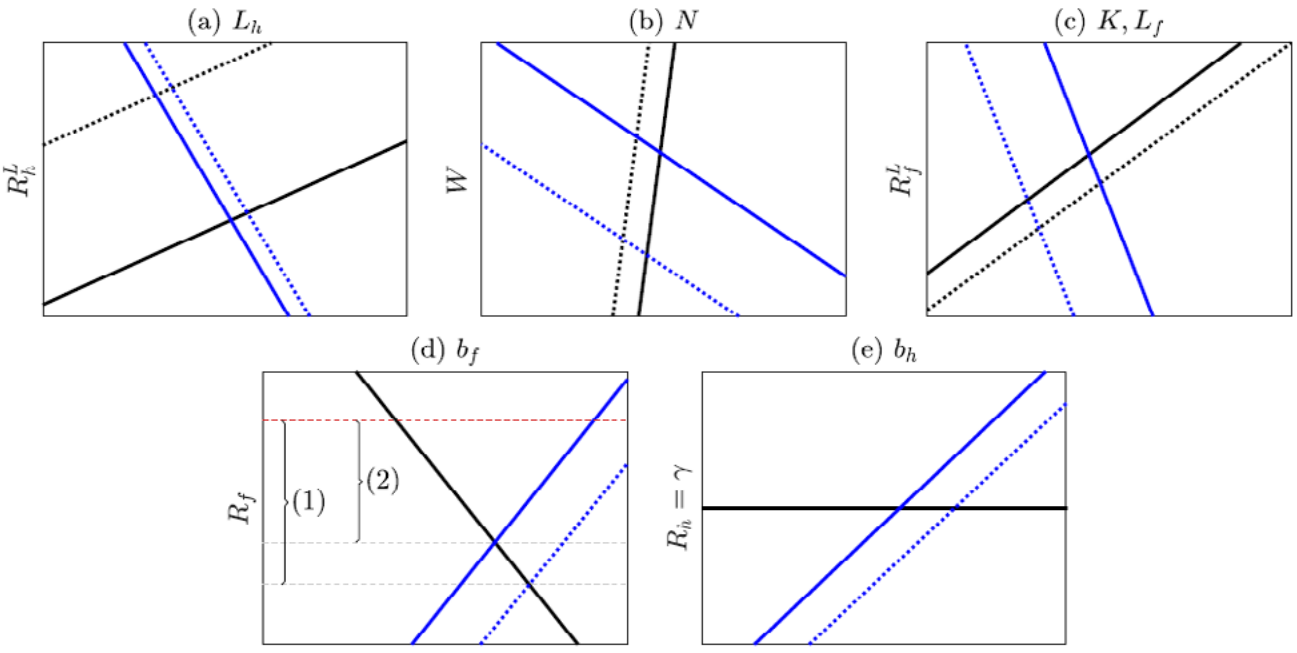
$$\min_{\{\widetilde{\psi}_{h1}\}} \quad \left| \mathcal{W}^h(\psi_{f1}, \widetilde{\psi}_{h1}) - \mathcal{W}^h(\psi_{f0}, \psi_{h0}) \right|$$

$$where \quad \mathcal{W}^h(\psi_f, \psi_h) = \mathcal{W}^{wh}(\psi_f, \psi_h) + \mathcal{W}^{bh}(\psi_f, \psi_h) + \mathcal{W}^{gh}(\psi_f, \psi_h).$$

Definition 3 states that the main objective of local monetary policy responses is to leave the welfare levels of the local economy at a fixed value of a given arbitrary initial state 0. As a matter of fact, this rule permits local monetary policy to fully counteract the effects of global monetary shocks over local welfare. That said, local monetary policy responses will have heterogeneous effects over the two elements that determine local aggregate welfare: production (that ultimately determines consumption levels) and consumption volatility. To see this, recall that total local welfare W^h is the sum

Figure 3.

Local monetary policy shock: Effects over markets of local response to hegemon's spillovers



Note: Figure shows market plots. Black curves identify supplies. Blue curves identify demands. Dotted lines identify previous equilibrium after hegemon monetary contraction, while solid lines identify new equilibrium after non-hegemon monetary expansion that increases funding spread ψ_h (response). The red dashed line in panel (d) identifies public investment return γ.

of the welfare levels of local workers W^{wh}, local bankers W^{bh} and the local government W^{gh}

$$\mathcal{W}_h = \underbrace{C_1^{wh} + C_2^{wh}}_{\mathcal{W}^{wh}} + \underbrace{\mathrm{E}\left[C_2^{gh}\right] - \mathrm{Var}\left[C_2^{gh}\right]}_{\mathcal{W}^{bh}} - \underbrace{C_2^{gh}}_{\mathcal{W}^{gh}},$$

which finally depend both on consumption levels and on consumption volatility.

Following the simple monetary policy rule proposed in definition 3, I take the situation before the hegemon monetary policy shock as the benchmark state –state 0 in definition 3– and study how does the local monetary response to the hegemon monetary contraction affects the equilibrium.

The effect of monetary policy response over main markets of the model is shown in figure 3. Panel (a) and (b) identify the direct effect of local monetary responses over local loans and local factor markets. As local liquidity conditions relax, local loans supply rises and the price of local loans declines. This generates an increase in local factor demand –labor N – as cheaper local loans can be used to pay for labor services. In equilibrium, there are also both a small reduction in the loans demand curve triggered by higher salaries and an increase in labor supply due to labor substitution effects, which are not paramount in the model and thus, will be sidelined.

As happened with global monetary policy shocks, increases in labor contracting of one factor of production rise demand for the other factor due to complementarity. When labor contracting increases, firms seek to contract higher levels of capital, causing capital demand to shift to the right and driving global loans return rate up. This is shown in panel (c). In equilibrium, there is a small drop in global loan supply so that this curve is pushed to the left. This result will be sidelined as is not the main focus of the paper. However, it is possible to make an educated guess and affirm that this phenomenon is a result of the fact that not only global loan rates, but also global safe rates are rising. As safe rates increase, the opportunity cost of creating global loans is higher so that global loan supply shifts to the left.

Panel (d) illustrates the way in which local monetary expansions affect global safe assets market equilibrium. As easier local liquidity conditions increase capital demand as well as its price, global banks are willing to divert funds from safe hegemon government debt to risky equity destined to create global loans. This reduces the demand of global safe assets, displacing its curve leftwards and rising the costs that the hegemon government should pay investors so that they are willing to accept holdings of its debt. This brings down the difference between the red dashed line –which as before identifies public investment return γ– and equilibrium debt cost, which is equivalent to a reduction in the exorbitant privilege.

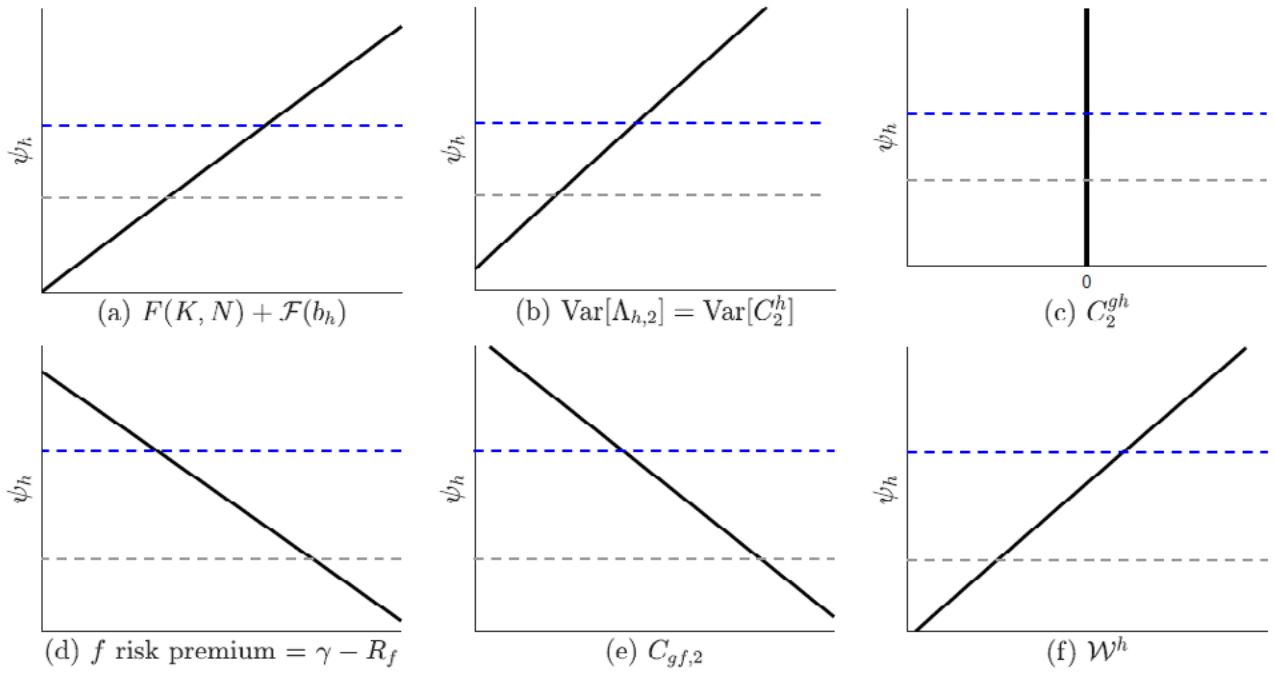
Panel (e) identifies the effects of the shock over the local safe asset market. As local bank's funding costs decrease and its incentives to create local loans increase, equilibrium local safe debt levels decrease. This happens as local banks are only able to increase local loans as long as they increase their equity holdings, what entails a reduction their local goverment debt holdings. As previously discussed, the cost of local government debt R_h remains constant.

Figure 4 illustrates equilibrium relations between local monetary policy levels and interest variables of the model, where the grey dashed lines indicate the funding spread ψh before the monetary shock while blue dashed lines indicate the local response to the hegemon's monetary contraction.

Following panel (a), local monetary response increases production levels. As discussed at the beggining of this section, this is consistent with the idea that local monetary authorities use their policy rates to counteract hegemon's production spillovers. Panel (b) shows that production boosting comes at the cost of increased local banker's portfolio volatility and consumption volatility as higher local loans are accompanied by

Figure 4.

Local monetary policy shock: Effects over interesting variables of local response to hegemon's spillovers



Note: Black solid lines identify equilibrium relations between referenced variables. Grey dashed lines identify local currency wholesale funding spread ψ_h before the shock while blue dashed lines identify the spread after the shock.

In open economy settings, this phenomenon may support the idea pointed out by Bordo and James(2017)

an increase in risky local equity holdings by local banks. In this sense, the local monetary response leads to a trade-off between boosting local production and reducing local consumption volatility. In closed economy settings, a similar dilemma between output boosting and financial vulnerabilities has been recently highlighted both by the works of Coimbra and Rey (2017) and Borio et al. (2019). In open economy settings, this phenomenon may support the idea pointed out by Bordo and James (2017) whereby a new dimension should be added to the traditional Mundellian trilemma, namely financial stability. For this reason, even if local monetary policy is autonomous in the sense of independent setting of funding costs, it is not autonomous in the sense of foreign shock isolation, as formerly suggested by traditional Mundellian arguments[16].

Panel (c) shows the dynamics of local government consumption which, as already discussed, is fixed at zero. Panels (d) and (e) depict local monetary policy effects over the exorbitant privilege. Local monetary shocks are associated with a decrease in hegemon's debt demand, so that the cost on this agent's debt will increase. The counterpart of this would be a reduction in risk premium, that in turn reduces the foreign government's monopoly/safety rent and its consumption levels. This last element highlights a possible source of local monetary spillovers over the hegemon's economy, and supports the idea that there may be gains of international monetary cooperation, not only for the local/non-hegemon economy but also for the global/hegemon economy.

Finally, panel (f) exhibits local monetary policy response effects over local welfare. As local monetary conditions relax and production increases, aggregate local welfare rises too, returning to its initial levels before the hegemon monetary policy shock. This is in fact consistent with the way in which the local monetary policy rule presented in definition 3 works.

4. Conclusions

This paper develops a novel framework that integrates monetary policy spillovers in an open economy model through safe asset market dynamics, characterizing the way in which the *safety-appetite* mechanism of transmission of monetary policy operates internationally, as well as its repercussions on international monetary policy spillovers.

[16] See Mundell (1963) and Fleming (1962) for the traditional Mundellian interpretation of monetary autonomy in open economies and Galí and Monacelli (2005) for New Open Economy Models interpretation of Mundellian arguments.

Monetary policy acts by changing the relation between risky asset returns and perceived risks, what alters portfolio decisions of risk-averse investors. In turn, this determines credit creation and factor contracting. Global currency funding needs by global banks confer a special role to the monetary policy of global currency issuers, so that monetary policy shocks of these authorities –referred as global monetary shocks– have repercussions over aggregate conceded loans by global banks. In a world with segmented local and foreign loan markets –i.e. with no perfect substitution between local and foreign loans–, where loans are an essential element of production, global loan shocks triggered by global monetary shocks have real effects over production. These are the *production* spillovers of monetary policy.

As global monetary shocks change portfolio composition of bankers, they also change the risks taken by these agents, what has consequences over aggregate risks. The repercussions over risk-taking of banks lie beyond countries' borders. These are the *risk-taking* spillovers of monetary policy.

The existence of international monetary spillovers drives local monetary policy to respond to foreign monetary shocks, what entails an inherent trade-off between increasing production and reducing consumption volatility. As global monetary conditions tighten, local welfare as well as local production declines leading local monetary authorities to run expansive monetary policy that increases welfare and production again by enhancing local loan creation. However, this will cause risky equity holdings of local agents –namely local banks– to increase, rising aggregate consumption volatility in the local economy.

Monetary conditions, both global and local, determine risk-averse investors' willingness to hold safe debt. In a setting where a particular government –the hegemon government– is a monopolist issuer of safe debt, changes in willingness to hold its debt modify its equilibrium levels of consumption, or its exorbitant privilege.

Finally, hegemon's monetary spillovers over local production and risk-taking, in tandem with local monetary spillovers over hegemon's government consumption, may give rise to arguments favouring the idea that there exist benefits associated with international monetary policy cooperation. Still, the rigorous characterization of the costs and gains of this proposal lie outside the scope of this work. Research strands on this particular area appear to be an attractive area of growing research (Agénor and Pereira da Silva, 2019; Matschke, 2020).

References

Agénor, Pierre Richard and Luiz Awazu Pereira da Silva, "Global Banking, Financial Spillovers, and Macroprudential Policy Coordination," Technical Report, Bank for International Settlements 2019.

Avdjiev, Stefan, Robert N. McCauley, and Hyun Song Shin, "Breaking free of the triple coincidence in international finance," Economic Policy, 2016, 31 (87), 409–451.

Barro, Robert J., "Rational Expectations and the Role of Monetary Policy," Journal of Monetary Economics, 1976, 2 (1), 1–32.

Basu, Suman, Emine Boz, Gita Gopinath, Francisco Roch, and Filiz Un- sal, "A Conceptual Model for the Integrated Policy Framework," Technical Report, International Monetary Fund 2020.

Bernanke, Ben and Mark Gertler, "Inside the Black Box: The Credit Channel of Monetary Policy Transmission," Journal of Economic Perspectives, 1995, 9 (4), 27–48.

,_ , _ , and Simon Gilchrist, "The financial accelerator in a quantitative business cycle framework," Handbook of Macroeconomics, 1999, 1 (C), 1341–1393.

Bordo, Michael and Harold James, "Capital Flows and International Order: Trilemmas and Trade-Offs from Macroeconomics to Political Economy and International Relations," in Atish Ghosh and Mahvash Qureshi, eds., From Great Depression to Great Recession: The Elusive Quest for International Policy Cooperation, Washington D.C., USA: International Monetary Fund, 2017, chapter 5, pp. 91–108.

Borio, Claudio and Haibin Zhu, "Capital regulation, risk-taking and monetary policy: A missing link in the transmission mechanism?," Journal of Financial Stability, 2012, 8 (4), 236–251.

_ , _ , Piti Disyatat, and Purichai Rungcharoenkitkul, "Monetary Policy Hysteresis and the Financial Cycle," Technical Report, Bank of International Settlements 2019.

, Robert N. McCauley, Patrick McGuire, and Vladyslav Sushko, "Covered interest parity lost: understanding the cross-currency basis," Technical Report, Bank of International Settlements 2016.

Bruno, Valentina and Hyun Song Shin, "Capital flows and the risk-taking channel of monetary policy," Journal of Monetary Economics, 2015, 71, 119–132.

Caballero, Ricardo, "On the Macroeconomics of Asset Shortages," 2006.

,_ , _ , and Emmanuel Farhi, "The Safety Trap," Review of Economic Studies, 2017, 85 (1), 223–274.

,_ , _ , and Pierre Olivier Gourinchas, "Global Imbalances and Policy Wars at the Zero Lower Bound," Technical Report, National Bureau of Economic Research 2015.

,_ , _ , and , "Safe Asset Scarcity and Aggregate Demand," American Economic Review, 2016, 106 (5), 513–518.

Cohen, Benjamin H., Dietrich Domanski, Ingo Fender, and Hyun Song Shin, "Global Liquidity: A Selective Review," Annual Review of Economics, 2017, 9 (1), 587–612.

Coimbra, Nuno and Hélène Rey, "Financial Cycles with Heterogeneous Intermediaries," Technical Report, National Bureau of Economic Research, Cambridge, MA 2017.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, "The Deposits Channel of Monetary Policy," The Quarterly Journal of Economics, 2017, 132 (4), 1819–1876.

Du, Wenxin, Joanne Im, and Jesse Schreger, "The U.S. Treasury Premium," Journal of International Economics, 2018, 112, 167–181.

Eichengreen, Barry, Exorbitant Privilege. The Rise and Fall of the Dollar and the Future of the International Monetary System, Oxford, UK: Oxford University Press, 2010.

Farhi, Emmanuel and Matteo Maggiori, "A Model of The International Monetary System," The Quarterly Journal of Economics, 2017, 133 (1), 295–355.

Fleming, John Marcus, "Domestic Financial Policies under Fixed and under Floating Exchange Rates," Technical Report, Staff Papers (International Monetary Fund) 1962.

Fornaro, Luca and Federica Romei, "The Paradox of Global Thrift," American Economic Review, 2019, 109 (11), 3745–3779.

Gabaix, Xavier and Matteo Maggiori, "International Liquidity and Exchange Rate Dynamics," The Quarterly Journal of Economics, 2015, 130 (3), 1369–1420.

Galí, Jordi and Tommaso Monacelli, "Monetary policy and exchange rate volatility in a small open economy," Review of Economic Studies, 2005, 72 (3), 707–734.

Gerko, Elena and Hélène Rey, "Monetary Policy in the Capitals of Capital," Journal of the European Economic Association, 2017, 15 (4), 721–745.

Gertler, Mark and Nobuhiro Kiyotaki, "Financial intermediation and credit policy in business cycle analysis," Handbook of Monetary Economics, 2010, 3 (C), 547–599.

and Simon Gilchrist, "The Role of Credit Market Imperfections in the Monetary Transmission Mechanism: Arguments and Evidence," The Scandinavian Journal of Economics, mar 1993, 95 (1), 43.

Goldberg, Linda and Cédric Tille, "Macroeconomic interdependence and the inter- national role of the dollar," Journal of Monetary Economics, 2009, 56 (7), 990–1003.

Gopinath, Gita, "The International Price System," Technical Report, National Bureau of Economic Research, Cambridge, MA 2015.

, Emine Boz, Camila Casas, Federico Díez, Pierre Olivier Gourinchas, and Mikkel Plagborg-Møller, "Dominant Currency Paradigm," American Economic Review, 2020, 110 (3), 677–719.

Gorton, Gary, "Private Clearinghouses and the Origins of Central Banking," Technical Report, Federal Reserve Bank of Philadelphia, Philadelphia, PA 1984.

and Lixin Huang, "Bank Panics and the Endogeneity of Central Banking," Journal of Monetary Economics, 2006, 53 (7), 1613–1629.

Gourinchas, Pierre Olivier, "The Dollar Hegemon? Evidence and Implications for Policy Makers," 2019.

and Hèléne Rey, "From World Banker to World Venture Capitalist: US External Adjustment and the Exorbitant Privilege," Technical Report, National Bureau of Economic Research 2007.

and Olivier Jeanne, "Global Safe Assets," Technical Report, Bank for Interna tional Settlements 2012.

, Héléne Rey, and Kai Truempler, "The Financial Crisis and the Geography of Wealth Transfers," Journal of International Economics, 2012, 88 (2), 266–283.

, , and Maxime Sauzet, "The International Monetary and Financial System," Technical Report, National Bureau of Economic Research, Cambridge, MA 2019.

, , and Nicolas Govillot, "Exorbitant Priviledge and Exorbitant Duty," Technical Report, Institute for Monetary and Economic Studies, Bank of Japan 2010.

, Walker Ray, and Dimitri Vayanos, "A Preferred-Habitat Model of Term Premia and Currency Risk," Technical Report, UC Berkeley 2020.

Greenwood, Robin, Samuel Hanson, and Jeremy Stein, "A Comparative-Advantage Approach to Government Debt Maturity," The Journal of Finance, 2015, 70 (4), 1683–1722.

Hill, Enoch and David Perez–Reyna, "Leverage Away Your Wedge: An Analysis of Banks' Impact on Output," Technical Report, Universidad de Los Andes - CEDE, Bogota, Colombia 2016.

and , "Financial development and occupational choice," Journal of Macroeconomics, 2017, 54 (B), 393–409.

Jiang, Zhengyang, Arvind Krishnamurthy, and Hanno Lustig, "Dollar Safety and The Global Financial Cycle," Technical Report, National Bureau of Economic Research, Cambridge, MA 2020.

, , and , "Foreign Safe Asset Demand and the Dollar Exchange Rate," Technical Report, National Bureau of Economic Research, Cambridge, MA 2020.

Jordá, Óscar, Moritz Schularick, Alan Taylor, and Felix Ward, "Global Financial Cycles and Risk Premiums," Technical Report, National Bureau of Economic Research, Cambridge, MA 2018.

Kekre, Rohan and Moritz Lenel, "Dollar Demand and International Risk Sharing," Technical Report 2020.

and , "Monetary Policy, Redistribution, and Risk Premia," Technical Report 2020.

Keynes, John Maynard, The General Theory of Employment, Interest, and Money, 1 ed., London: Macmillan & Co, 1936.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen, "The Aggregate Demand for Treasury Debt," Journal of Political Economy, 2012, 120 (2), 233–267.

and Hanno Lustig, "Mind the Gap in Sovereign Debt Markets: The U.S. Treasury basis and the Dollar Risk Factor," Technical Report, Jackson Hole Symposium of the Federal Reserve Bank of Kansas City 2019.

Laughlin, James Laurence, "The Aldrich-Vreeland Act," Journal of Political Economy, 1908, pp. 489–513.

Lenel, Moritz, "Safe Assets, Collateralized Lending and Monetary Policy," Technical Report 2017.

Lucas, Robert E., "Expectations and the Neutrality of Money," Journal of Economic Theory, 1972, 4 (2), 103–124.

Maggiori, Matteo, Brent Neiman, and Jesse Schreger, "The Rise of the Dollar and Fall of the Euro as International Currencies," AEA Papers and Proceedings, 2019, 109, 521–526.

, , and , "International Currencies and Capital Allocation," Journal of Political Economy, 2020, 128 (6), 2019–2066.

Matschke, Johannes, "National Interests, Spillovers and Macroprudential

Coordination," Technical Report 2020.

Miranda-Agrippino, Silvia and Hélène Rey, "US Monetary Policy and the Global Financial Cycle," Technical Report, National Bureau of Economic Research, Cambridge, MA 2015.

Miron, Jeffrey, "Financial Panics, the Seasonality of the Nominal Interest Rate, and the Founding of the Fed," The American Economic Review, 1986, 76 (1), 125–140.

Mundell, Robert, "Capital Mobility and Stabilization Policy under Fixed and Flexible Exchange Rates," The Canadian Journal of Economics and Political Science, 1963, 29 (4), 475–485.

Nagel, Stefan, "The Liquidity Premium of Near-Money Assets," The Quarterly Journal of Economics, 2016, 131 (4), 1927–1971.

Ocampo, José Antonio, Resetting The International Monetary (Non) System, Oxford, UK: Oxford University Press, 2017.

Rey, Hélène, "Dilemma not Trilemma: The Global Financial Cycle and Monetary Policy Independence," Technical Report, National Bureau of Economic Research, Cambridge, MA 2015.

, "International Channels of Transmission of Monetary Policy and the Mundellian Trilemma," Technical Report, National Bureau of Economic Research, Cambridge, MA 2016.

Rogoff, Kenneth, "Can international monetary policy cooperation be counterproductive?," Journal of International Economics, 1985, 18 (3-4), 199–217.

Sargent, Thomas and Neil Wallace, ""Rational" Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," Journal of Political Econ- omy, 1975, 83 (2), 241–254.

Sheng Shen, Leslie, "Global vs. Local Banking: A Double Adverse Selection Problem," 2019.

Warburg, Paul, The Federal Reserve System: Its Origin and Growth, New York: Macmillan & Co, 1930.

Appendix

Appendix A. Market Clearing in Goods Markets

Market clearing condition in period 1 implies

$$\underbrace{\omega + l_h}_{\text{domestic goods supply}} = \underbrace{C_1^{wh} + b_h + K}_{\text{domestic absorption}} - \underbrace{(\omega + l_f + d_f - b_f)}_{\text{net imports}}.$$

As there is no production in this period, resources come either from banker's wealth endowments ω, from local/ global wholesale funding –funds which are provided elastically by central banks–, and from global deposits – provided elastically in the international capital markets–. Local aggregate supply equals local banker's wealth and local wholesale funding, as shown in the left hand side of the previous equation. Domestic absorption equals local household's goods purchases, local banker's safe asset purchases and capital. Governments and bankers only consume in the second period. All foreign resources that fund the local economy are grouped in the net imports term.

Replacing (3.1), (3.2), (3.3), the first period constraint of (2.1) and the first period constraint of (2.3) we get

$$\omega + \lambda_2 s_h = WN - d_h + \omega - s_h + (\lambda_1 + \lambda_2 + 1)\, s_f - \left[\omega + \lambda_2 s_f + \lambda_1 s_f - (\omega - s_f)\right],$$

canceling terms

$$\lambda_2 s_h = WN - d_h - s_h.$$

Recall that equation (3.1) states that $d_h = \lambda_1 s_h$, such that one may reorganize the previous expression as

$$(1 + \lambda_2 + \lambda_1)\, s_h = WN,$$

where $WN = L_h$ as local loans are equal to total labor payments. Thus,

$$(1 + \lambda_2 + \lambda_1)\, s_h = L_h,$$

which is true[17], such that first period budget constraint holds.

For the second period, market clearing requires that

$$F(K,N) + \mathcal{F}(b_h) - \Omega\delta s_h = \left[C_2^{wh} + C_2^{bh} + C_2^{gh} + R_h^d(1-\psi_h)l_h \right] + \left[R_f^d d_f + C_2^{bf} + C_2^{gf} + R_f^d(1-\psi_f)l_f - (\mathcal{F}(b_f) - \Omega\delta s_f) \right].$$

Local aggregate supply –left hand side– is equivalent to firms and local government production, less the local equity random shock. Domestic absorption –first term of right hand side– equals local workers', bankers' and government's consumption plus local wholesale funding payments. Net exports –second term of the right hand side– account for foreign demand for goods –the sum of payments to foreign deposits, global bankers' consumption, foreign goverment's consumption and payments for foreign wholesale funding– less foreign aggregate supply –which equals foreign government production less global equity random shock–. Replacing by firm's production function, household's second period constraint shown in (2.1), bankers' consumption/portfolio value shown in (2.3) and governments' production function shown in (2.5) we get

$$ZK^{\alpha}N^{1-\alpha} + \gamma b_h - \Omega\delta s_h = \left[R_h^d d_h + \pi_h + R_h b_h - \Omega\delta s_h + \gamma b_h - R_h b_h + R_h^d(1-\psi_h)l_h \right] + \left[R_f^d d_f + \pi_f + R_f b_f - \Omega\delta s_f + \gamma b_f - R_f b_f + R_f^d(1-\psi_f)l_f - \gamma b_f + \Omega\delta s_f \right],$$

canceling terms

$$ZK^{\alpha}N^{1-\alpha} = \left[R_h^d d_h + \pi_h + R_h^d(1-\psi_h)l_h \right] + \left[R_f^d d_f + \pi_f + R_f^d(1-\psi_f)l_f \right].$$

Replacing bank's benefits depicted in equation (3.4)

$$ZK^{\alpha}N^{1-\alpha} = \left[R_h^d d_h + \left(R_h^L - R_h^d\right)\lambda_1 s_h + \left(R_h^L - R_h^d(1-\psi_h)\right)\lambda_2 s_h + R_h^L s_h + R_h^d(1-\psi_h)l_h \right] + \left[R_f^d d_f + \left(R_f^L - R_f^d\right)\lambda_1 s_f + \left(R_f^L - R_f^d(1-\psi_f)\right)\lambda_2 s_f + R_f^L s_f + R_f^d(1-\psi_f)l_f \right],$$

replacing d_j and l_j for equations (3.1) and (3.2)

$$ZK^{\alpha}N^{1-\alpha} = \left[R_h^d\lambda_1 s_h + \left(R_h^L - R_h^d\right)\lambda_1 s_h + \left(R_h^L - R_h^d(1-\psi_h)\right)\lambda_2 s_h + R_h^L s_h + R_h^d(1-\psi_h)\lambda_2 s_h \right] + \left[R_f^d\lambda_1 s_f + \left(R_f^L - R_f^d\right)\lambda_1 s_f + \left(R_f^L - R_f^d(1-\psi_f)\right)\lambda_2 s_f + R_f^L s_f + R_f^d(1-\psi_f)\lambda_2 s_f \right],$$

[17] See equation (3.3).

and by canceling terms we get

$$ZK^{\alpha}N^{1-\alpha} = (\lambda_1 s_h \mid \lambda_2 s_h \mid s_h) R_h^L \mid (\lambda_1 s_f \mid \lambda_2 s_f \mid s_f) R_f^L. \qquad (\Lambda.1)$$

Prices R^L_b and R^L_f in (A.1) should be consistent with a solution for firms' problem. Solution to (2.2) gives the following first order conditions

$$R_h^L = \frac{(1-\alpha)Z}{W}\left(\frac{K}{N}\right)^{\alpha}$$

$$R_f^L = \alpha Z\left(\frac{N}{K}\right)^{1-\alpha},$$

both of which can be replaced in (A.1), obtaining

$$ZK^{\alpha}N^{1-\alpha} = (\lambda_1 s_h + \lambda_2 s_h + s_h)\left[\frac{(1-\alpha)ZK^{\alpha}}{WN}\cdot\frac{1}{N^{\alpha-1}}\right] + (\lambda_1 s_f + \lambda_2 s_f + s_f)\left[\frac{\alpha ZN^{1-\alpha}}{K}\cdot\frac{1}{K^{-\alpha}}\right].$$

Equation (3.3) states that $L_b = (\lambda_1 s_b + \lambda_2 s_b + s_b)$ and $L_f = (\lambda_1 s_f + \lambda_2 s_f + s_f)$. Total labor (capital) payments should equal local (foreign) loans, such that one may simplify the previous expression as

$$\begin{aligned} ZK^{\alpha}N^{1-\alpha} &= \left[(1-\alpha)ZK^{\alpha}\cdot\frac{1}{N^{\alpha-1}}\right] + \left[\alpha ZN^{1-\alpha}\cdot\frac{1}{K^{-\alpha}}\right], \\ &= ZK^{\alpha}N^{1-\alpha} \end{aligned}$$

proving that second period budget constraint holds.

Appendix B. Proof of Lemma 1

Replacing banks' balance sheet constraint in banks' profits, one obtains

$$\begin{aligned} \pi_j &= R_j^L(d_j + l_j + s_j) - R_j^d d_j - R_j^d(1-\psi_j)l_j \\ &= \left(R_j^L - R_j^d\right)d_j + \left(R_j^L - R_j^d(1-\psi_j)\right)l_j + R_j^L s_j. \end{aligned}$$

If $R^L_j < R^d_j$, bank's profits will be decreasing in the number of deposits demanded, so that bank's demand of deposits will be zero. In the interesting case where $R^L_j > R^d_j$, bank's profits are increasing both in deposits and on wholesale funding. The case for deposits is trivial. For wholesale funding, note that spread $\psi_j \in (0, 1)$ such that $R^d_j > R^d_j(1 - \psi_j)$, what indicates that $R^L_j > R^d_j(1 - \psi_j)$. As a corollary, bank profits are increasing in wholesale funding l_j.

Banks will then demand deposits and wholesale funds up to the maximum limit they are allowed, so that their leverage and capital requirements constraints bind.

Appendix C. Proof of Proposition 1

Without leverage constraints and capital requirements, the problem of bank $j \in f, h$ becomes

$$\max_{\{L_j, d_j, l_j\}} \pi_j(s_j) = R^L_j L_j - R^d_j d_j - R^d_j(1 - \psi_j) l_j \qquad \text{(C.1)}$$

subject to the balance sheet constraint

$$L_j = d_j + l_j + s_j. \qquad \text{(C.2)}$$

Replacing equation (C.2) in (C.1) and reorganizing terms, one obtains the following expression for the bank's problem

$$\max_{\{d_j, l_j\}} \pi_j(s_j) = \left(R^L_j - R^d_j\right) d_j + \left(R^L_j - R^d_j(1 - \psi_j)\right) l_j + R^L_j s_j. \qquad \text{(C.3)}$$

The analysis will only focus on the interesting case in which $R^L_j \geq R^d_j(1 - \psi_j)$, as cases in which wholesale fund rates are higher to loan rates are uncommon.

Lemma C.1. *In the case in which monetary policy sets $\psi_j > 0$, for an equilibrium to exist, it should be the case that $R^L_j = R^d_j(1 - \psi_j)$ and banks will elastically demand wholesale funds while demanding zero deposits.*

Proof: If $R^L_j > R^d_j(1 - \psi_j)$, banks will demand infinite amount of wholesale funding, as profits –depicted in equation (C.3)– will be increasing in this variable. This is not consistent with an equilibrium.

By contrast, if $R^L_j = R^d_j(1 - \psi_j)$, banks will be indifferent in the amount of wholesale funds that they demand. Given this fact, equilibrium in wholesale funding markets indicates that these agents should demand all wholesale funds supplied at prevalent return rates. This is consistent with an elastic demand for wholesale funding. Equality between loan return rates and wholesale funding rates implies that $R^L_j < R^d_j$. This is evident by recalling the initial assumption in lemma C.1according to which $\psi_j > 0$. In this case, bank's profits denoted by equation (C.3) are decreasing in deposits levels, so that banks will demand zero deposits.

In the case in which there exists a positive spread between deposits' cost and wholesale funding costs, one can draw conclusions on the effective supply of loans by using lemma C. 1in conjunction with bank's budget constraint, equation (C.2). Let ls denote the supply of wholesale funds (which do not depend on bank's decisions), loan supply in the case where $\psi_j > 0$ will be equal to

In this case, loan supply is not entirely determined by bankers' equity allocations

$$L_j = l_j^s + s_j$$

Note that, in this case, loan supply is not entirely determined by banker's equity allocations sj. In this sense, portfolio decisions, and specially risk-aversion of bankers, is not a fundamental driver of loan creation and credit.

Lemma C.2. *In the case in which monetary policy set $\psi_j = 0$, for an equilibrium to exist, it should be the case that $R_j^L = R_j^d$ and banks will elastically demand both wholesale funds and deposits.*

Proof: Analogous to the case depicted in lemma C.1. If $R_j^L > R_j^d$, banks will demand both infinite amount of deposits and infinite amount of wholesale funds, as profits – depicted in equation (C.3)– will be increasing in both variables. This is not consistent with an equilibrium. If $R_j^L = R_j^d$, banks will be indifferent both in the amount of deposits and in the amount of wholesale funds that they demand. Given this fact, equilibrium in deposits and wholesale funding markets suggests that these agents should demand all deposits and all wholesale funds supplied at prevalent return rates. This is consistent with both an elastic demand of deposits and of wholesale funding.

Conclusions on the effective supply of loans when the spread between deposits and wholesale funding is zero can be drawn by using lemma C.2 in conjunction with bank's budget constraint, equation (C.2). Let d^s denote the supply of deposits (which do not depend on bank's decisions), loan supply in the case where $\psi_j = 0$ is equal to

$$L_j = d_j^s + l_j^s + s_j, \tag{C.4}$$

such that loans are not entirely determined by equity allocations s_j nor by banker's portfolio decisions and risk-aversion.

Appendix D. Banker's Optimality Condition

Replacing banker's portfolio value in its utility function –see equation (2.3)–, one may express banker's welfare W^{bj} as

$$\begin{aligned}\mathcal{W}^{bj} &= \mathrm{E}\left[\Lambda_{j,2}\right] - \frac{\chi}{2}\mathrm{Var}\left[\Lambda_{j,2}\right] \\ &= \mathrm{E}\left[\pi_j(s_j) + R_j b_j - \Omega\delta s_j\right] - \frac{\chi}{2}\mathrm{Var}\left[\pi_j(s_j) + R_j b_j - \Omega\delta s_j\right], \qquad \text{(D.1)}\end{aligned}$$

such that banker's utility ultimately depends on both the expected value and the variance of its portfolio holdings.

Equation (3.4) states that

$$\pi_j(s_j) = \left(R_j^L - R_j^d\right)\lambda_1 s_j + \left[R_j^L - R_j^d(1-\psi_j)\right]\lambda_2 s_j + R_j^L s_j.$$

The model assumes that loans, public debt and deposits contracts are enforceable, such that there is no uncertainty on private loans return rates R^L_j, public debt return rates R_j, nor on deposit interest rates R^d_j. This also implies that there is no uncertainty in bank's profits $\pi_j(s_j)$ such that one may simplify equation (D.1) as

$$\begin{aligned}\mathcal{W}^{bj} &= \pi_j(s_j) + R_j b_j - \mathrm{E}\left[\Omega\delta s_j\right] - \frac{\chi}{2}\mathrm{Var}\left[\Omega\delta s_j\right] \\ &= \pi_f(s_j) + R_j b_j - \delta s_j \mathrm{E}(\Omega) - \frac{\chi}{2}\,\delta^2\, s_j^2\,\mathrm{Var}(\Omega).\end{aligned}$$

Replacing b_j using banker's first period budget constraint, we obtain an expression for W^{bj} that uniquely depends on banker's equity allocations s_j

$$\mathcal{W}^{bj} = \pi_j(s_j) + R_j(\omega - s_j) - \delta s_j \mathrm{E}(\Omega) - \frac{\chi}{2}\delta^2 s_j^2 \mathrm{Var}(\Omega),$$

so that global bankers problem may be written as

$$\max_{\{s_j\}} \quad \mathcal{W}^{bj} = \pi_j(s_j) + R_j(\omega - s_j) - \delta s_j \mathrm{E}(\Omega) - \frac{\chi}{2}\delta^2 s_j^2 \mathrm{Var}(\Omega).$$

Random equity shock Ω is independent of equity allocations, such that first order condition of the previous problem is

$$\pi_j{}'(s_f) = R_f + \delta s_f \mathrm{E}(\Omega) + s_f \chi \delta^2 \mathrm{Var}(\Omega),$$

which is equivalent to equation (3.5).

Hysteresis From Monetary Policy Mistakes: How Bad Could It Be?

March 26, 2021

Andrés O. Dávila

*I owe special gratitude to Hernando Zuleta for his devoted guidance, advice, patience and unconditional support in all the steps of this paper. I am also grateful to Santiago Caicedo, Andrés Alvarez, Marc Hofstetter, Andrés Zambrano, Manuel Fernández, Jimena Hurtado, Luis Eduardo Arango, Tomás Rodríguez, David Perez-Reyna, Jose Ignacio Lopez, Miguel Urrutia, Adriana Camacho, Matilde Angarita, Sergio González, Andrea Figueroa, Andrés Gallegos, Diego Botero, Sergio Florez, Andrea Salazar, José Miguel Quintero, Daniel Espinosa, Mateo Rodriguez, Sofia Castro, two anonymous referees from the Asobancaria contest for the best M.Sc. Thesis in Banking and Monetary topics, and seminar participants at LAMES, Universidad de Los Andes, Asobancaria and Universidad del Rosario for helpful insights and useful suggestions. This paper is forthcoming as a Working Paper of the Documentos CEDE series at Universidad de Los Andes. Please, contact me at my E-mail if you wish to cite this paper.
† E-mail: ao.davila10@uniandes.edu.co.

Abstract

What would happen if the central bank makes a mistake facing a crisis? This paper argues that it would leave scars in the long-run trend of production. If monetary policy is not expansionary-enough during crises, an inefficient rise of the interest rate intensifies the scarring effects of recessions. The hysteresis effect comes from higher innovation costs that induce a drop in productivity growth, an indiscriminate firms' exit process, and a rise in unemployment. This article presents a theoretical model that rationalizes these mechanisms. The theory suggests that, in the long-run, even though growth recovers to its pre-shock rate and the economy converges to full firms' survival and full employment, the long-term output level is persistently lower than the level it would have reached in the absence of errors. How costly are monetary policy mistakes? The short answer is that they are infinitely costly.

Keywords: Hysteresis, Monetary Policy, Endogenous Growth, Productivity, Firms' Exit, Unemployment.

JEL Codes: E52, E58, O11, O40, O41, O42, O47.

I. Introduction

What would happen if the central bank makes a mistake facing a recession? Could this have persistent effects on economic activity? This paper presents a long-run endogenous growth model with non-neutral monetary policy to assess these inquiries. The research question this article intends to answer is if monetary policy mistakes can lead to hysteresis, which refers to a permanent deviation of the long-run production trend from its pre-shock level after a recession. A monetary policy mistake is a non-expansionary-enough response to recessions. For instance, the deepest possible error is a monetary policy grip during a crisis. In the model, recessions lead to hysteresis depending on the reaction of the monetary authority. In particular, a central bank focused solely on inflation during bad times makes a monetary policy mistake that generates an inefficient rise of the interest rate, which intensifies the scarring effects of recessions. On the contrary, an optimal countercyclical monetary stimulus during crises saves the economy from hysteresis while hitting the inflation target in the long-term.

Nonetheless, why shouldn't the central bank react with enough strength to recessions? Is it possible that monetary policy deviates from the most convenient response? It depends on the flexibility and discretion the monetary authority enjoys. In particular, government failures that bind the central bank to a regime focused only on inflation targeting, even during bad times, could restrain countercyclical monetary policy and escalate hysteresis. In this sense, the regime that governs central banking goals could be inconsistent with the demands of recessions, tying monetary policy to an inefficient answer.

The experience of OECD countries during the crises lived in the 1980s illustrates this argument (Ball et al., 1999).[1] One stylized fact stands: the central banks that place most of their efforts in attaining the ongoing crisis prevent output loss and a structural change of the unemployment level (Ball, 1997; Ball et al., 1999; Ball, 2001). On the contrary, the countries where monetary policy failed to resist the temptations to deal with inflationary pressures during the recessions suffered permanent costs on potential output and unemployment (Ball, 1997, 2001).[2]

Moreover, evidence around the responsibility of monetary policy in the scarring effects of recessions has motivated a recent concern in central banking: should monetary policy include the long-run economic performance among its goals? (Summers, 2014; Wilkins, 2014; Tucker, 2017; Coeure, 2017; Orphanides and Williams, 2011). This paper suggests that exercising stabilization policy focused on protecting the economy from hysteresis during bad times is crucial. Moreover, it contributes to a new debate around the nature of the models used in central banking (Adrian, 2018; Krugman, 2018), presenting a new model to examine this issue.

The theoretical approach follows Aghion and Howitt (1990) and shows permanent effects on the long-run trend from monetary policy decisions during recessions. The supply side of the economy involves three production sectors: *i.* ideas, *ii.* capital intermediate goods, and *iii.* a final good. The demand side comes from households subject to an exogenous shock that affects savings and changes loanable funds. The production of ideas uses credit from the savings market to invest in research and development (henceforth R&D) and innovate. The innovation then creates growth gains in

[1] There is also evidence of scarring effects in the aftermath of the great recession (Ball, 2014; Summers, 2014; Yagan, 2019). The relationship with monetary policy remains open for future research.

[2] Appendix A.1.1 describes another interesting example: the end-of-century crisis in Colombia during 1998 and 1999. Furthermore, Appendix A.1.2 presents a statement from an interview with Urrutia and a contrast with Yellen (2015). Urrutia emphasized that a lesson learned from the end-of-century crisis is that raising the short-run interest rate too much has adverse effects in the long-run.

the intensive margin, from vertical growth of each firm's productivity, and in the extensive margin, from horizontal growth of the firms' mass. Besides, the final good production is subject to downward real wage rigidities that generate unemployment due to firms' exit. Therefore, negative shocks that affect the ideas sector's resources hurt vertical and horizontal growth and create unemployment.

The long-run trend falls and never gets to recover to its pre-shock counterpart due to monetary policy mistakes. Thus, the model suggests that monetary policy should be sufficiently expansive during recessions to prevent hysteresis, demanding flexible central banking regimes that allow sacrificing inflation during crises.

In the face of these adverse consequences of shocks, a central bank that acts through the credit channel reacts following its monetary policy regime. Depending on the central bank's design, monetary policy can be countercyclical to avoid hysteresis or procyclical with the only purpose of hitting the inflation target. This reaction defines the scarring effects of recessions. Finally, in the long-run, after the shocks and the policy reactions, the economy grows in a Balanced Growth Path (henceforth BGP) unique general equilibrium.

The model's main result is that the absence of an expansionary-enough monetary policy during crises creates hysteresis. The scarring effects arise from higher innovation costs that induce a fall in productivity growth, an indiscriminate firms' exit process, and a rise in unemployment. Moreover, in the long-run, even though growth recovers to its pre-shock rate and the economy converges to full firms' survival and full employment, the long-term output level is persistently lower than the level it would have reached in the absence of errors. The long-run trend falls and never gets to recover to its pre-shock counterpart due to monetary policy mistakes. Thus, the model suggests that monetary policy should be sufficiently expansive during recessions to prevent hysteresis, demanding flexible central banking regimes that allow sacrificing inflation during crises.

Furthermore, the theory's fundamental mechanism relies on the need of the ideas sector to use resources from the savings market to innovate. There is evidence on the demand for credit of innovating firms to finance R&D expenditure and its relationship with the business cycle that sustains this assumption (Aghion et al., 2012a,b; Giebel and Kraft,

2018; Mancusi and Vezzulli, 2014; Lee et al., 2015). Precisely, in the model, the loanable funds market supplies credit for investment in R&D, charging a marginal cost for resources equal to the interest rate. Hence, when a recession destroys available resources, a mistaken reaction of the central bank leads to a surge in the interest rate and a higher cost of R&D expenditure, lowering the success probability of the creation of ideas. The lower frequency of innovations caused by the monetary policy error hurts average productivity and generates hysteresis.

The contribution of this paper is threefold. First, the model suggests a new channel for hysteresis in the falls of productivity among surviving firms, a mechanism overlooked by the literature. Second, the theory shows long-run effects of monetary policy through the credit channel. Third, this is the first paper that finds that even if in the long-run growth recovers and the economy converges to full employment and full firms' survival, monetary policy mistakes create hysteresis.

Therefore, this work is related to three strands of the literature. First, it contributes to the literature of the monetary policy effects on growth (Jorda et al., 2020; Aghion et al., 2012b; Bergoeing et al., 2002; Moran and Queralto, 2018; Younsi and Nafla, 2019). As revised by Twinoburyo and Odhiambo (2018), the vast majority of empirical papers find that monetary policy has a relationship with growth both in the short-run and in the long-run. Furthermore, recent contributions have widely challenged the long-run neutrality of monetary policy (Michl and Oliver, 2019; Garga and Singh, 2020; Jorda et al., 2020; Gali, 2020). This paper contributes to the theoretical alternatives used to understand this relationship with an endogenous long-run growth model that relates hysteresis to the central bank's decisions. Moreover, the central bank's modeling strategy is also innovative: it shows long-run effects of monetary policy operating through the credit market.

Second, the literature on hysteresis has focused on the effect of aggregate macroeconomic shocks, such as recessions, over unemployment, potential output, and firms' exit (Blanchard and Summers, 1986a; Cerra and Saxena, 2008; Abiad et al., 2009; Ball, 2014; Eslava et al., 2010; Dosi et al., 2018; Kienzler and Schmid, 2014; Logeay and Tober, 2006; Mourougane, 2017; Sturn, 2014). Ever since the seminal article of Blanchard and Sum-mers (1986b), a large number of papers have found long-term unemployment costs from short-run shocks. This paper's theoretical framework to analyze hysteresis is novel. It focuses on the long-run side of the issue, building over an endogenous growth model instead of on the usual Neo-Keynesian approach, which gives a more leading role to innovation (Garga and Singh, 2020; Dosi et al., 2010, 2018; Mourougane, 2017). Similarly, another contribution is the study of scarring effects with endogenous potential output growth, a significant feature of the model.

Besides, the literature on the scarring effects of recessions highlights the active role that central banking plays in the birth of hysteresis, centering on two primary mechanisms: firms' exit and unemployment (Ball, 1997; Stockhammer and Sturn, 2011; Kienzler and Schmid, 2014; Blanchard, 2018; Gali, 2020). Nonetheless, the literature on the non-neutrality of monetary policy in the long-run has focused on the effects of central banks' decisions in productivity (Jorda et al., 2020; Garga and Singh, 2020). Hence, this paper intends to provide a theoretical framework consistent with three mechanisms to explain hysteresis and its interaction with monetary policy: *i.* productivity costs, *ii.* firms' destruction, and *iii.* unemployment.

Finally, monetary policy literature has also explored the effects of policy decisions on unemployment and potential output hysteresis (Jorda et al., 2020; Garga and Singh, 2020; Gali, 2020; Aghion et al., 2012b; Michl and Oliver, 2019; Bilbiie et al., 2014; Chu and Ji, 2016; Moran and Queralto, 2018; Stockhammer and Sturn, 2011; Younsi and Nafla, 2019). The closest papers in this field differ in the focus of the analysis and the theoretical approach (Jorda et al., 2020; Garga and Singh, 2020; Gali, 2020; Aghion et al., 2012b; Stockhammer and Sturn, 2011). This paper's first difference is the distinction between three potential mechanisms for hysteresis in productivity, unemployment, and firms' exit. The second difference is the focus on the government failure side, due to monetary policy, rather than on the market failure side identified by financial constraints. The last difference is the emphasis on productivity and innovation as a critical channel where monetary policy interacts with the economy's long-run productive capacity.

This is the first paper that highlights the role of productivity costs from weaker vertical growth as a channel of the scarring effects of monetary policy actions during crises. The primary theoretical prediction emphasizes that, even if growth, employment, and firms survival fully recover after a recession, the productivity costs generate hysteresis. Thus, the results imply that changes in productivity growth among surviving firms during crises are a key channel to understand the scarring effects of recessions.

2. A Model of Hysteresis and Monetary Policy

This section presents the preferences and optimal choices of the households, the firms and the central bank in the model. Time in this setting is continuous[3]. In the long-run, the economy grows in a BGP unique equilibrium characterized by the natural interest rate r^* and a constant growth rate g^*. Depending on the monetary policy regime, the central bank's choices can keep the long-run trend in a situation consistent with the equilibrium or exacerbate the scarring effects of recessions.

2.1 Households

The economy is populated by N households comprising homogeneous agents. Each representative household seeks to maximize its expected utility U with CRRA preferences in an infinite horizon. The household is the owner of assets a_t, supplies labor L inelastically, and chooses consumption c_t. The budget constraint of the problem is subject to exogenous shocks ε_t that affect the value of assets.[4] Unemployment is distributed evenly on each household and operates only in the intensive margin. The simplifying assumption is that the household is only partially affected by unemployment, so a fraction ut is unemployed. The household problem is:

$$\max_{c_t} U = \mathbb{E}_t \int_0^{\infty} e^{-\rho t} \left[\frac{c_t^{1-\zeta} - 1}{1-\zeta} \right] \quad s.t. \quad \dot{a}_t = a_t r_t + a_t \varepsilon_t + w_t(1-u_t)L - c_t, \qquad (2.1)$$

where ρ is the impatience discount rate, ζ is an inverse measure of the intertemporal elasticity of substitution, r_t is the savings market interest rate that determines the return on assets, and w_t is the wage paid to the employed fraction of the household $(1-u_t)L$. The shock is i.i.d. $\varepsilon t \sim$ iid with mean 0 within a support $\varepsilon_t \in (-1, 1)$.[5]

The shock works like a demand shock as it alters household income through exogenous losses or gains in the value of assets. Depending on the magnitude of the shock ε_t, the economy has both recessions and booms around the long-run trend of the BGP equilibrium. The negative shocks constitute recessions since there exists a deep link between crises and asset losses that produce wealth destruction (De Nardi et al., 2011; Bosworth, 2012; Huo and Rios-Rull, 2013). The analysis focuses on recessions $\varepsilon_t < 0$, as this is the part of the cycle where hysteresis from monetary policy mistakes may rise.

The optimal decision of the household leads to the Euler equation:[6]

$$\frac{\dot{c}_t}{c_t} - \mathbb{E}_t \left[\frac{1}{\zeta}(r_t + \varepsilon_t - \rho) \right], \qquad (2.2)$$

[3] For any variable x_t in the model, define $\{x_t\}_{t=0}^{\infty} = \{x_t : t \in \mathbb{R}^{\geq 0}\}$. So each aggregate variable x_t evolves in continuous time.
[4] This technique is equivalent to shocking the returns of assets.
[5] This assumption is necessary to assure that the negative part of the cycle does not imply a total destruction of consumption $c_t = 0$. The values of the parameters in the shocks' distribution should be consistent with a support $\varepsilon_t \in (-1, 1)$.
[6] The transversality condition of optimal control theory holds, $\lim_{t\to\infty} \Lambda_t a_t = 0$ where Λ_t is the Hamiltonian multiplier.

so the household dissents from a smooth path of consumption because of changes in the interest rate r_t or shocks ε_t. A negative shock $\varepsilon_t < 0$ reduces savings as households try to compensate for the falls on current consumption. Hence, the uncertainty associated with recessions affects consumption volatility and depresses the value of the assets that the private banks receive as savings.

2.2 Firms

Firms in the model follow the structure of Aghion and Howitt (1990) and Romer (1990). This set enables endogenous growth both in the quality of each capital good variety, for vertical growth, and in the expansion of capital goods, for horizontal growth. There are three production sectors: *i.* ideas, *ii.* intermediate capital goods, and *iii.* a final consumption good. The interaction between kinds of firms is done in a dynamic strategic setting fulfilled in each period *t*. First, ideas firms invest in R&D using resources obtained from the savings market as a credit to innovate and acquire a patent on a specific quality of capital goods. If the ideas firms achieve to innovate, average productivity increases thanks to vertical and horizontal improvements. Next, a capital goods firm buys the patent from the successful ideas firm and obtains a monopoly thanks to property rights over the latest quality of varieties. This new quality of capital goods makes former versions of existing varieties obsolete, feeding creative destruction. Finally, the final good firm produces subject to downward wage rigidities using labor and a continuum of existing capital goods varieties. The supply side partial equilibrium is solved by backward induction.[7]

2.2.1 Final Good Firm

The final good firm produces the consumption good in a competitive market. Production y_t is done with Cobb-Douglas technologies using labor and a continuum of different varieties of capital goods supplied by a mass of intermediate firms. Each capital good has a specific quality and is produced by a unique firm with monopolistic power. The production process is subject to downward wage rigidities, which generate unemployment.[8] The final good firm solves:

[7] See Aghion and Howitt (2008); Klette and Kortum (2004); Aghion et al. (2014a) and Aghion et al.(2014b) for further details on the design of the strategic setting and more aspects of the Schumpeterian growth theory.

[8] Wage rigidities are modelled following the usual approach in the literature as explained by Blanchard and Gali (2010); Shimer (2005); Hall (2005); Gertler and Trigari (2009); Hall et al. (1975); Rees (1951) and Yellen (1995). The appeal of this assumption is that it allows observing involuntary unemployment, the rate u_t in the household problem.

$$\max_{E_t,k_{it}} \Pi_t = F_t^{1-\alpha} \int_{\phi_t}^{1} A_{il}^{1-\alpha} k_{it}^{\alpha} di - w_t F_t - \int_{\phi_t}^{1} p_{it} k_{it} di \quad s.t. \quad \frac{\dot{w}_t}{w_t} \geq \kappa. \qquad (2.3)$$

The firm maximizes its profits Π_t, chooses employed labor E_t, demands capital goods k_{it} with an specific productivity A_{it} for certain variety i, and pays wages wt and the capital goods price p_{it}. Supply of capital goods comes from intermediate firms and is used for production in a continuum of size $1 - \phi_t$, where ϕ_t is the quantity of destroyed intermediate firms. This assumption shows the channel of firms destruction that the model intends to provide. The parameter α represents an inverse measure of the degree of complementarity between the different types of capital goods. The intensity of wage rigidities depends on the rate of stickiness κ, which satisfies $-1 < \kappa \leq 0$.

The firm makes optimal decisions subject to the sticky wages constraint. When downward wage rigidities are binding $(\frac{\dot{w}_t}{w_t} = \kappa)$, the wage decreases at a constant rate κ rom an observe wage $\bar{w}_t$, setting the minimum wage possible in the labor market at each period. The wage $\bar{w}_t$ is the equilibrium wage of the labor market in the infinitesimally close previous period to a fall in the labor marginal product.[9] In the sticky wages scenario, demand for labor does not equilibrate to full employment. Note that κ represents the stickiness degree of wages, so that a high κ -closer to perfectly rigid wages $\kappa = 0$ increases unemployment costs from a negative shock to labor marginal product. When wage rigidities are not binding $(\frac{\dot{w}_t}{w_t} > \kappa)$, the economy operates in full employment $E_t = L$ and wages equal the marginal product of labor.

First order conditions give the demand functions for labor and capital goods,

$$w_t = \max\left((1-\alpha)E_t^{-\alpha} \int_{\phi_t}^{1} A_{it}^{1-\alpha} k_{it}^{\alpha} di \quad ; \quad \tilde{w}_t e^{\kappa t}\right), \qquad (2.4)$$
$$p_{it} = \alpha E_t^{1-\alpha} A_{it}^{1-\alpha} k_{it}^{\alpha-1}.$$

[9] As time is continous, this wage would be the equilibrium wage achieved by the economy in the infinitesimally close period to a moment when labor demand falls. It could be called the wage achieved *around* the moment t of a shock on labor marginal productivity.

From the pre-shock wage $\bar{w}_t$ wages decrease at a rate $\frac{\dot{w}_t}{w_t} = \kappa$ as long as the wage rigidities are binding. In this sense, labor demand equals the maximum between the marginal product of labor and the decreasing wage around a shock.

Employment in the labor market is given by:[10]

$$E_t = \begin{cases} L & \frac{\dot{w}_t}{w_t} > \kappa \\ \left[\frac{(1-\alpha) \int_{\phi_t}^{1} A_{it}^{1-\alpha} k_{it}^{\alpha} di}{\bar{w}_t e^{\kappa t}} \right]^{\frac{1}{\alpha}} & \frac{\dot{w}_t}{w_t} = \kappa, \end{cases} \tag{2.5}$$

that defines the rate of unemployment $u_t = 1 - \frac{E_t}{L}$. Note that when wage rigidities are not binding, the labor market equilibrates to full employment $u_t = 0$. Then, sticky wages can create unemployment because of a shock in labor marginal productivity, for instance from a change in firms' destruction ϕ_t.

A sufficiently high fall in labor's marginal product shifts the equilibrium wage in the absence of rigidities below the level set by sticky wages.

Therefore, a sufficiently high fall in the marginal product of labor around the period t due to a change in firms' destruction ϕ_t, creates unemployment $u_t > 0$ and fix wages wt at the corner solution $\bar{w}_t e^{\kappa t}$. If the fall in the marginal product of labor is not large enough, wages will adjust such that wage rigidities are not binding $\frac{\dot{w}_t}{w_t} > \kappa$. In that case, employment equilibrates at the interior solution with full employment $E_t = L$.

***Lemma* 2.1** For a $\dot{\phi}_t > 0$ sufficiently high, such that $\frac{\dot{w}_t}{w_t} = \kappa$, then $u_t > 0$. If $\dot{\phi}_t \leq 0$ or $\dot{\phi}_t > 0$ is not high enough, such that $\frac{\dot{w}_t}{w_t} > \kappa$, then $E_t = L$.

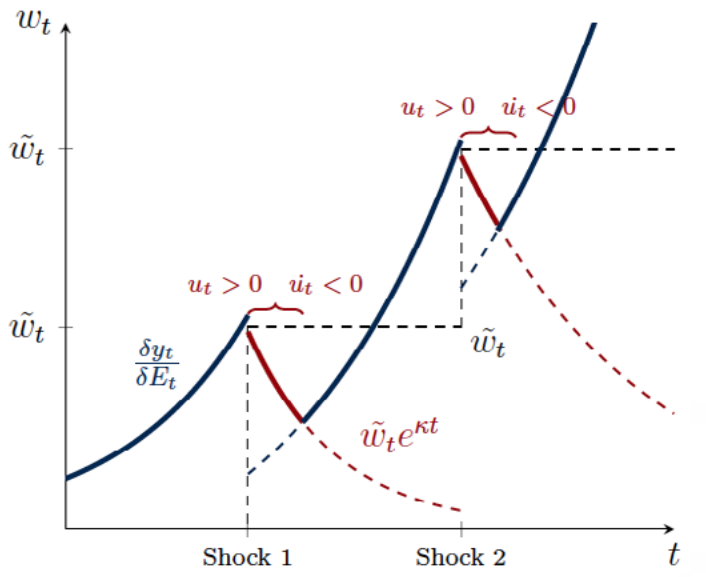
Proof. See appendix A.3.1.

To exemplify Lemma 2.1, assume an exogenous shock on destroyed firms ϕ_t that decreases labor marginal product.[11] A sufficiently high fall in labor's marginal product shifts the equilibrium wage in the absence of rigidities below the level set by sticky wages. If this is the case, the wages adjustment is not strong enough to equilibrate labor demand and supply, so employment decreases with the shock. If the fall is not sufficiently large, wages' adjustment is enough to keep the labor market at full employment.

[10] Employment is $E_t = (1 - u_t)L$. Both expressions for employment are used indistinctly.
[11] This shock would be endogenous to the recession and monetary policy after introducing horizontal growth for firms' creation and destruction.

Figure 1.

Labor Demand with Sticky Wages



Note: The Figure shows the behavior of labor demand. The functions are built for a given set of parameters in equations 2.4 and 2.5. The thick plot is labor demand and wages, as shown in equation 2.4. Note that $\frac{\delta y_t}{\delta E_t}$ represents the interior solution of the final good firm, while $\tilde{w}_t e^{\kappa t}$ is the corner solution. Recall that $\tilde{w}_t$ is the equilibrium wage achieved by the economy in the infinitesimally close period to a moment when labor demand falls. The Figure plots two exogenous shocks on labor marginal product to show the expected behavior of labor demand. This shocks can be interpreted as exogenous shocks on ϕ_t that decreases the firms mass.

Consider the situation depicted in Figure 1 to illustrate the dynamics of labor demand in a constraint scenario. Note that it is possible to observe multiple negative shocks across time. Also, note that $\tilde{w}_t$ changes each time a shock arises but is constant between periods with shocks. As described by Figure 1, a shock to firms' mass decreases the labor marginal productivity below the pre-shock wage level. Here sticky wages become binding and labor demand equals a decreasing exponential function at the rate κ. In this scenario, unemployment rises discontinuously, with a magnitude depending on the gap between labor demand and the labor marginal product. Eventually, the marginal product of labor recovers and exceeds the pre-shock sticky wage. Then, after rising, unemployment decreases until it disappears when labor demand returns to the marginal product of labor, and wage rigidities are not binding anymore. Thus, the labor market equilibrates again to full employment. Note that if the slope

of the sticky wages labor demand is flat $\varkappa = 0$, unemployment lasts longer and starts from a higher level.[12]

In this sense, the magnitude of the unemployment costs described in lemma 2.1 depends on the degree of stickiness $\varkappa$. The closer $\varkappa$ is to perfectly downward rigid wages $\varkappa = 0$, the higher and longer are the unemployment costs from a change in marginal productivity of labor.[13] Finally, the second first order condition in equation 2.4 sets the demand for each variety i of capital goods, which depends on its specific quality and the level of complementarity between varieties. This optimal conditions set the demand for factors of production of the final good firm.

2.2.2 Intermediate Capital Goods Firms

The intermediate capital goods firms produce in a monopolistic framework, thanks to market power received from property rights over a specific quality. Each intermediate firm i supplies a unique variety of capital goods k_{it} with a specific productivity A_{it}. The uniqueness of each capital good allows each firm to produce as a monopoly, facing the total demand for an individual variety of the capital good. The monopolists solve:

$$\max_{k_{it}} \hat{\Pi}_{it} = p_{it}k_{it} - k_{it} \quad s.t. \quad p_{it} = \alpha E_t^{1-\alpha} A_{it}^{1-\alpha} k_{it}^{\alpha-1}, \tag{2.6}$$

[12] Note that unemployment takes more time to fade away with a value of $\varkappa$ closer to 0 as the curves of marginal product and sticky wages take longer to cross. Also note that $\exists \phi^*_t$ s.t. $E_t = L$, where $\phi^*_t \in [0, 1]$.

[13] See Appendix A.2.1 for a discrete time example of how wage rigidities operate.

where the choice of each firm is constrained by the complete demand for capital goods of variety i from the final good firm (as in equation 2.4). Total value of the firm *i* production comes from the amount of capital goods of variety *i* sold k_{it} and the monopolist price paid by the final good firm p_{it}. The cost of production is assumed to be linear on final output with a marginal cost of 1, so each unit of capital goods produced imply an expenditure of the same size in final good quantities. The optimal decision of the monopolist leads to:

$$\begin{aligned} p_{it} &= \frac{1}{\alpha}, \\ k_{it} &= \alpha^{\frac{2}{1-\alpha}}(1-u_t)LA_{it}, \\ \hat{\Pi}_{it} &- (1-\alpha)\alpha^{\frac{1\,|\,\alpha}{1-\alpha}}(1-u_t)LA_{it}. \end{aligned} \tag{2.7}$$

Note that, as the optimization problem is homogeneous, all the capital goods varieties have the same price. The higher is the substitutability between capital goods α, the lower is the mark-up over the marginal cost and the lower is the paid price p_{it}. Offered capital goods kit vary depending on the specific productivity and the employed labor force, due to complementarity between factors of production.

The optimal supply of capital goods k_{it} implies a partial equilibrium in final good production. The final good firm consumes all the offered capital goods and combines them with the employed fraction of workers $(1-u_t)L$ to produce. Given that each capital good has its own level of productivity -an specific quality- final production would depend on average productivity.

Definition **2.1** Define average productivity among operating firms $1-\phi_t$ as a weighted average along the individual productivities A_{it} of each variety *i* of capital goods

$$A_t\,(1-\phi_t) - \int_{\phi_t}^{1} A_{it}di \tag{2.8}$$

where A_t is total factor productivity and $1-\phi_t$ is the size of the firms' mass.

Consequently, the interaction between choices of the monopoly in equation 2.7 and the production function of the final good firm problem in equation 2.3, using Definition 2.1, designates that the final good production function is:[14]

$$y_t = \alpha^{\frac{2\alpha}{1-\alpha}}(1-u_t)L(1-\phi_t)A_t. \tag{2.9}$$

[14] Note that the economy GDP is defined from the resources constraint $y_t = Nc_t + \int_{\phi_t}^{1} k_{it}di + \int_{\phi_t}^{1} R_{it}di$, such that $GDP_t = y_t - \int_{\phi_t}^{1} k_{it}di$. So $GDP_t = \alpha^{\frac{2\alpha}{1-\alpha}}(1-\alpha^2)(1-u_t)L(1-\phi_t)A_t$. Henceforth, as final good production yt and GDP_t dynamics are proportional, the analysis focuses on final production level and growth with no loss of generality.

In the partial equilibrium between the consumption good firm and the intermediate capital goods firms, final production depends on unemployment u_t, the size of surviving firms $1-\phi_t$, total factor productivity A_t and a set of parameters that capture the level of complementarity between capital varieties. This leads to a growth rate of final production:

$$\frac{\dot{y}_t}{y_t} = \frac{\dot{A}_t}{A_t} - \frac{\dot{\phi}_t}{1-\phi_t} - \frac{\dot{u}_t}{1-u_t}, \tag{2.10}$$

so that total output grows $\frac{\dot{y}_t}{y_t} > 0$ with productivity growth $\frac{\dot{A}_t}{A_t}$, and decreases with firms destruction $\frac{\dot{\phi}_t}{1-\phi_t}$ and unemployment expansions $\frac{\dot{u}_t}{1-u_t}$. The growth engine, as in Aghion and Howitt (1990) and Romer (1990), comes from innovations produced in the ideas sector, that allows the capital goods firms to offer a certain quality level.

A capital goods firm then has to pay for a patent that grants property rights over the offered quality. The patent is purchased from the ideas firm at a cost $\hat{P}_{it}$ and sets the productivity A_{it} of a capital goods variety. Each capital goods firm spends the total amount of its benefits $\hat{\Pi}_{it}$ in this purchase, as otherwise another firm could offer a higher payment for the patent and obtain the property rights over that specific quality. The cost of patents is given by:

$$\hat{P}_{it} = \hat{\Pi}_{it} = (1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L\Lambda_{it}. \tag{2.11}$$

2.2.3 Ideas Firms and Credit Market Firms

Innovation and productivity growth come from investment in R&D made by firms in the ideas sector that operate in perfect competition. These firms create a quality of capital goods i with specific productivity A_{it}. With this purpose, the firm uses resources R_{it} obtained from a credit of cost C_{it} offered by the savings market, which lending funds come from the household assets a_t and resources deployed by the monetary authority M_t. Credit is managed by a private bank that works as an intermediary between available funds and the demand for loans in a competitive market. The benefits function of the private bank is:

$$\bar{\Pi}_{it} = \mathcal{C}_{it} - r_t R_{it}, \tag{2.12}$$

where r_t is the rewarded interest rate on lending funds. Due to perfect competition in the credit market, the benefits of the private bank are $\bar{\Pi}_{it} = 0$ and the cost of credit is given by $C_{it} = r_t R_{it}$.

The private bank charges a marginal cost equal to the real interest rate r_t for the quantity of borrowed resources.[15] With this, the loanable funds market clears. Hence, the private bank is not an agent but works as a market clearing condition for credit supply and demand. This condition constraints the production process of the ideas firms, as they must pay the credit in order to obtain resources.

In this scenario, the representative ideas firm chooses how much to invest in in-novation, according to certain success probability μ_{it}, with the purpose of acquiring a patent. The investment costs are given by expenditure in R&D R_{it} and the cost of credit C_{it}. The ideas firms solve the problem:

$$\begin{aligned} \max_{R_{it}} \tilde{\Pi}_{it} = \mu_{it}\hat{P}_{it} - R_{it} - \mathcal{C}_{it} \quad s.t. \quad & \mu_{it} = \lambda \left(\frac{R_{it}}{A^*}\right)^{\sigma}, \\ & \hat{P}_{il} = (1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_l)LA_{it}, \\ & \mathcal{C}_{it} = r_t R_{it}. \end{aligned} \tag{2.13}$$

Each ideas firm maximizes its profits $\tilde{\Pi}_{it}$, choosing how much resources R_{it} to invest in innovation. Given the success probability of innovation μ_{it}, if a firm innovates and increases the quality of a variety i of capital goods, it receives the price of the patent $\hat{P}_{it}$ that equals the monopolistic benefits of a firm in the intermediate sector. The probability of being successful is standard, where greater investment increases the likelihood of innovation, but targeting a high level of goal productivity A^* decreases the chances of succeeding. In the probability function, σ is interpreted as the elasticity of investment on innovation and λ captures the productivity of resources expenditure.

First order conditions of this problem set the optimal investment in R&D:

$$R_{it} = \left[\frac{\lambda\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L}{1+r_t}\right]^{\frac{1}{1-\sigma}} A_{it}. \tag{2.14}$$

Investment is increasing in the resulting productivity from the innovation process A_{it} and in the level of employment E_t, as this affects the price at which that patent is sold. Note that investment in innovation is decreasing in the interest rate r_t, because it affects the marginal cost of innovation due to the credit cost. Besides, the relationship between R&D expenditure and the interest rate depends in the elasticity of investment on innovation σ, the the productivity of resources λ, and the terms collected by the patents price. This amount of investment is consistent with an ideas success probability that sets innovation effort:[16]

$$\mu_t = \lambda^{\frac{1}{1-\sigma}} \left[\frac{\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L}{1+r_t}\right]^{\frac{\sigma}{1-\sigma}}. \tag{2.15}$$

Note that innovation effort is also increasing in employment and decreasing on the interest rate. Then, both investment in R&D R_{it} and innovation effort μ_t depend negatively on the interest rate of the credit market r_t. These results from equations 2.14 and 2.15 are summed up in Lemma 2.2.

[15] Also note that all the resources are obtained from the private bank's credit.

[16] The decentralized solution for innovation effort in equation 2.15 is suboptimal from the social planner perspective under certain conditions. Appendix A.2.2 presents both the planner's problem and solution, and the conditions on the interest rate that lead to suboptimal innovation.

***Lemma* 2.2** The innovation outcomes R_{it} and μ_t hold $\frac{\partial R_{it}}{\partial r_t} < 0$ and $\frac{\partial \mu_t}{\partial r_t} < 0$.

Proof. It is straightforward from equations 2.14 and 2.15. See appendix A.3.2 for details.

Note that, as the marginal cost of investment in the ideas sector is increasing in the interest rate r_t, a change in the supply of credit that raises the interest rate charged for resources r_t decreases expenditure in R&D R_{it} and innovation frequency μ_t. Therefore, a rise in the interest rate hurts innovation outcomes and average productivity.

2.3 Vertical and Horizontal Growth

The firms setting presented in the previous section shows a structure that allows average productivity to expand both in the intensive and the extensive margin. As the creation of ideas improves the quality of capital goods, productivity grows in the intensive margin with the success probability of innovation. On the other hand, the firms' mass dynamics also depends on the innovation effort, setting the amounts of capital goods varieties that contribute to growth in the extensive margin. In this sense, the model includes both vertical growth from individual productivity upgrades and horizontal growth from the creation of new capital varieties.[17]

This strategy is appealing from a theoretical perspective as it allows to test both growth mechanisms, from Aghion and Howitt (1990) and Romer (1990), in the presence of shocks and in the BGP. Furthermore, evidence on the scarring effects of recessions highlights the relevance of firms' net exit as a central mechanism for hysteresis (Ouyang, 2009; Eslava et al., 2010). For this reason, including horizontal growth is helpful as it accounts for net destruction of firms, in contrast to the permanent entry and exit of the typical Schumpeterian approach (Aghion et al., 2014b).

In the intensive margin, vertical growth comes from upgrades in individual productivity. The investment decision that determines innovation effort from the ideas firms defines the productivity growth rate in the economy.

[17] Some examples of models that also allow both types of growth are Young (1998) and Aghion et al.(2019). In general, horizontal growth exhausts in the long-run and only vertical expansions prevail, a prediction also true in this model. This paper's primary difference comes from the reasoning of varieties as firms, where firms' exit separates production from its efficient level. See Aghion and Howitt (2008) for more details.and the conditions on the interest rate that lead to suboptimal innovation.

***Definition* 2.2** Innovation creates productivity upgrades with certain probability μ_t, so the law of motion for each variety specific productivity A_{it} is

$$\dot{A}_{it} = \begin{cases} \gamma A_t & \mathbb{P}(\mu_t) \\ 0 & \mathbb{P}(1-\mu_t), \end{cases} \tag{2.16}$$

where A_t is total factor productivity in the economy, γ is the size of the jump in productivity thanks to the innovation, and μ_t is the innovation success probability. A fraction μ_t of the firms' mass $1 - \phi_t$ innovates and reaches a higher productivity at period t, while the individual productivity of firms in the fraction $1 - \mu_t$ equals the current total factor productivity. The expected productivity growth is given by

$$\mathbb{E}_t \left[\frac{\dot{A}_t}{A_t} \right] = \mu_t \gamma, \tag{2.17}$$

So productivity grows depending on the frequency of innovations μ_t and on the size of the upgrades in quality γ.

Then, the success probability of ideas firms increases productivity. As a fraction of firms innovate and increase their individual productivity, average productivity increases in the long-run.[18] This improvement in the productivity of each capital good variety is the source of vertical growth in production.

Productivity gains of capital goods varieties have spillover effects over all the ideas sector, not only on the firms that manage to innovate. As the productivity of the firms that do not innovate equals total productivity, innovation effort benefits the whole economy when average productivity increases thanks to successful firms. Similarly, newborn firms acquire the average productivity of existing varieties, so they also enjoy the innovation upgrades. These simplifying assumptions discard the possibility of increasing heterogeneity between more and less successful varieties in the production of ideas.

In the extensive margin, horizontal growth comes from expansions in the firms' mass. As the amount of capital goods depends on the quantity of available patents, the innovation process creates learning externalities on firms and varieties creation. The assumption is that inventing an idea creates externalities due to a learning process during the patent transaction. Externalities foster that the interaction with the innovation sector of a capital goods firm improves the surviving probability of the firms in the market. Thus, more investment in R&D on existing varieties of capital goods generates useful knowledge for the production of new varieties and the survival of firms.

[18] This is true due to the law of large numbers. The continuum that defines how many firms are attempting to innovate is large enough to increase average productivity, calculated as a weighted average of individual productivities among surviving firms.

***Definition** 2.3* The law of motion for firms destruction $\dot{\phi}_t$ is defined from learning externalities of ideas creation to firms survival:

$$\dot{\phi}_t = \begin{cases} (\theta - \Omega\mu_t)(1 - \phi_t) & 0 < \phi_t \leq 1 \\ \max(\theta - \Omega\mu_t, 0) & \phi_t = 0, \end{cases} \tag{2.18}$$

where θ is an exogenous natural rate of destruction, Ω is the degree to which learning externalities are translated to firms survival, and μ_t is optimal innovation effort from ideas firms. This defines the level and change of the lower bound of firms mass ϕ_t, so $\dot{\phi}_t > 0$ is net destruction of firms and $\dot{\phi}_t < 0$ is net creation.

Then, the higher the frequency of innovation μt, the easier firms survive in the market $\dot{\phi}_t < 0$. This implies that new ideas increase the quality of capital goods and improve the survival probability of intermediate firms. Note that there exists a $\hat{\mu}_t = \frac{\theta}{\Omega}$ where the firms' mass stops moving $\dot{\phi}_t = 0$ for all values of ϕ_t. Also note that stopping firms dynamics $\dot{\phi}_t = 0$ when there are no destroyed firms $\phi_t = 0$ implies maximum firms survival. Horizontal growth then arrives at its maximum installed capacity. Under Lemma 2.1, all the firms are producing, there is no more destruction of firms, and the economy operates at full employment.[19] On the other side, note that a complete destruction of firms $\phi t = 1$ implies no entry of firms $\dot{\phi} t = 0$ and absence of capital goods for production, so the economy dies.

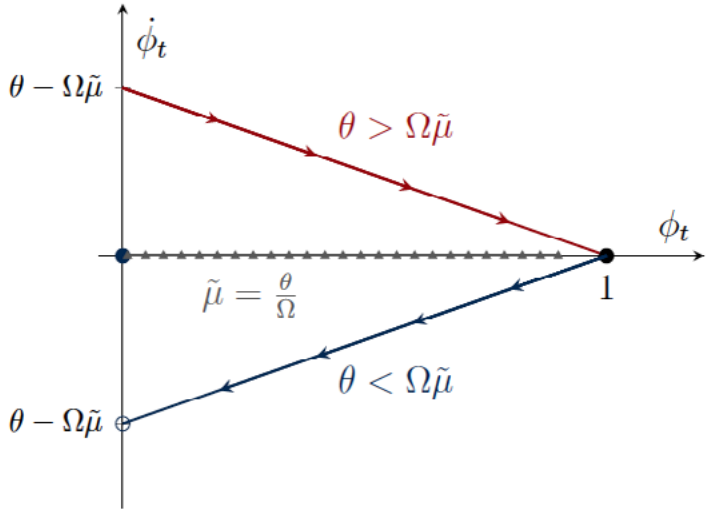
The dynamics given by equation 2.18 are represented in Figure 2. The Figure shows the lower bound of firms' mass ϕ_t against its change $\dot{\phi}_t$ for three possible scenarios that define horizontal growth, assuming a constant rate of innovation effort $\tilde{\mu}$. Note that the dynamics of firms destruction depend on the relationship between how easily innovation effort translates into creation $\Omega\tilde{\mu}$ and the natural rate of firms' exit θ. If innovation effort is not sufficiently large, the natural rate of destruction overcomes learning externalities $\theta > \Omega\tilde{\mu}$, and eventually all intermediate firms in the economy die. On the other hand, if learning externalities for a given innovation effort dominate $\theta < \Omega\tilde{\mu}$, the firms' mass grows, and the

[19] Unemployment converges to 0 in the absence of firms' destruction that affects the marginal product of labor. This is the usual case in Aghion and Howitt (1990): full firms' survival and full employment.

economy converges to a scenario with full firms' survival. In this case, the quantity of destroyed firms is $\dot{\phi}_t = 0$ and destruction stops $\dot{\phi}_t = 0$. All the possible firms for different varieties of capital goods are operating, and horizontal growth is depleted.

Figure 2.

Firms' Destruction Convergence Based on Innovation Effort



Note: The Figure shows firms' mass dynamics, of equation 2.18, possible scenarios for a given innovation effort $\tilde{\mu}$. The Figure plots three cases. First, a case when innovation effort is not large enough to exceed natural destruction of firms given by $\theta > \Omega\,\tilde{\mu}$. Second, a case when innovation effort is high enough, so that $\theta < \Omega\,\mu$. Finally, the parametric condition on innovation effort $\tilde{\mu} = \frac{\theta}{\Omega}$ that makes destroyed firms ϕ_t cease to increase or decrease. Arrows signal the direction of the dynamics of ϕ_t from an initial point to one of its bounds $\phi_t \in [0, 1]$.

Note that eventually the economy reaches a point where the expansion or reduction of capital goods varieties stops. If learning externalities are strong enough for a sufficiently large innovation effort to hold $\Omega\mu_t > \theta$, in the long-run all firms survive $\phi_t = 0$. In this situation, firms' destruction $\dot{\phi}_t$ decreases with destroyed firms ϕ_t until it converges to the maximum level of firms survival $\phi_t = 0$. This result is summarized in Lemma 2.3.[20]

***Lemma* 2.3** For μ_t sufficiently high such that $\Omega\mu t > \theta$, then $\phi_t = 0$. and $\dot{\phi}_t = 0$ in the long-run.

Proof. It is straightforward from Definition 2.3. See appendix A.3.3 for details.

[20] Note that a fall of μ_t would de-accelerate the speed at which the economy arrives to full firms survival. However, as long as the condition of Lemma 2.3 holds, eventually all varieties of capital goods would operate in a firms' mass of size 1.

The economy reaches full firms survival $\phi_t = 0$ and $\dot{\phi}_t = 0$ in the long-term if innovation effort μ_t is high enough to fulfil the necessary and sufficient condition $\Omega\mu_t > \theta$.

The notion that a high level of research within a sector is related to lower exit probability of firms inspires these assumptions on horizontal growth. The literature on firms' dynamics has widely studied the rate of firms' exit θ and its underlying roots (Agarwal and Gort, 2002; Clementi and Palazzo, 2016; Acemoglu et al., 2018; Marcus, 1967). The learning externalities Ω assumption is supported by results that show that firms with higher R&D expenditure have a lower exit probability (Madrid et al., 2015; Cefis and Marsili, 2006; Akcigit et al., 2016; Jung et al., 2018). Then, Definition 2.3 highlights that more research is related to a higher chance of survival in the market.

Thanks to endogenous productivity growth and learning externalities from innovation on firms' creation, more investment in R&D generates vertical improvements in the quality of the capital goods and new varieties from a horizontal expansion of the firms' mass. As in Definition 2.1, average productivity would depend on how many firms survive $1 - \phi_t$ and on total productivity A_t, both ruled by the success probability of ideas μ_t.

Finally, from the result of optimal innovation effort of equation 2.15 and the definitions for vertical and horizontal growth in equations 2.17 and 2.18, the model suggests bliss points for growth gains depending on the interest rate.[21] Destroyed firms have a minimum in $\phi_t = 0$ that is stable when firms' dynamics stop $\dot{\phi}_t = 0$. Here, all the firms survive, and destruction ends due to learning externalities. Besides, innovation effort can reach a maximum at $\mu_t = 1$, where all firms that try to innovate manage to create an idea and R&D's success probability within the ideas sector becomes 1. Therefore, two points on the interest rate define the limits of growth gains: the full firms' survival interest rate r_t^{ϕ} and the full innovation interest rate r_t^{μ}.

These bliss points delimit the space of influence of the interest rate over growth engines. Depending on their location relative to the interest rate consistent with the BGP, the natural real interest rate r^*, vertical and horizontal growth operate at a certain depth in the long-run. The full firms' survival interest rate consistent with $\dot{\phi}_t = 0$. and $\phi_t = 0$ is:

[21] Note that ϕ_t depends on μ_t that depends in r_t, so growth depends directly in the interest rate.

$$r_t^{\phi} = \left(\frac{\Omega}{\theta}\right)^{\frac{1-\sigma}{\sigma}} \lambda^{\frac{1}{\sigma}} \sigma\,(1-\alpha)\,\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L - 1. \qquad (2.19)$$

Any interest rate below r_t^{ϕ} does not create benefits in terms of firms survival, so horizontal growth is exhausted after this point. Moreover, from Lemma 2.3 it is clear that under the correct parametric conditions and a desirable innovation effort, in the BGP the natural interest rate equals the full firms' survival interest rate $r_t^{\phi} = r^*$, as all firms survive in the long-run.

The full innovation interest rate consistent with $\mu_t = 1$ is:

$$r_t^{\mu} = \lambda^{\frac{1}{\sigma}} \sigma\,(1-\alpha)\,\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L - 1, \qquad (2.20)$$

so an interest rate beneath r_t^{μ} does not propagate any productivity gains, as all firms that try to innovate succeed and vertical growth is determined by the jump in productivity γ. Nevertheless, the creative destruction mechanism that feeds endogenous vertical growth must hold in the long-run. Thus, the innovation process always operates in an uncertain scenario $\mu_t < 1$, and a fractions of firms do not innovate. Consequently, the long-run natural interest rate must be greater than the full innovation interest rate $r^* > r_t^{\mu}$.

Therefore, in the BGP equilibrium the full firms' survival interest rate must be higher than the full innovation interest rate $r_t^{\phi} > r_t^{\mu}$. Note that if the firms natural rate of destruction is lower than the learning externalities pass-through $\Omega > \theta$, then it holds that $r_t^{\phi} > r_t^{\mu}$. This parametric condition is a reasonable assumption in terms of firms' dynamics: it ensures that in the long-run, all firms survive while the innovation process remains uncertain. Moreover, it admits the BGP equilibrium relations $r_t^{\phi} = r^*$ and $r_t^{\mu} < r^*$ in reference to the long-run natural real interest rate r^*.

***Lemma* 2.4** $\exists r_t^{\phi}$ such that $\phi_t = 0$ and $\dot{\phi}_t = 0$, and $\exists r_t^{\mu}$ such that $\mu_t = 1$. If $\Omega > \theta$, then $r_t^{\phi} > r_t^{\mu}$, and in the long-run $r_t^{\phi} = r^*$ and $r_t^{\mu} < r^*$.

Proof. See appendix A.3.4.

2.4 Central Bank and Monetary Policy

The central bank decides monetary supply emission $\dot{M}_t$ to minimize a loss function $\mathcal{L}_t$ that represents its preferences over monetary policy objectives. A policy regime that sets the goals of central banking rules its behavior. Optimal monetary policy is defined as a rule that motivates the central bank to react throughout recessions and avoid long-run costs in production. In contrast, a monetary policy mistake arises from a short-sighted central bank, in which preferences the only policy goal is inflation. In this scenario, monetary policy is not sufficiently expansionary during recessions, which rises the interest rate too much, creating costs over innovation and permanent effects on the long-run trend.

To achieve its policy objectives, the central bank operates along the credit channel, as described by Bernanke and Gertler (1995) and Mishkin (1996).[22] When the economy receives a shock, the central bank changes the available resources in the savings market deploying monetary supplies. Monetary emission then changes the interest rate, allowing the central bank to hit both inflation and growth with monetary policy.

2.4.1 Inflation, The Credit Channel and The Power of Monetary Policy

Inspired in Aghion et al. (2019), inflation in the model is calculated as the residual from monetary emission of real output growth. Consider a simple quantitative equation of money with constant velocity of monetary aggregates. The identity implies that the growth rate of monetary supply $\frac{\dot{M}_t}{M_t}$ is distributed between inflation π_t and output growth $\frac{\dot{y}_t}{y_t}$ Then, residual inflation is given by:

$$\pi_t = \frac{\dot{M}_t}{M_t} - \frac{\dot{y}_t}{y_t}. \qquad (2.21)$$

Note that inflation is interpreted as excessive monetary emission relative to production growth. As the production side does not have prices that affect agents' decisions, inflation springs only because of a disproportionate monetary mass creation. However, introducing residual inflation creates a cost for the

[22] In detail, monetary policy would work in the bank lending transmission mechanism. However, the reasoning focuses on the effect of the supply of bank loans on the general level of the interest rate, rather than on the size of the external finance premium.

central bank of undershooting the interest rate to shallow levels. As an unreasonably high monetary expansion would produce positive inflation, there are no incentives to set interest rates too low.[23]

Monetary policy operates along the credit channel. The central bank decides monetary mass emission to the savings market loanable funds. Thus, the savings market equilibrium constrains the central bank's power. The market clearing condition on the savings market is given by the equilibrium between supply and demand of credit:

$$\int_{\phi_t}^{1} R_{it} di = N a_t + M_t. \qquad (2.22)$$

The demand for credit is given by the ideas firms that require resources for production. All the available funds are distributed along the amount of ideas firms, that equals the mass of invented varieties of capital goods $1 - \phi_t$, i.e. the demand for credit is total expenditure on R&D. The supply of credit, as explained earlier, comes from available lending funds composed by assets a_t that the N households deposit as loanable funds, and by monetary aggregates M_t deployed by the central bank. Remark that monetary emission is viewed as a gift of resources from the central bank to the savings market.[24]

Note that the price that clears the equilibrium of the savings market given by equation 2.22 is the real interest rate r_t. Using equation 2.14 for investment in R&D R_{it}, the interest rate charged by the private banks that clears the market is:

$$r_t = \left(\lambda\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L\right)\left[\frac{(1-\phi_t)A_t}{Na_t + M_t}\right]^{1-\sigma} - 1. \qquad (2.23)$$

The central bank changes credit supply using emission $\dot{M}_t$ to control the real interest rate r_t that clears the savings market. Then, a positive monetary emission $\dot{M}_t > 0$ increases credit supply through monetary aggregates M_t and decreases the interest rate r_t.[25] Also note that, as recessions $\varepsilon_t < 0$ affect household savings and reduce assets creation $\dot{a}_t$, a negative shock reduces assets deposits at in the savings market. This leads to a decrease in credit supply and an increase in the cost of credit. In this scenario, the central bank's intervention can rebalance the credit supply and prevent a rise in the interest rate.

Furthermore, the central bank aims to close a gap between observed growth $\frac{\dot{y}_t}{y_t}$ and potential output growth $\frac{\dot{\bar{y}}_t}{\bar{y}_t}$, essentially after shocks. Equation 2.9 suggests that production changes because of *i*. firms' exit, *ii*. unemployment, and *iii*. productivity deviations. Based on Lemma 2.3, in the BGP, the firms' mass arrives to full firms survival. Likewise, from Lemma 2.1, even though the unemployment rate is positive $u_t > 0$ during recessions, the labor market operates in full employment $u_t = 0$ during regular times and booms, where wage rigidities are not binding.[26] Thus, equation 2.9 implies that potential output equals:

---

[23] This is the easiest way of introducing inflation as a cost of undershooting the interest rate. The quantitative equation of money $M_t \bar{V} = P_t y_t$ sustains the residual inflation assumption. It is possible to defend an equivalent argument introducing the price of the consumption good Pt and calculating inflation from nominal output $M_t = P_t y_t$.

[24] Helicopter drops to household income is another strategy for monetary policy transmission. However, this alternative also demands a propagation channel through the savings market. Thus, the savings market remains essential, and the credit channel is the most direct approach for monetary policy transmission.

[25] Note that $\frac{\partial r_t}{\partial \dot{M}_t} < 0$ and the elasticity of the interest rate to monetary supply is negative

[26] Recall from Figure 1 that unemployment is positive but decreasing after a sharp rise caused by the shock, so the labor market will always return to full employment.

$$\tilde{y}_t = \alpha^{\frac{2\alpha}{1-\alpha}} L A_t. \tag{2.24}$$

Potential output $\tilde{y}_t$ depends on total factor productivity A_t, labor supply L, and the same set of parameters that capture intermediate firms' mark-up. Unemployment and firms' destruction do not play a role in setting the efficient level of production. Hence, potential output grows at the same rate as productivity growth:

$$\frac{\dot{\tilde{y}}_t}{\tilde{y}_t} = \frac{\dot{A}_t}{A_t}. \tag{2.25}$$

In usual times, when the economy is at the BGP, the growth gap is 0. However, it widens to negative values during recessions, triggering unemployment and firms' destruction.[27] Booms, on the other side, heat the economy by boosting the path towards full employment and full firms' survival, and by increasing average productivity with more innovation. At last, from equations 2.9 and 2.24, it is clear that both observed and potential output depend on productivity A_t. Recessions and monetary policy mistakes affect both observed and potential output growth by hitting productivity growth.

Therefore, the supply side equilibrium will also constrain the central bank's range of action. In particular, in line with Lemma 2.2 the effect of monetary emission in the real interest rate connects monetary policy with the innovation outcomes R_{it} and μ_t. Since both vertical and horizontal growth depend on the probability of success in the creation of ideas, as stated in Definitions 2.2 and 2.3, and according to Lemma 2.1 employment only changes because of shocks on labor demand through the firms' mass $1 - \phi_t$, under equation 2.10 growth $\frac{\dot{y}_t}{y_t}$ changes with the central bank's decisions. Monetary policy has an impact on innovation effort μ_t, which governs growth. Hence, the interaction between the credit market and the supply side outlines the monetary policy power over growth.[28]

[27] Even though firms' exit and unemployment don't affect growth in the BGP, they certainly play a role in defining the macroeconomic stabilization during crisis, as the growth gap depends strictly in firms' exit and unemployment $\left(\frac{\dot{y}_t}{y_t} - \frac{\dot{\tilde{y}}_t}{\tilde{y}_t}\right) = -\frac{\dot{\phi}_t}{1-\phi_t} - \frac{\dot{u}_t}{1-u_t}$.

[28] This means that the model comprises an implicit Phillips Curve that relates monetary aggregates with growth. This feature of the model answer directly to Lucas Jr (1972), as it allows to relate monetary actions with supply -thanks to the credit market- without forcing a correlation between inflation and output. This result prevails as far as monetary actions answer strictly to shocks.

2.4.2 The Central Bank

The central bank reacts by changing monetary aggregates M_t with emission $\dot{M}_t$ to minimize a loss function $\mathcal{L}_t$. Although the central bank has the responsibility of accomplishing its targets, it does not have the authority to define the policy goals on its preferences. On the contrary, its design is set outside of the economy by institutional and legal means. Thus, the framework that mandates monetary actions may be inconsistent with the economy's state and needs, which opens space to monetary policy mistakes. In this sense, the analysis of monetary policy decisions would contrast two schemes of preferences that drive different policy platforms.

Historically, output and inflation are the usual monetary policy goals, justified by evidence based on Taylor (1993). Hence, the central banking designs would focus on efforts to hit growth and inflation. Under a certain policy scheme, monetary policy relies on the credit channel to use emission as its main instrument. Moreover, the central bank's decisions are subject to the residual inflation identity, the savings market equilibrium, the labor market equilibrium, and the supply side partial equilibrium.

The central bank decides after observing the shock in household income εt, so it acknowledges its effect on savings and assets creation. Thus, the central bank observes that shocks deviate the economy from its initial position at the BGP. The analysis centers on negative shocks $\varepsilon_t < 0$, to capture recessions as the source of hysteresis. In response to shocks, the central banking tool is a monetary policy platform for emission $\dot{M}_t$ to accomplish the objectives in a specific loss function $\mathcal{L}_t$.

Definition 2.4 A monetary policy platform $\dot{M}'_t$ for a loss function that represents the preferences of the central bank $\mathcal{L}'_t$ is

$$\dot{M}'_t = \operatorname{argmin}_{\dot{M}_t} \mathcal{L}'_t \qquad (2.26)$$

where $\mathcal{L}'_t \geq 0$ holds. So, from $\min(\mathcal{L}'_t) = 0$ choosing $\dot{M}'_t$ implies $\mathcal{L}'_t = 0$.

From Definition 2.4, it is clear that the chosen policy platform suggests a monetary policy rule specific to the central bank design. Reconvene that changes in central banking objectives depend on the institutional setting, so the policy is constrained by which priorities the central bank is trying to fulfill. Thus, the approach to determine optimal policy and errors is assessing different arguments in the loss function.

The model then suggests criteria to classify as a mistake any deviation of the optimal platform, resulting from a set of preferences inconsistent with the economy's best scenario in terms of guarding the long-run trend.

In the spirit of Barro and Gordon (1983), the central bank's decisions result from solving the policy platform for a loss function $\mathcal{L}_t$. The primary idea of optimal monetary policy is to elude hysteresis. The central bank can achieve this purpose through its effects on growth. The model then suggests criteria to classify as a mistake any deviation of the optimal platform, resulting from a set of preferences inconsistent with the economy's best scenario in terms of guarding the long-run trend. As a result, the definition of optimal monetary policy is a platform that protects the economy from hysteresis.

Definition 2.5 Hysteresis driven optimal monetary policy is a monetary policy platform $\dot{M}^*_t$ for a loss function $\mathcal{L}^*_t = f\left(\frac{\dot{y}_t}{y_t}, \frac{\dot{\tilde{y}}_t}{\tilde{y}_t}, \pi^*_t\right)$ such that

$$\dot{M}^*_t = \operatorname{argmin}_{\dot{M}_t}\left[\left(\frac{\dot{y}_t}{y_t} - \frac{\dot{\tilde{y}}_t}{\tilde{y}_t}\right)^2 + \operatorname{Var}\left(\frac{\dot{\tilde{y}}_t}{\tilde{y}_t}\right) + (\pi^*_t)^2\right], \tag{2.27}$$

where $\frac{\dot{y}_t}{y_t}$ is observed growth, $\frac{\dot{\tilde{y}}_t}{\tilde{y}_t}$ is efficient production growth, and π^*_t is long-run inflation.

The optimal monetary policy implies choosing monetary emission $\dot{M}^*_t$ to minimize the growth gap $\frac{\dot{y}_t}{y_t} = \frac{\dot{\tilde{y}}_t}{\tilde{y}_t}$, maintain potential output growth balanced $\operatorname{Var}\left(\frac{\dot{\tilde{y}}_t}{\tilde{y}_t}\right) = 0$, and keep long-run inflation in a target assumed in $\bar{\pi}_t = 0$. Moreover, in this regime, the central bank weighs all its purpose in the same proportion. The solution of equation 2.27 applying Definition 2.4 conducts to an optimal monetary policy rule:

$$\frac{\dot{M}_t}{M_t}^* = \frac{\dot{y}_t}{y_t} - \frac{N a_t \varepsilon_t}{M_t}. \tag{2.28}$$

The efficient policy rule implies that the central bank must emit sufficient monetary aggregates to counterweight the assets' losses due to recessions $Na_t\varepsilon_t$. This balances credit supply whenever a negative shock destroys loanable funds of the households, holding the savings market and the interest rate in a position compatible with the BGP evolution. The optimal response is to practice expansionary monetary policy during crises. Lemma 2.5 describes the central bank's expected behavior in periods with shocks and its effect on credit supply.

Lemma 2.5 Let $\tilde{t} = \{t : \varepsilon_t \neq 0 \wedge t \in \mathbb{R}^{\geq 0}\}$, $\Gamma(\cdot)$ map the changes in $\tilde{t}$, and CS_t be credit supply. $\Gamma(\dot{M}^*_t) = -Na_t\varepsilon_t$ and $N\Gamma(\dot{a_t}) = Na_t\varepsilon_t$. Thus, $\dot{M}_t = \dot{M}^*_t$ implies $\Gamma(CS_t) = 0$.

Proof. See appendix A.3.5.

Optimal monetary policy perfectly compensates for all the assets losses generated from negative shocks. During bad times, the economy separates from its regular situation at the BGP. In these extraordinary periods, an optimal central bank reacts with extraordinary changes in emission to rebalance credit supply and sterilize the shock propagation. The central bank then manages to keep the economy at the BGP trend even during extraordinary times.[29] Lastly, in the absence of shocks, the central bank emits at a constant rate consistent with production growth, which guarantees that inflation hits the target in the long-run.

Achieving the central bank's objectives in the optimal regime implies that monetary policy evades hysteresis, and restrains the shock's effects. Four results arise from following the optimal monetary policy response $\dot{M}^*_t$. First, the interest rate r_t keeps in the BGP consistent level r^* thanks to the rebalance of credit supply. Second, the growth gap equals 0, so there is no firms' destruction $\phi_t = 0$ and the labor market keeps in full employment $u_t = 0$. Third, the BGP growth rate does not change with the recession, so potential growth $\frac{\dot{y}_t}{y_t}$, productivity growth $\frac{\dot{A}_t}{A_t}$, and innovation effort μ_t are constant. Finally, long-run inflation π^*_t hits the inflation target $\bar{\pi}_t = 0$. Thus, the economy continues growing at the same rate during shocks, and in the long-run inflation hits its target. Lemmas 2.6, 2.7, 2.8, and 2.9 formalize these results.

***Lemma* 2.6** If $\dot{M}_t = \dot{M}^*_t$, then $r_t = r^*$.

Proof. See appendix A.3.6.

***Lemma* 2.7** If $\dot{M}_t = \dot{M}^*_t$, then $\frac{\dot{y}_t}{y_t} = \frac{\dot{\bar{y}}_t}{\bar{y}_t}$. Therefore, $\phi_t = 0$ and $u_t = 0$.

Proof. See appendix A.3.7.

***Lemma* 2.8** If $\dot{M}_t = \dot{M}^*_t$, then $\text{Var}(\frac{\dot{\bar{y}}_t}{\bar{y}_t}) = 0$, $\text{Var}(\frac{\dot{A}_t}{A_t}) = 0$, and $\text{Var}(\mu_t) = 0$.

Proof. See appendix A.3.8.

***Lemma* 2.9** If $\dot{M}_t = \dot{M}^*_t$, then $\pi^*_t = \bar{\pi}_t = 0$.

Proof. See appendix A.3.9.

Moreover, the optimal answer to recessions $\dot{M}^*_t$ demands sacrificing to achieve the inflation target in the short-run, as expansionary monetary policy requires excessive emission. Nevertheless, Lemma 2.9 shows that inflation gets to its target in the long-run. Therefore, the inflation costs are significantly lower than the hysteresis output loss that would arise from a recession in the absence of a vigorous response.

Monetary policy mistakes result from a central banking regime whose objectives devote no attention to crises' potential long-run costs. The central bank's design then embodies a short-sighted central bank as an inflexible regime controls its macroeconomic policy vision, limiting its attention solely to inflation, even during recessions. The myopia of the monetary authority is costly during crises as it causes hysteresis. The risk is that an unyielding policy scheme ties the central bank's hands to defend a purpose that is not the main cause of economic unbalance throughout the adverse shocks.

In particular, the monetary policy error origin is a loss function that depends only on current inflation. Thus, to keep inflation on target, the central bank avoids the excessive emission needed for an expansionary monetary policy devised to

[29] Note that this expected behavior is consistent with a countercyclical Taylor rule, for instance $r^O_t = r* + \psi\varepsilon_t$. See Appendix A.2.3 for more details.

face crises. Consequently, interest rates rise too much in recessions, hurting the long-run trend. Hence, the definition of a monetary policy mistake is a platform that deviates from the optimal decision because of a fixation on inflation targeting.

***Definition* 2.6** An inflation driven monetary policy mistake is a monetary policy platform $\dot{M}^E_t$ for a loss function $\mathcal{L}^E_t = f(\pi_t)$ such that

$$\dot{M}^E_t = \operatorname{argmin}_{\dot{M}_t} (\pi_t)^2 , \tag{2.29}$$

where π_t is observed residual inflation.

Monetary policy mistakes designate choosing monetary emission $\dot{M}^E_t$ with the only purpose of setting and maintaining inflation on target $\bar{\pi}_t = 0$. Using equation 2.21, the solution to the monetary policy mistake regime yields to a monetary policy rule of a short-sighted central bank:

$$\frac{\dot{M}_t}{M_t}^E = \frac{\dot{y}_t}{y_t}. \tag{2.30}$$

The monetary policy authority does not see nor react to recessions. The policy rule states that the central bank adopts an emission equal to production growth to keep inflation stable on target, even during crises, and blocks it from exercising expansionary monetary policy. Moreover, as emission must match production growth, in the absence of the counterweight suggested by the optimal rule, emission falls when recessions depress growth: monetary policy becomes procyclical. In attempting to keep inflation on target, the central bank applies a contractionary monetary policy during crises.

Therefore, following Lemma 2.2, a shock in the presence of the monetary policy mistake described by equation 2.30 produces permanent effects on production because of a lack of expansionary monetary policy that prevents the interest rate from rising. The central bank fails to keep the interest rate constant after a shock because the recession destroys a fraction of credit supply and hurts growth, furthering a monetary contraction. Hence, the fear of creating inflation with excessive emission restrains the central bank from a countercyclical policy. Lemma 2.10 shows this result.

***Lemma* 2.10** If $\dot{M}_t = \dot{M}_t^E$ and $\varepsilon_t < 0$, then $\pi_t = \bar{\pi}_t = 0$ but $\dot{r}_t > 0$ and $\dot{\mu}_t < 0$.

Proof. See appendix A.3.10.

A monetary policy mistake $\dot{M}^E_t$ sets inflation π_t on target $\bar{\pi}_t = 0$ but reacts to negative shocks $\varepsilon_t < 0$ with a contractionary monetary policy. This generates a rise in the interest rate rt that harms innovation effort μ_t and creates productivity costs, originating hysteresis.

Thus, the monetary policy error comes from fear of excessive emission during recessions because of an extreme focus on inflation. Since any deviation from the optimal platform can be considered an error, a monetary policy mistake is a not sufficiently expansionary monetary policy during crises. The regime focused only on inflation makes the deepest possible error: it performs a monetary policy grip during bad times. Instead of mitigating the crises, the central bank intensifies the consequences of the shocks.[30] Any policy reaction more contractionary than this maximum mistake would not make sense in light of the classical objectives of monetary policy. Hence, optimal and error platforms delimit a set of monetary policy errors that are suboptimal compared to the reaction that protects the economy from hysteresis.

***Definition* 2.7** The monetary policy errors set $\tilde{S}$ contains policy platforms $\dot{M}_t$ that arenot sufficiently expansionary relative to $\dot{M}^*_t$. The lower bound of the set is the deepest possible error $\dot{M}^E_t$.

$$\tilde{S} = \{\dot{M}_t : \dot{M}_t^E \leq \dot{M}_t < \dot{M}_t^* \wedge t \in \mathbb{R}^{\geq 0}\}. \tag{2.31}$$

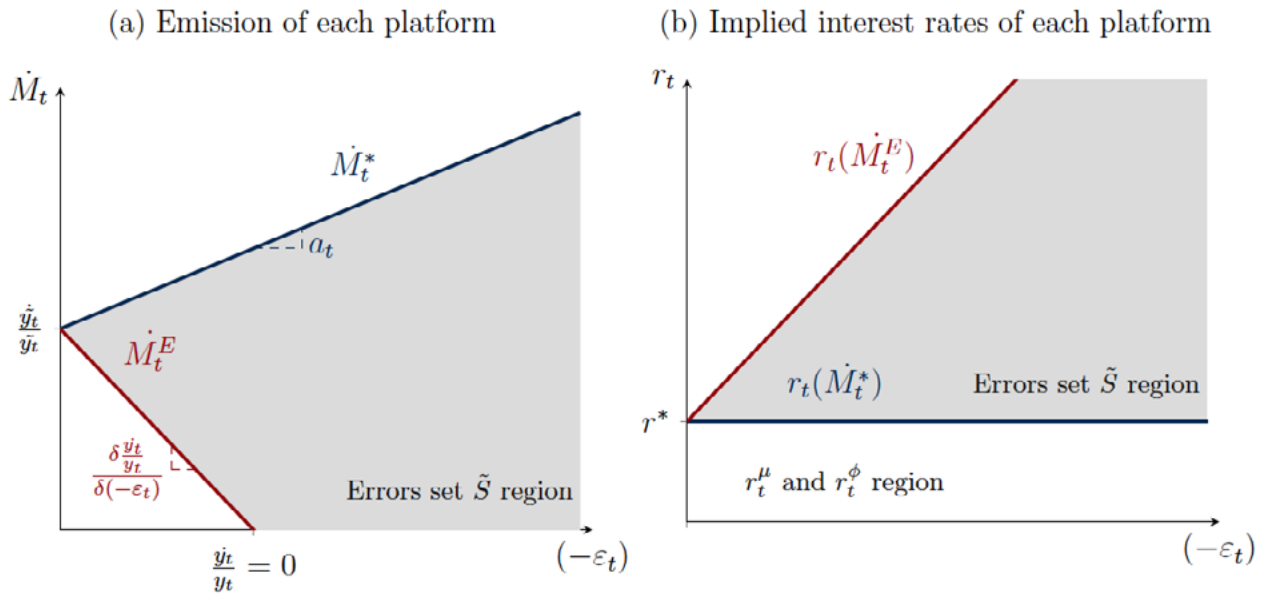
If the monetary policy platform is optimal or a mistake depends on the central bank's regime. The main difference between an optimal policy reaction and the error deviation is the attention to shocks. The optimal decision is to center the monetary efforts on alleviating the effects of crises, especially the latent costs in the long-run trend. In this sense, expansionary monetary policy is the main instrument to succeed in facing recessions. If the expansion is not strong

[30] Note that this behavior is consistent with a Taylor rule centered only on inflation, for example $r^E_t = r + \psi'\pi_t$, or with a procyclical Taylor rule $r^E_t = r* - \psi\varepsilon_t$. See Appendix A.2.3 for more details.

enough, the drop in savings caused by harmful shocks generates a rise in interest rates in a particularly bad moment. Figure 3 exemplifies the contrast between the responses of a mistake vis-a-vis an optimal decision.

Figure 3.

Optimal Decision and Monetary Policy Mistake Representation



Note: The Figure plots the behavior of emission $\dot{M}_t$ and interest rates rt suggested by equations 2.28 and 2.30 assuming $M_t = 1$. For simplicity, assume that the relationship between emission $\dot{M}_t$ and the magnitude of the recession $(-\varepsilon_t)$ is linear. In absence of shocks $\varepsilon_t = 0$ the economy is at the initial position in the BGP. Note that with the counterweight of $\dot{M}^*_t$, growth is not related to the shock, while in $\dot{M}^E_t$ the shock does affect growth. The bounds of Lemma 2.4 r^μ_t and r^ϕ_t belong to a region between the zero lower bound 0 and the natural interest rate of the BGP r^*, so $r^\mu_t, r^\phi_t \in [0, r^*]$, and their position depends on the adopted policy platform. Remark that in the long-run $r^\phi_t = r^*$. The illustration shows only the positive quadrant.

Both panels in Figure 3 show the expected behavior of optimal $\dot{M}^*_t$ and error $\dot{M}^E_t$ policy platforms and the region of the errors set $\tilde{S}$. Panel (a) depicts the level of emission $\dot{M}_t$ in response to different magnitudes of a recession $(-\varepsilon_t)$.[31]

[31] Recall that the analysis is centered in $\varepsilon_t < 0$.

The optimal intensity of expansionary monetary policy increases as the recession deepens, while the mistaken reaction contracts emission through crises with procyclical policy. Consequently, the probability of making a mistake grows with the magnitude of the shock.

Panel (b) portrays the reaction of interest rates r_t to the monetary actions employed to handle the shock. If the central bank engages in the optimal policy, the interest rate is constant for all shock values and equals the long-run natural interest rate r^*. On the contrary, the monetary contraction of the error policy generates a disproportionate rise in the interest rate. From Lemma 2.2, any overshooting of the interest rate creates productivity costs. Therefore, the policy platforms that belong to the errors set $\tilde{S}$ cause an inefficient rise in the interest rate and hurt the long-run trend.

3. General Equilibrium and Theoretical Predictions

The general equilibrium comprised by the household, the firms, and the central bank determines the economy's long-run performance. This section presents the definition and characterization of the BGP equilibrium. The existence and uniqueness of an equilibrium identified by a constant and common growth rate g^*, depends on the existence of a unique constant natural long-run interest rate r^*. The theoretical predictions uncover the discussion on the implications of negative shocks and monetary policy reactions to recessions. The last part of this section outlines the relationship of monetary policy mistakes and hysteresis in a straightforward general equilibrium argument, essentially based on the partial equilibrium results.

3.1 Balance Growth Path Equilibrium

Definition 3.1 presents the BGP equilibrium concept for the economy in the long-run, assuming that the final consumption good is the numeraire $p_{yt} = 1$.

***Definition* 3.1** Define the BGP general equilibrium as a set of growth paths for prices $\{r_t, w_t, p_{it}, \hat{P}_{it}\}^{\infty}_{t=0}$ and quantities $\{c_t, y_t, a_t, E_t, k_{it}, \phi_t, R_{it}, \pi_t, M_t\}^{\infty}_{t=0}=0$ in the absence of shocks $\varepsilon_t = 0$, that fulfill the conditions:

i. Wages w_t, patents prices $\int_0^1 \hat{P}_{it}$, consumption c_t, production y_t, assets a_t, supply of capital goods $\int_0^1 k_{it}$, and monetary aggregates M_t grow at the BGP equilibrium constant growth rate g^*.

ii. The consumption, capital goods, patents, savings, and labor markets clear.

iii. Optimal decisions determine the behavior of households, the final good firm, capital goods firms, ideas firms, and the central bank. The strategic setting between firms' sectors collapses to a subgame perfect equilibrium.

Several implications characterize the BGP equilibrium of Definition 3.1. Based on Lemma 2.3, under the parametric condition $\Omega > \theta$ for learning externalities and the natural destruction rate, the long-run equilibrium has stable full firms' survival, $\phi_t = 0$ and $\dot{\phi}_t = 0$. Consequently, as explained earlier, from Lemma 2.1 it is clear that downward wage rigidities are not binding and the labor market works at full employment in the long-run $u_t = 0$. Production in the BGP then equals potential output $\tilde{y}_t$ and the economy grows at a constant productivity growth rate $g^* = \frac{\dot{A}_t}{A_t}$. Thus, the long-run growth engine is only vertical innovations that cause productivity growth.

The absence of shocks and the BGP constant growth rate g^* condition indicate that both assets a_t and monetary aggregates M_t grow at the same constant rate. Therefore, credit supply flowing to the innovation sector through the private banks grows at the same rate than average productivity.[32] Under these conditions, the interest rate of the savings market in equation 2.23 is constant and complies with the BGP claims. Then, the interest rate equals the long-run natural interest rate r^*.

From Definition 2.2 and Lemma 2.2 it is explicit that the equilibrium growth rate g^*depends on the interest rate through its effect on the equilibrium innovation effort μ^*. This growth rate g^* denotes the existence and uniqueness of the equilibrium. Hence, the BGP equilibrium relies on the existence of a unique constant natural interest rate r^* that guarantees the equilibrium conditions.

***Proposition* 3.1** There exists a unique constant natural long-run interest rate r^* consistent with the BGP equilibrium conditions of Definition 3.1.

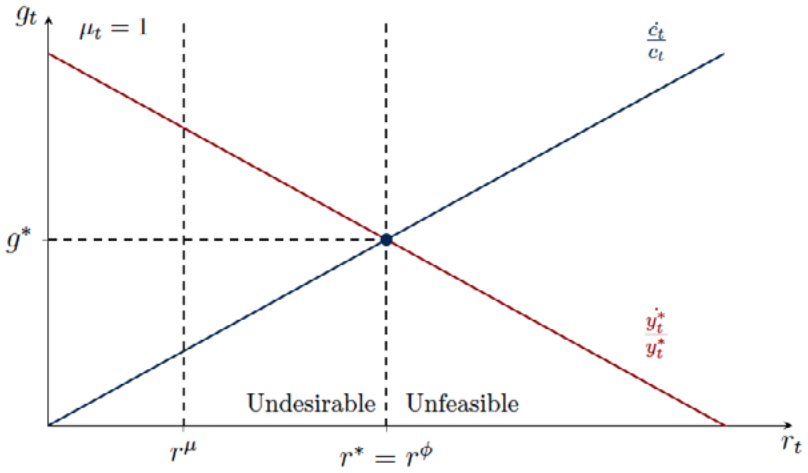
Proof. Balanced growth between consumption and production in the BGP identifies a unique interest rate r_t, $\exists! r_t : \frac{\dot{c}_t}{c_t} = \frac{\dot{y}_t^*}{y_t^*}$. The proof relies upon the necessary parametric condition $1 > \sigma(1 + \rho)$. See appendix A.3.11 for details.

Under Proposition 3.1, there exists a unique BGP equilibrium characterized by g^*. Figure 4 illustrates this equilibrium growth rate g^* and the equilibrium natural interest rate r^*. The natural interest rate r^* achieves a balance between the incentives on savings $\frac{\dot{c}_t}{c_t}$ from the returns on assets, and the costs on growth $\frac{\dot{y}_t}{y_t}$ of more expensive credit. This is the only desirable and feasible equilibrium point, which defines a unique constant long-run growth rate g^* in the BGP. Any other interest rate would be untenable in the long-run: a higher one is unfeasible in terms of production, and a lower one is undesirable from the consumption perspective.

[31] From $\phi_t = 0$ and $\dot{\phi}_t = 0$, and Definition 2.1 is clear that average productivity is A_t.

Figure 4.

The BGP growth rate and the Long-run Natural Interest Rate



Note: The Figure illustrates the BGP equilibrium between consumption growth $\frac{\dot{c}_t}{c_t}$ and production growth $\frac{\dot{y}_t^*}{y_t^*}$. The plotted functions are for growth values gt and the interest rate r_t. g^* is the BGP growth rate, r^* is the natural long-run interest rate, r^ϕ is the full firms' survival interest rate, and r^μ is the full innovation interest rate. The results signal regions for unfeasible, undesirable, and with no growth gains cases. All variables are assumed in the BGP characterization. Recall that the parametric conditions $\Omega > \theta$ and $1 > \sigma(1 + \rho)$ must hold. The $\frac{\dot{y}_t^*}{y_t^*}$ function is plotted assuming $\sigma = 0.5$.

Remark that the parametric condition and consequences of Lemma 2.4 hold in the BGP. As shown by Figure 4, to sustain the equilibrium's consistency, the natural interest rate equals the full firms' survival bliss point $r^* = r^\phi$, but is higher than the full innovation interest rate $r^* > r^\mu$. Therefore, the equilibrium success probability in the ideas sector is constant and $\mu^* < 1$. This result assures a constant growth rate g^*. Creative destruction remains as the single growth source. Lastly, note that any interest rate below the bliss point r^μ does not generate growth gains.[33]

Furthermore, market clearing conditions must equally hold in the BGP. Markets clear if prices listed in Definition 3.1 equilibrate supply and demand. Due to Walras Law, conditions on *i.* capital goods, *ii.* patents, *iii.* savings, and *iv.* labor markets are enough to satisfy a general equilibrium. Thus, the market clearing conditions are the equilibrium values of the capital varieties price p, the patents price $\dot{P}_{it}$, the natural interest rate r^*, and wages w_t.

The partial equilibrium in the capital goods market is defined from the constant monopolist price $p = \frac{1}{\alpha}$ of the first order condition in equation 2.7. In the

[33] Also note that, because of full employment, equations 2.20 and 2.19 map constant bliss points for growth gains from changes in the interest rate, r^μ and r^ϕ.

same order, the patents market equilibrium is achieved when the monopolist spends all its benefits in the purchase of a patent of price $\hat{P}_{it}$, presented in equation 2.11. Note that in the general equilibrium scenario, full employment raises the demand for factors of production, so the patents price $\hat{P}_{it}$ and the supply of capital varieties k_{it} reach a maximum.

The natural interest rate r^* comes from equation 2.23 at full employment and with full firms' survival. Under this interest rate for loanable funds, the savings market is in equilibrium. Besides, reconvene that the equilibrium interest rate is also equal to r^ϕ at full employment in equation 2.19. As well, the labor market equilibrium clears based on labor demand of the final good firm first order conditions in equation 2.4. Specially, labor demand in the BGP equals the marginal product of labor with full firms survival, as wage rigidities do not apply, setting employment at labor supply.[34] Subsequently, the market clearing conditions assure that the economy satisfies the resources constraint:[35]

$$y_t = Nc_t + \int_0^1 R_{it}di + \int_0^1 k_{it}di. \tag{3.1}$$

In the described equilibrium, the first order conditions that summarize the optimality of all agents apply. In particular, equations 2.2, 2.4, 2.7, 2.14, and 2.15 capture the optimal behavior of firms and households. The central bank optimality would depend on the specific monetary policy regime chosen to evaluate. Thus, either rule 2.28 or rule 2.30 can describe the monetary policy platform.

At first, the chosen platform should mean different outcomes in the economy state. However, in the absence of shocks to the BGP, monetary aggregates M_t must grow at a constant rate of g^*, independently of the adopted rule. Thus, in the BGP the central bank never practices extraordinary emission and achieves its goals $\mathcal{L}_t = 0$. As both monetary aggregates and production grow at the same rate, inflation in the BGP arrives and maintains at its target $\bar{\pi}_t = 0$.

The equilibrium of Definition 3.1 captures the economy's usual state: agents interact and decide ruled by the BGP characterization's incentives and conditions. When the economy is in equilibrium, monetary policy does not play a role beyond inflation targeting. Moreover, in the balanced growth equilibrium the optimal and error platforms are the same and achieve the same goal. In regular times, the central bank cannot make a mistake, and it does not have to worry about anything different from inflation. This reasoning could support the idea that monetary policy is unnecessary in this economy. Nonetheless, what happens if there are adverse shocks that produce recessions and distance the economy from the pre-shock long-run trend?

[34] In particular, wages in the BGP are given by $w_t = (1-\alpha)\alpha^{\frac{2\alpha}{1-\alpha}}A_t$.
[35] It collapses to $y_t = Nc_t + \left[\left(\frac{\lambda\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}L}{1+r^*}\right)^{\frac{1}{1-\sigma}} + \alpha^{\frac{2}{1-\alpha}}L\right]A_t$.

3.2 Shocks Around the Balance Growth Path

The economy departs from the BGP because of exogenous shocks on household income ε_t. The shocks cause a business cycle, as they create fluctuations of allocated quantities consistent with expansions and contractions of aggregate economic activity (Zarnowitz, 1992). Thus, the economy suffers volatility around the BGP due to recessions and booms. Recall that on average shocks are 0, but they arise unpredictably and transitorily within the support $\varepsilon_t \in (-1, 1)$ in some periods. Even though the results focus on negative shocks that cause recessions $\varepsilon_t < 0$, this section describes the intuition around both phases of the cycle, centering the analysis in a situation in the absence of monetary policy.

First, a positive shock $\varepsilon_t > 0$ generates a boom due to a rise in assets wealth that increases credit supply. The exogenous gains on the savings market increase available resources for investment in ideas and reduce the private banks' interest rate. Thus, from Lemma 2.2 it is plain that booms improve the ideas environment and benefit innovation effort. This chain of events boosts growth during the positive phase of the cycle. Therefore, it is plausible that booms leave permanent and positive effects on the long-run trend. As growth improves during some periods, when the shock vanishes $\varepsilon_t = 0$ the economy returns to the BGP growth rate g^* in a higher level of production.[36]

[36] In some periods the slope of y_t is higher $\frac{\dot{y}_t}{y_t} > g^*$ and increases the level of production before returning to the usual slope g^*.

Nonetheless, two key issues bound the long-run benefits of booms. Once the economy reaches full firms' survival and full employment in the BGP, increases in innovation effort are useless to improve horizontal growth or employment. Hence, from Lemmas 2.3 and 2.1 it is clear that booms' growth gains caused by a fall in the interest rate benefit the long-run trend only through quality improvements. Moreover, as in Lemma 2.4, the magnitude of the benefits depends on how ample is the available space to undershoot the interest rate before exhausting growth gains, at a position equal or lower than the full innovation bliss point r^{μ}. Note that the distance between the natural interest rate r^* and the full innovation interest rate $r\mu$ that mediate booms' permanent effects is a purely parametric issue and depends on the model's chosen parametrization for Ω and θ.[37]

Furthermore, in the absence of monetary policy, the behavior of inflation from equation 2.21 implies a deflationary effect from booms. Then, an overheating economy does not generate inflation costs. Nevertheless, the model's purpose is not to explain booms but to understand the hysteresis effect from the interaction of recessions with monetary policy. A plausible extension that may help analyze this side of the cycle is introducing inflationary costs from falls in the interest rate, expanding the definition of inflation with lower aggregations of the monetary mass. Overall, as far as the model's scope goes, business cycles' potential benefits are much lower than the potential costs: the dynamics of unemployment and firms' destruction operates only during crises, while the expected effects of booms depend on the size of possible growth gains designated by the parameters.

Now, a negative shock $\varepsilon_t < 0$ generates a recession because of exogenous assets' losses in the savings market. The fall of households savings that flow to credit supply diminishes available resources for investment in R&D, which increases the marginal cost of loanable funds. In this case, the shift in credit supply generates a rise in the interest rate that, following Lemma 2.2, harms innovation outcomes. Thus, the negative phase of the cycle shrinks growth, which potentially injures the long-run trend. Contrary to booms, when the shock vanishes $\varepsilon_t = 0$ and the economy returns to the BGP, the growth rate recovers to g^* in a long-run trend inferior to its pre-shock counterpart.[38]

Therefore, the model directly links recessions with hysteresis. Given that growth recovers at most to g^*, the healing is not strong enough to counterpoise the output losses of crises. Moreover, the size of the long-run scars depends on the depth of the productivity upgrades, firms exit, and unemployment costs. Relative to the direct effect on quality improvements of a lesser innovation effort during recessions, the magnitude of the hysteresis effect grows with a firms' destruction process independent from each firm's expected productivity, which pushes unemployment up.

[37] Figure 4 of the equilibrium characterization and Lemma 2.4 help to illustrate this point.
[38] In this case, the slope of yt during crises is lower $\frac{\dot{y}_t}{y_t} < g^*$, which decreases the production level before returning to the usual slope $g*$.

In the absence of monetary policy, recessions hurt the long-run trend. Are the scars of recessions inevitable? It depends on the central bank. As mistakes arise because of bad reactions to crises, the chosen platform can withdraw the shock with countercyclical policy or conduct the economy towards hysteresis and a weaker long-run trend. In general, monetary policy mistakes would exacerbate the magnitude of the predicted hysteresis consequences, while optimal monetary policy controls the volatility around the equilibrium state. This is the main topic of the next section.

3.3 Hysteresis from Monetary Policy Mistakes

Consider a negative shock around the BGP $\varepsilon_t < 0$ in the presence of an active central bank. As Lemmas 2.6, 2.7, 2.8, and 2.9 suggest, the optimal platform protects the economy from long-run costs. As well, based on Lemma 2.10 the fear of inflation from the sub-optimal central bank during a crisis leads to long-run scars. This section assesses the theoretical predictions and the intuitive lessons from the central bank's optimal $\dot{M}^*_t$ and mistaken $\dot{M}^E_t$ responses facing a recession.

First, suppose that the central bank regime is hysteresis driven, as in Definition 2.5, and follows the optimal monetary policy rule $\dot{M}^*_t$ In equation 2.28. Based on Lemma 2.5, when a negative shock $\varepsilon_t < 0$ reduces credit supply from household assets, the optimal rule commands that the central bank perfectly balances the resources available for R&D investment. This utterly countercyclical policy succeeds in countering the recessions' effects and keeping the economy in the BGP equilibrium.

***Proposition* 3.2** If the central bank follows the optimal rule $\dot{M}^*_t$, then it performs expansionary monetary policy during recessions $\varepsilon_t < 0$ and offsets the resource losses of the credit market. Thus, the interest rate stays at the natural interest rate r^* consistent with the BGP, keeping growth stable at the pre-shock rate g^*. Therefore, the economy does not suffer hysteresis, and the long-run trend prevails recessions.

Proof. See appendix A.3.12.

After negative shocks, the central bank emits enough monetary aggregates to the credit market to restrain adverse consequences on R&D investment. This

reaction keeps the interest rate constant on its natural level. Thus, the economic conditions stay at the pre-shock BGP scenario with an identical growth rate. As monetary policy avoids growth costs from recessions, the long-run trend does not fall, and the central bank saves the economy from hysteresis.

In this case, the central bank prevents the spread of the shock to the economy. Recessions do not cause firms' destruction nor unemployment, potential output does not decline, and productivity growth keeps stable. Moreover, after the shock vanishes and the savings market recovers, the optimal rule dictates an inflation targeting behavior, withdrawing useless excessive emission. Consequently, emission falls to the usual level of the BGP, and the central bank attains the inflation target in the long-run.

Now suppose that the central bank follows an obsessed with inflation monetary policy regime, like in Definition 2.6. In this case, the chosen emission platform $\dot{M}_t^E$ of equation 2.30 drives an inflation driven monetary policy mistake. Then, if a negative shock $\varepsilon_t < 0$ hits the savings market, the central bank performs procyclical policy to achieve the inflation target $\bar{\pi}_t = 0$. As stated in Lemma 2.10, the fear of inflation from excessive emission of the mistaken regime leads to permanent effects on productivity.

***Proposition* 3.3** The monetary policy mistake $\dot{M}_t^E$ drives hysteresis due to a monetary grip during a recession $\varepsilon_t < 0$.

Proof. See appendix A.3.13.

The deepest error policy platform motivates a monetary contraction during bad times. Consequently, the interest rate rises because of a fall in assets from the shock and a procyclical shrinkage of monetary aggregates. Available resources for R&D investment of the savings market fall, and innovation effort decreases. This lower frequency of ideas creation entails a reduction in the growth rate. Thus, growth declines during recessions and the long-run trend shifts to an inferior level. Moreover, after the shock fades away, growth recovers to the pre-shock BGP rate, but the trend never recovers to its previous path, i.e. the economy suffers hysteresis.

The mistaken procyclical policy's permanent effects come from costs in productivity upgrades, firms exit, and unemployment. If the rise of the interest rate is strong enough to push innovation effort to unusually weak levels, firms' destruction springs creating unemployment due to sticky wages, which amplifies the monetary policy error's harmful effects on growth. However, when the economy recovers to the BGP, unemployment disappears, and all the firms survive. Thus, firms' destruction and unemployment work as complementary channels of the total productivity losses from the monetary policy mistake, magnifying the fall of the long-run trend.

In this sense, even though growth recovers to the BGP pre-shock rate and the economy converges to full firms' survival and full employment after the recession, a monetary policy mistake causes hysteresis. In particular, the harsh effect on total productivity would always leave scars on long-run production. This result implies that changes in productivity improvements during recessions are a key channel to understand hysteresis, highlighting the role of individual productivity among surviving firms, rather than the usual unemployment and firms exit approach (Ouyang, 2009; Eslava et al., 2010; Blanchard and Summers, 1986b; Ball et al., 1999; Stockhammer and Sturn, 2011; Gali, 2020).

Note that the same logic of the deepest monetary policy error $\dot{M}_t^E$ in proposition 3.3 applies to any emission platform in the errors set $\tilde{S}$. A non-expansionary enough monetary policy platform $\dot{M}_t < \dot{M}^*_t$ means that the central bank does not recover the loss of resources entirely, so the interest rate tightens during a recession. Thus, any monetary policy mistake that performs a poor expansion during a crisis leads to a fall in the long-run trend that never gets to recover. The magnitude of the output loss from hysteresis in a set of periods grows as emission gets closer to the severest possible mistake $\dot{M}_t^E$. Performing contractionary monetary policy during a recession is the worst possible scenario, but an insufficient monetary expansion would also have permanent costs. The central bank's decision must be sharp on the optimal policy to avoid hysteresis.

Furthermore, the hysteresis prediction does not rely on the persistence of errors. On the contrary, a transitory monetary policy failure is irreparable and leaves long-run scars. The path dependence of productivity growth in Definition 2.2 emphasizes the impossibility of fixing an error's consequences adopting the optimal platform. The linkage between productivity increases and previous total productivity hinders a recovery to the pre-shock productive capacity after a transitory mistake. Besides, recovering from hys-teresis demands a rise in growth above the equilibrium rate. Although, the central bank's incentives forbid overshooting growth with an undershot of the interest rate due to in-flationary costs. Therefore, after a crisis, growth recovers but is never higher than the balanced growth rate g^*, so the economy cannot return to its pre-shock path.

The results presented in this section suggest that monetary policy is non-neutral in the long-run. If the central bank makes the mistake of letting interest rates rise too much during bad times, permanent cost arises due to harmful effects on available resources for investment in R&D. This monetary policy error is plausible when the central bank suffers from a government failure that ties its

hands to a regime inconsistent with crises' economic conditions. Moreover, an inconvenient design restricts an expansionary-enough monetary policy during recessions, which is critical to assess potential hysteresis threats. Focusing on the wrong priorities while leaving aside crises unbalances hurts the long-run trend. This prediction opens a debate on the flexibility of monetary policy regimes, the primary subject of the next section.

4. Further Remarks for Central Banking

This paper's theoretical predictions imply a direct recommendation for central banking: monetary policy should be expansive during bad times. Moreover, the expansion should be sufficiently large to avoid inefficient rises in the interest rate that may harm credit costs and available resources for investment, properly in R&D. Central banking can achieve this as long as monetary policy enjoys some degree of discretion to perform potentially inflationary expansions when recessive shocks arise.

This paper's theoretical predictions imply a direct recommendation for central banking: monetary policy should be expansive during bad times.

Evading hysteresis with countercyclical policy produces short-run inflationary costs due to excessive emission. Nevertheless, these costs seem to be insignificant compared to the hysteresis threat. Furthermore, once monetary policy has shielded the long-run trend from recessions, the central bank can focus its strengths on reaching the inflation target in the long-run. Centering too much on inflation targeting, especially during recessions, is risky. It may restrain the central bank's disposition to save the economy from long-run scars, a monetary policy mistake.

Moreover, in the light of Aghion et al. (2019) results on measurement error of inflation estimates, an obsession with inflation targeting is even riskier. If the current approach to measuring inflation leads to overestimated values because of creative destruction, then central banking's anti-inflationary bias is misplaced. Thus, inflation targets may be too low to attain in crises, increasing the hysteresis costs of making a mistake, as the procyclical effort would be stronger.

The distinction between the optimal policy and errors relies on the outline of different monetary policy regimes. Then, why does the central bank always keeps the same design even though it hurts the economy? In the end, it could perform a strict Pareto improvement changing its policy scheme to the hysteresis driven objectives. It is not that easy. Usually, the central bank's intentions are predefined by an institutional setting, harnessing the capacity of suiting goals to needs. Thus, monetary policy regimes should be more flexible during crises, particularly regarding the fixation on inflation targeting.

Besides, the theory suggests another lesson for central banking during crises: after saving the economy from hysteresis, the central bank must take care of inflation. In particular, it must focus on long-run inflation. If the central bank's credibility keeps expectations anchored during recessions, it should not fear over-the-target inflation strictly from expansions during bad times. In the long-term, inflation should be stable in the target, as monetary policy affects long-run inflation mostly through expectations rather than using monetary grips (Hazell et al., 2020).

Furthermore, the results of this paper open questions for future research. The natural step forward is to test the empirical relevance of the theoretical findings. Fortunately, the set of testable hypotheses inferred by the model can be evaluated using aggregated cross-country data or firm-level microdata. Another potential route of research is extending the model with heterogeneous agents in the exposure to unemployment to consider the potential effects on inequality. In this setting, monetary policy mistakes could affect both growth and income distribution, leading to hysteresis with inequality and poverty traps. Finally, one last question that arises is the implications of inflationary costs. An alternative to evaluate this issue is introducing nominal wage rigidities that restrain the adjustment of real wages and create unemployment due to an inflationary environment.

5. Conclusions

This paper presents a long-run endogenous growth model where monetary policy can attenuate scarring effects from recessions or intensify them. In particular, depending on the regime that governs the central bank's decision, the expansionary reaction to crises may fall short to restrain permanent effects. Moreover, a central bank that focuses solely on inflation during bad times performs procyclical policy and exercises a monetary policy tightening during a recession, which is the deepest possible monetary policy error. In this case, the theoretical framework highlights three main mechanisms for hysteresis: a slowdown in productivity growth, an indiscriminate firms' exit process, and a rise of unemployment. Finally, after the monetary policy error and the shock fade away, even though growth recovers its pre-shock rate and the economy

returns to full firms' survival and full employment, the long-run trend never recovers to its previous level.

In this sense, the theoretical predictions of the model highlight two main lessons. First, despite firms' exit and unemployment are two key channels to understand the scarring effects of recessions, the productivity losses due to fewer vertical improvements among surviving firms also identify a crucial mechanism overlooked by the literature. Second, monetary policy must be sufficiently expansionary during recessions. The short-run inflation costs of this countercyclical policy are certainly lower than the infinite costs of hysteresis. Monetary policy is non-neutral in the long-run, so setting explicit policy objectives in potential growth for stabilization policy is key to deal with the aftermath of crises. On the contrary, the central bank could make an infinitely costly mistake without a chance for vindication.

References

Abiad, A. G., Balakrishnan, R., Koeva Brooks, P., Leigh, D., and Tytell, I. (2009). What's the damage? medium-term output dynamics after banking crises. IMF working papers, pages 1–37.

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation, and growth. American Economic Review, 108(11):3450–91.

Adrian, T. (2018). Remarks on optimal monetary policy. IMF Speeches.

Agarwal, R. and Gort, M. (2002). Firm and product life cycles and firm survival. American Economic Review, 92(2):184–190.

Aghion, P., Akcigit, U., and Howitt, P. (2014a). The economics of growth new chapter, chapter Chapter 5: The Schumpeterian Model. Re-worked version of chapter 4.

Aghion, P., Akcigit, U., and Howitt, P. (2014b). What do we learn from schumpeterian growth theory? In Handbook of economic growth, volume 2, pages 515–563. Elsevier.

Aghion, P., Askenazy, P., Berman, N., Cette, G., and Eymard, L. (2012a). Credit constraints and the cyclicality of r&d investment: Evidence from france. Journal of the European Economic Association, 10(5):1001–1024.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2019). Missing growth from creative destruction. American Economic Review, 109(8):2795–2822.

Aghion, P., Farhi, E., and Kharroubi, E. (2012b). Monetary policy, liquidity, and growth. NBER Working Papers.

Aghion, P. and Howitt, P. (1990). A model of growth through creative destruction. NBER Working Papers.

Aghion, P. and Howitt, P. W. (2008). The Economics of Growth. MIT press.

Akcigit, U., Celik, M. A., and Greenwood, J. (2016). Buy, keep, or sell: Economic growth and the market for ideas. Econometrica, 84(3):943–984.

Ball, L. (2001). Laurence ball comment on "why price stability?" of otmar issing. ECB panel discussion.

Ball, L., Mankiw, N. G., and Nordhaus, W. D. (1999). Aggregate demand and long-run unemployment. Brookings papers on economic activity, 1999(2):189–251.

Ball, L. M. (1997). Disinflation and the nairu. In Reducing inflation: Motivation and strategy, pages 167–194. University of Chicago Press.

Ball, L. M. (2014). Long-term damage from the great recession in oecd countries. NBER Working Papers.

Barro, R. J. and Gordon, D. B. (1983). A positive theory of monetary policy in a natural rate model. Journal of political economy, 91(4):589–610.

Bergoeing, R., Kehoe, P. J., Kehoe, T. J., and Soto, R. (2002). A decade lost and found: Mexico and chile in the 1980s. Review of Economic Dynamics, 5(1):166–205.

Bernanke, B. S. and Gertler, M. (1995). Inside the black box: the credit channel of monetary policy transmission. Journal of Economic perspectives, 9(4):27–48.

Bilbiie, F. O., Fujiwara, I., and Ghironi, F. (2014). Optimal monetary policy with endogenous entry and product variety. Journal of Monetary Economics, 64:1–20.

Blanchard, O. (2018). Should we reject the natural rate hypothesis? AEA.

Blanchard, O. and Gali, J. (2010). Labor markets and monetary policy: A new keynesian model with unemployment. American economic journal: macroeconomics, 2(2):1–30.

Blanchard, O. J. and Summers, L. H. (1986a). Hysteresis and the european unemployment problem. NBER macroeconomics annual, 1:15–78.

Blanchard, O. J. and Summers, L. H. (1986b). Hysteresis in unemployment. NBER Working Papers.

Bosworth, B. (2012). Economic consequences of the great recession: Evidence from thepanel study of income dynamics. Boston College Center for Retirement Research Working Paper, (2012-4).

Cefis, E. and Marsili, O. (2006). Survivor: The role of innovation in firms' survival. Research policy, 35(5):626–641.

Cerra, V. and Saxena, S. C. (2008). Growth dynamics: the myth of economic recovery. American Economic Review, 98(1):439–57.

Chu, A. C. and Ji, L. (2016). Monetary policy and endogenous market structure in a schumpeterian economy. Macroeconomic Dynamics, 20(5):1127–1145.

Clementi, G. L. and Palazzo, B. (2016). Entry, exit, firm dynamics, and aggregate fluctuations. American Economic Journal: Macroeconomics, 8(3):1–41.

Coeure, B. (2017). Scars or scratches? hysteresis in the euro area. Speech at the International Center for Monetary and Banking Studies, Geneva, 19.

De Nardi, M., French, E., and Benson, D. (2011). Consumption and the great recession. Technical report, National Bureau of economic research.

Dosi, G., Fagiolo, G., and Roventini, A. (2010). Schumpeter meeting keynes: A policyfriendly model of endogenous growth and business cycles. Journal of Economic Dynamics and Control, 34(9):1748–1767.

Dosi, G., Pereira, M. C., Roventini, A., and Virgillito, M. E. (2018). Causes and consequences of hysteresis: aggregate demand, productivity, and employment. Industrial and Corporate Change, 27(6):1015–1044.

Eslava, M., Galindo, A. J., Hofstetter, M., and Izquierdo, A. (2010). Scarring recessions and credit constraints: evidence from colombian plant dynamics. Documento CEDE, (2010-27).

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. American economic review, 105(10):3150–82.

Gali, J. (2020). Insider-outsider labor markets, hysteresis and monetary policy. Technical report, National Bureau of Economic Research.

Garga, V. and Singh, S. R. (2020). Output hysteresis and optimal monetary

policy. Journal of Monetary Economics.

Gertler, M. and Trigari, A. (2009). Unemployment fluctuations with staggered nash wage bargaining. Journal of political Economy, 117(1):38–86.

Giebel, M. and Kraft, K. (2018). Bank credit supply and firm innovation. ZEW-Centre for European Economic Research Discussion Paper, (18-011).

Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. American economic review, 95(1):50–65.

Hall, R. E., Sims, C., Solow, R., and Gordon, R. (1975). The rigidity of wages and the persistence of unemployment. Brookings Papers on Economic Activity, 1975(2):301– 349.

Hazell, J., Herre˜no, J., Nakamura, E., Steinsson, J., et al. (2020). The slope of the phillips curve: evidence from us states. Unpublished manuscript.

Huo, Z. and Rios-Rull, J.-V. (2013). Paradox of thrift recessions. Technical report, National Bureau of Economic Research.

Jorda, O., Singh, S. R., and Taylor, A. M. (2020). The long-run effects of monetary policy. Technical report, National Bureau of Economic Research.

Jung, H., Hwang, J., and Kim, B.-K. (2018). Does r&d investment increase sme survival during a recession? Technological Forecasting and Social Change, 137:190–198.

Kienzler, D. and Schmid, K. D. (2014). Hysteresis in potential output and monetary policy. Scottish Journal of Political Economy, 61(4):371–396.

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. Journal of political economy, 112(5):986–1018.

Krugman, P. (2018). The economic future isn't what it used to be (wonkish). New York Times Opinion.

Lee, N., Sameen, H., and Cowling, M. (2015). Access to finance for innovative smes since the financial crisis. Research policy, 44(2):370–380.

Llano, J. and Urrutia, M. (2012). Los actores en la crisis economica de fin de siglo. Ediciones Uniandes-Universidad de los Andes.

Logeay, C. and Tober, S. (2006). Hysteresis and the nairu in the euro area. Scottish Journal of Political Economy, 53(4):409–429.

Lucas Jr, R. E. (1972). Expectations and the neutrality of money. Journal of economic theory, 4(2):103–124.

Madrid, M. J. A., Garcia-Tabuenca, A., and Galvez, C. S. (2015). R & d investments and firm survival across regions. In Innovation and Entrepreneurship in the Global Economy. Edward Elgar Publishing.

Mancusi, M. L. and Vezzulli, A. (2014). R&d and credit rationing in smes. Economic Inquiry, 52(3):1153–1172.

Marcus, M. (1967). Firms' exit rates and their determinants. The Journal of Industrial Economics, pages 10–22.

Michl, T. R. and Oliver, K. M. (2019). Combating hysteresis with output targeting. Review of Keynesian Economics, 7(1):6–27.

Mishkin, F. S. (1996). The channels of monetary transmission: lessons for monetary policy. Technical report, National Bureau of Economic Research.

Moran, P. and Queralto, A. (2018). Innovation, productivity, and monetary policy. Journal of Monetary Economics, 93:24–41.

Mourougane, A. (2017). Crisis, potential output and hysteresis. International Economics, 149:1–14.

Orphanides, A. and Williams, J. (2011). Monetary policy mistakes and the evolution of inflation expectations. Technical report, National Bureau of Economic Research.

Ouyang, M. (2009). The scarring effect of recessions. Journal of Monetary Economics, 56(2):184–199.

Perez-Reyna, D. (2017). Historia del banco de la republica: crisis de 1999. Capitulo 10. Historia del Banco de la Republica. Crisis de 1999. Pag.: 437-463.

Rees, A. (1951). Wage determination and involuntary unemployment. Journal of Political Economy, 59(2):143–153.

Romer, P. M. (1990). Endogenous technological change. Journal of political Economy, 98(5, Part 2):S71–S102.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. American economic review, 95(1):25–49.

Stockhammer, E. and Sturn, S. (2011). The impact of monetary policy on unemployment hysteresis. Applied Economics, 44(21):2743–2756.

Sturn, S. (2014). Macroeconomic policy in recessions and unemployment hysteresis. Applied Economics Letters, 21(13):914–917.

Summers, L. (2014). U.s. economic prospects: Secular stagnation, hysteresis, and the zero lower bound. National Association for Business Economics Conference.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In Carnegie-Rochester conference series on public policy, volume 39, pages 195–214. Elsevier.

Tucker, P. (2017). What is macroprudential policy for? making it safe for central bankers. Making It Safe for Central Bankers (March 2, 2017). BIS Paper, (91).

Twinoburyo, E. N. and Odhiambo, N. M. (2018). Monetary policy and economic growth: a review of international literature. Journal of Central Banking Theory and Practice, 7(2):123–137.

Urrutia, M. (2020). Interview with andr´es d'avila on the long-run effects of monetary policy. April 24 of 2020.

Wilkins, C. (2014). Monetary policy and the underwhelming recovery. In Remarks by Carolyn Wilkins at the CFA Society Toronto meeting, volume 22.

Yagan, D. (2019). Employment hysteresis from the great recession. Journal of Political Economy, 127(5):2505–2558.

Yellen, J. (1995). Efficiency wage models of unemployment. In Essential readings in economics, pages 280–289. Springer.

Yellen, J. L. (2015). The outlook for the economy. Speech at the Providence Chamber of Commerce, Providence, Rhode Island, May, 22.

Young, A. (1998). Growth without scale effects. Journal of political economy, 106(1):41– 63.

Younsi, M. and Nafla, A. (2019). Financial stability, monetary policy, and economic growth: panel data evidence from developed and developing countries. Journal of the Knowledge Economy, 10(1):238–260.

Zarnowitz, V. (1992). What is a business cycle? In The business cycle: Theories and evidence, pages 3–83. Springer.

A. Appendix

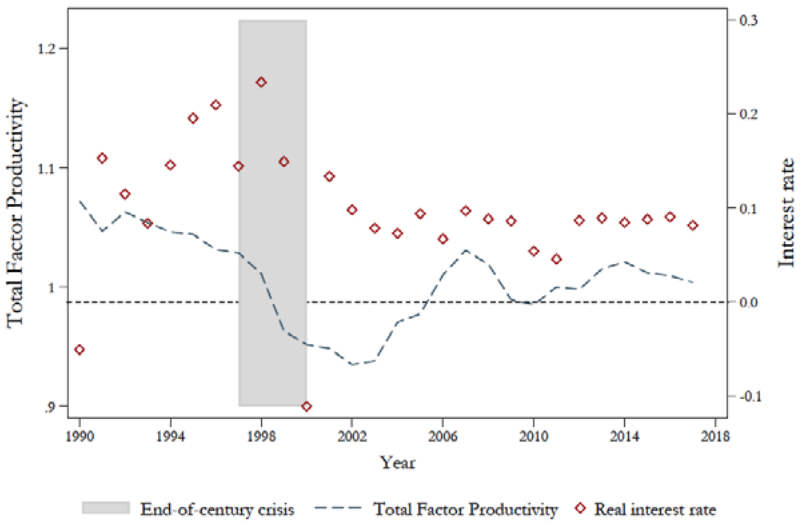
A.1 Motivation

A.1.1 Stylized Fact: Interest Rates and Productivity in Colombia

In 1998 and 1999, Colombia lived the end-of-century crisis in a harsh environment for central banking. A reversal in external capital flows triggered a recession, magnified by unbalances in the housing and financial sectors and a currency crisis (Llano and Urrutia, 2012; Perez-Reyna, 2017). The central bank followed an exchange rate band regime at the time, which impulsed a procyclical policy during 1998 while interest rates rose (Llano and Urrutia, 2012; Perez-Reyna, 2017). Furthermore, the central bank could not perform countercyclical policy before 1999, as defending the exchange rate band was keen to avoid the potential costs of an accelerated devaluation in an uncertain context (Perez-Reyna, 2017). At the end of the recession, the unemployment rate had reached 20%, and GDP per capita had fallen by 5.96% (Llano and Urrutia, 2012). Moreover, the aftermath of the recession left scars from inefficient exit of productive firms (Eslava et al., 2010).

Figure A.1.

Productivity, Interests Rates and the End-of-Century Crisis



Note: The Figure shows time series for productivity and the real interest rate. Productivity comes from the Peen World Table Total Factor Productivity (TFP) index from Feenstra et al.(2015), and the real interest rate comes from World Bank World Development Indicators. The shaded region identifies the years of the end-of-century crisis, according to Llano and Urrutia (2012): 1998 and 1999. The plotted data is for Colombia between 1990 and 2018.

Figure A.1 shows the evolution of Total Factor Productivity and the real interest rate in Colombia between 1990 and 2018. Productivity has a clear decreasing trend between 1990 and the early 2000s, a fall that intensifies during the end-of-century crisis. The interest rate was exceptionally high in these years: between 1993 and 2002, the real interest rate was above 10% and reached a maximum of 23.37% in 1998. After these years of crisis with high interest rates, productivity continued to decrease until 2002. Overall, productivity growth has not been strong enough to exceed the productivity level lived at the beginning of the crisis. This stylized fact points to a feasible relationship between weak productivity growth and interest rate tightenings during bad times.

The indicative evidence from the end-of-century crisis motivates the question about the role of monetary policy during recessions and its effects on the long-term, a feature evaluated by the theory in this paper.

A.1.2 The Central Bankers' Attention in the Long-Run

Urrutia was the governor of the Colombian central bank, *El Banco de la República*, during the end-of-century crisis. Urrutia kindly participated in an interview about the lessons learned for central banking from the recession. In particular, the interview's primary purpose was to recognize the importance that a former central banker places onto the long-run effects of monetary actions.

A quotation from the interview summarizes Urrutia's position regarding the long-run effects of monetary policy.

> ***"From the crisis, we learn that central banking should be careful with the short-run interest rate. Rising the short-interest rate too much has negative effects in the long-run. ... Central bankers frequently write about their experience of crises and strongly take into account this issue."***
>
> ***Urrutia (2020)***

Furthermore, Urrutia is not alone in this perception. In 2016, the former chair of the Federal Reserve Yellen stressed a similar point in a speech.

"Federal Reserve actions to strengthen the recovery may not only help bring our economy back to its productive potential, but it may also support the growth of productivity and living standards over the longer run."

Yellen (2015)

Both quotations of former governors of central banks highlight the relevance of the permanent effects of monetary policy.

A.2 Further Explanations of Theoretical Details

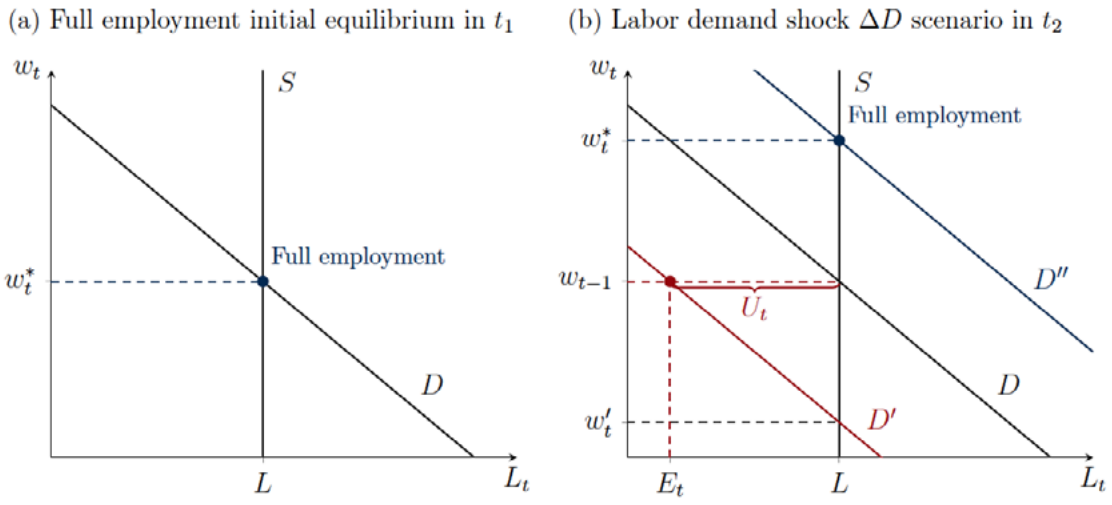
A.2.1 Wage Rigidities

Suppose that time is discrete. The discrete time equivalent for the downward wage rigidities constraint in 2.3 is $w_t \geq \tilde{\kappa} w_{t-1}$. Suppose that the wage resulting from the first order conditions of the final good firm is w^*_t. Then, wages are given by:

$$w_t = \max\left(w_t^*; \tilde{\kappa} w_{t-1}\right). \qquad \text{(A.1)}$$

Figure A.2.

Discrete Time Wage Rigidities Example



Note: The Figure shows two scenarios for a partial equilibrium labor market in presence of perfect wage rigidities $\tilde{\kappa} = 1$. S is constant labor supply that equals L and D is a linear labor demand function. Panel (a) shows the full employment initial equilibrium at time t_1. Panel (b) shows an scenario with two fictional labor demand shocks at time t_2: a negative shock that shifts demand to D' and a positive shock that shifts demand to D''.

Consider the situation depicted in Figure A.2, where the degree of wage stickiness is $\tilde{\kappa} = 1$. Note that in the discrete time version $\tilde{\kappa}$ has a slightly different interpretation. While in the continuous time version κ is the growth rate of sticky wages, $\tilde{\kappa}$ represents how far can wages be from its previous value. Thus, perfectly rigid wages are given by $\kappa = 0$ in continuous time and $\tilde{\kappa} = 1$ in discrete time.

Panel (a) shows the full employment equilibrium before any shock at time t_1. Note that in this scenario, demand D and supply S equilibrium sets a wage equal to w^*_t and employed labor equal to total labor supply L. Panel (b) depicts a situation at time t_2 where a negative shock shifts demand downwards to D' and a positive shock shifts demand upwards to D''. Note that a positive shock that increases labor demand rises equilibrium wages but does not affect employment, as the wage rigidities constraint is not binding. On the other side, a negative shock on labor demand pushes wages down, turning on downward wage rigidities. In absence of sticky wages ($\tilde{\kappa} = 0$) the labor market would equilibrate to full employment and a wage w'_t . However, as this wage is lower than the previous one w_{t-1}, the wage remains unchanged while employment falls to E_t. Due to sticky wages, a negative shock in labor demand produces unemployment U_t.

A.2.2 The Social Planner

The social planner maximizes consumption c_t at period t subject to the resources con-straint of the economy. The level of consumption from the resource constraint is:

$$c_t = \frac{1}{N}\left[y_t - \int_{\phi_t}^{1} k_{it}di - \int_{\phi_t}^{1} R_{it}di\right]. \qquad (A.2)$$

According to Aghion et al. (2014a), solving the social planner's problem requires a two steps process. First, the social planner chooses the quantity of capital intermediate goods k_{it} to maximize consumption, taking the productivity of each variety A_{it} and in-vestment in R&D R_{it} as given. Define $\tilde{c}_t(A_{it}, R_{it})$ as the maximum level of consumption for any A_{it} and R_{it} pair. Therefore, the social planner first solves:

$$\tilde{c}_t(A_{it}, R_{it}) \equiv \max_{k_{it}} \left\{\frac{1}{N}\left[y_t - \int_{\phi_t}^{1} k_{it}di - \int_{\phi_t}^{1} R_{it}di\right]\right\}. \qquad (A.3)$$

The resulting consumption of the social planner is:

$$\tilde{c}_t(A_{it}, R_{it}) = \frac{1}{N}\left[\alpha^{\frac{\alpha}{1-\alpha}}(1-\alpha)(1-u_t)L(1-\phi_t)A_t - \int_{\phi_t}^{1} R_{it}di\right]. \qquad (A.4)$$

In the second step of the solution, the social planner chooses innovation effort μ_{it} evaluating the expected consumption for successful innovations vis-a-vis failures in the creation of ideas. Thus, the social planner decides the optimal rate of innovation solving:

$$c_i^* \equiv \max_{\mu_{it}} \{\mu_{it}\tilde{c}_t(\gamma A_t, R_{it}) + (1-\mu_{it})\tilde{c}_t(A_t, R_{it})\} \quad s.t. \quad \mu_{it} = \lambda\left(\frac{R_{it}}{A^*}\right)^{\sigma}. \qquad (A.5)$$

The solution of the planner's innovation problem sets the optimal innovation rate:

$$\mu_t^* = \left[\alpha^{\frac{\alpha}{1-\alpha}}(1-\alpha)(1-u_t)L(1-\phi_t)\sigma\lambda^{\frac{1}{\sigma}}(\gamma-1)\right]^{\frac{\sigma}{1-\sigma}}. \qquad (A.6)$$

Comparing the planner's innovation rate of equation A.6 with innovation effort of equation 2.15 suggests that innovation effort is suboptimal $\mu_t < \mu^*_t$ if:

$$r_t > \frac{\alpha^{\frac{1}{1-\alpha}}}{(1-\phi_t)(\gamma-1)} - 1. \qquad (A.7)$$

The contrast between the two monetary policy schemes for optimal and error policy platforms can be related to a simple Taylor rules framework inspired on Taylor (1993).

Therefore, the decentralized solution implies underinvestment in R&D if the interest rate satisfies condition A.7.

A.2.3 Taylor Rules for Monetary Policy

The contrast between the two monetary policy schemes for optimal and error policy platforms can be related to a simple Taylor rules framework inspired on Taylor (1993). The optimal monetary response relies in counterweighting the losses in assets from recessions. So, the central bank makes expansionary monetary policy during bad times. Thus, optimal monetary policy is consistent with a countercyclical Taylor rule of the form:

$$r_t = r^* + \psi\varepsilon_t, \qquad (A.8)$$

where r_t is the real interest rate, r^* is the natural interest rate and ε_t is the shock to households income. Note that ψ captures the sensibility of monetary policy actions to recessions, which is implicit in the optimal platform: the central bank perfectly recovers credit supply when negative shocks arise.

This same idea can be extended to a classic Taylor rule in terms of the production gap and inflation:

$$r_t = r^* + \psi\left(y_t - y_t^*\right) + (1-\psi)\left(\pi_t - \bar{\pi}_t\right), \qquad (A.9)$$

where $(y_t - y^*_t)$ is the output gap and $(\pi_t - \bar{\pi}_t)$ is the inflation gap. With $\psi \in (0, 1)$ this Taylor rule implies countercyclical monetary policy, i.e. a fall of production below its efficient level motivates expansionary monetary policy. As the optimal platform suggests a perfect magnitude for the countercyclical policy, the reaction function of equation 2.28 infers an optimal value of the sensibility to recessions ψ^*. In the same fashion as the platforms framework, any weaker than the optimal reference expansionary monetary policy $\psi < \psi^*$ is a monetary policy mistake.

A.3 Proofs

A.3.1 Proof of Lemma 2.1

***Lemma* 2.1** For $\dot{\phi}_t > 0$ sufficiently high such that $\frac{\dot{w}_t}{w_t} = \kappa$, then ut > 0. If $\dot{\phi}_t \leq 0$ or $\dot{\phi}_t > 0$ is not high enough such that $\frac{\dot{w}_t}{w_t} > \kappa$, then $E_t = L$.

Proof. Let $\hat{w}_t$ be the labor marginal product of equation 2.4 in the absence of sticky wages,

$$\hat{w}_t = (1-\alpha)E_t^{-\alpha}\int_{\phi_t}^{1} A_{it}^{1-\alpha}k_{it}^{\alpha}di.$$

It is straightforward that, keeping everything else constant, $\dot{\phi}_t > 0 \rightarrow \dot{\hat{w}}_t < 0$, as $\frac{\partial \hat{w}_t}{\partial \phi_t} < 0$. Under the sticky wages constraint $\frac{\dot{w}_t}{w_t} \geq \kappa$, a sufficiently high $\dot{\phi}_t > 0$ that decreases $\hat{w}_t$ implies that wage rigidities become binding $\frac{\dot{w}_t}{w_t} = \kappa$ and wages stick to $w_t = \tilde{w}_t e^{\kappa t}$. Note that, under the downward wage rigidities it holds that $\tilde{w}_t e^{\kappa t} > \hat{w}_t$. Therefore, from equation 2.5 for employment, it is clear that $E_t < L$ and $u_t > 0$. Finally, note that in the absence of a high enough $\dot{\phi}_t > 0$, which includes $\dot{\phi}_t \leq 0$, wage rigidities are not binding $\frac{\dot{w}_t}{w_t} > \kappa$ and $w_t = \hat{w}_t$ Thus, from equation 2.5 is direct that $E_t = L$.

A.3.2 Proof of Lemma 2.2

***Lemma* 2.2** The innovation outcomes R_{it} and μ_t hold $\frac{\partial R_{it}}{\partial r_t} < 0$ and $\frac{\partial \mu_t}{\partial r_t} < 0$.

Proof. Start from equations 2.14 and 2.15:

$$R_{it} = \left[\frac{\lambda\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L}{1+r_t}\right]^{\frac{1}{1-\sigma}} A_{it} \quad \text{and} \quad \mu_t = \lambda^{\frac{1}{1-\sigma}}\left[\frac{\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L}{1+r_t}\right]^{\frac{\sigma}{1-\sigma}}.$$

Define $\Xi \equiv \sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}(1-u_t)L$. Note that:

$$\frac{\partial R_{it}}{\partial r_t} = -\left(\frac{1}{1-\sigma}\right)(\lambda\Xi)^{\frac{1}{1-\sigma}}(1+r_t)^{\frac{\sigma-2}{1-\sigma}} < 0$$

$$\frac{\partial \mu_t}{\partial r_t} = -\left(\frac{\sigma}{1-\sigma}\right)\lambda^{\frac{1}{1-\sigma}}\Xi^{\frac{\sigma}{1-\sigma}}(1+r_t)^{-\frac{1}{1-\sigma}} < 0.$$

Thus, is clear that $\frac{\partial R_{it}}{\partial r_t} < 0$ and $\frac{\partial \mu_t}{\partial r_t} < 0$ given that the parameters hold $\sigma \in (0, 1)$, $\lambda > 0$ and $\alpha \in (0, 1)$.

For instance, a fall in credit supply that rises r_t lowers both R_{it} and μ_t.

A.3.3 Proof of Lemma 2.3

***Lemma* 2.3** For μ_t sufficiently high such that $\Omega\mu_t > \theta$, then $\phi_t = 0$ and $\dot{\phi}_t = 0$ in the long-run.

Proof. The proof relies in showing that $\Omega\mu_t > \theta \rightarrow \phi_t = 0 \wedge \dot{\phi}_t = 0$. Let $\tilde{\mu}$ be a sufficiently high innovation effort such that $\Omega\tilde{\mu} > \theta$. Start from equation 2.18 of Definition 2.3:

$$\dot{\phi}_t = \begin{cases} (\theta - \Omega\mu_t)(1-\phi_t) & 0 < \phi_t \leq 1 \\ \max(\theta - \Omega\mu_t, 0) & \phi_t = 0. \end{cases}$$

Recall that the size of destroyed firms ϕ_t is bounded $\phi_t \in (0, 1)$. Note that, for $0 < \phi_t \leq 1$, it is true that $\frac{\partial \dot{\phi}_t}{\partial \mu_t} < 0$, so that if $\mu_t = \tilde{\mu}$ and $\Omega\mu_t > \theta$, then $\dot{\phi}_t < 0$. Also note that with $\dot{\phi}_t < 0$, ϕ_t approaches 0 as $t \to \infty$. If the condition $\Omega\mu t > \theta$ holds, in the long-run ϕ_t eventually reaches $\phi_t = 0$. If so, from equation 2.18 it is clear that in the long-run $\phi_t = \max(\theta - \Omega\mu t, 0)$. Under $\Omega\mu_t > \theta \to \theta - \Omega\mu_t < 0$, in the long-run is also true that $\dot{\phi}_t = 0$. Therefore, if μ_t holds $\Omega\mu_t > \theta$, then $\phi_t = 0$ and $\dot{\phi}_t = 0$ when $t \to \infty$.

Figure 2 illustrates the argument of the proof for Lemma 2.3.

A.3.4 Proof of Lemma 2.4

Lemma 2.4 $\exists r^{\phi}_t$ such that $\phi_t = 0$ and $\dot{\phi}_t = 0$, and $\exists r^{\mu}_t$ such that $\mu_t = 1$. If $\Omega > \theta$, then $r^{\phi}_t > r^{\mu}_t$, and in the long-run $r^{\phi}_t = r^*$ and $r^{\mu}_t < r^*$.

Proof. The existence of the interest rates is straight forward from equations 2.19 and 2.20:

$$r_t^{\phi} = \left(\frac{\Omega}{\theta}\right)^{\frac{1-\sigma}{\sigma}} \lambda^{\frac{1}{\sigma}} \sigma (1-\alpha) \alpha^{\frac{1+\alpha}{1-\alpha}} (1-u_t) L - 1 \quad \text{and} \quad r_t^{\mu} = \lambda^{\frac{1}{\sigma}} \sigma (1-\alpha) \alpha^{\frac{1+\alpha}{1-\alpha}} (1-u_t) L - 1.$$

r^{ϕ}_t comes from applying $\phi_t = 0$ and $\dot{\phi}_t = 0$ to equation 2.18 after replacing μ_t with equation 2.15. r^{μ}_t comes from clearing equation 2.15 under $\mu_t = 1$. From equations 2.19 and 2.20 is straightforward that $\Omega > \theta \to r^{\phi}_t > r^{\mu}_t$. From Lemma 2.3 it is clear that in the long-run $r_t = r^{\phi}_t$. As in the long-run the interest rate equals the natural interest rate $r_t = r^*$, it is true that $r^{\phi}_t = r^*$. Finally, as $r^{\phi}_t > r^{\mu}_t$ and $r^{\phi}_t = r^*$, therefore $r^* > r^{\mu}_t$.

A.3.5 Proof of Lemma 2.5

***Lemma* 2.5** Let $\tilde{t} = \{t : \varepsilon_t \neq 0 \wedge t \in \mathbb{R}^{\geq 0}\}$, $\Gamma(\cdot)$ map the changes in $\tilde{t}$, and CS_t be credit supply. $\Gamma(\dot{M}^*_t) = -Na_t\varepsilon_t$ and $N\Gamma(\dot{a}_t) = Na_t\varepsilon_t$. Thus, $\dot{M}_t = \dot{M}^*_t$ implies $\Gamma(CS_t) = 0$.

Proof. Starting from credit supply of equation 2.22, define $CS_t \equiv Na_t + M_t$. Recall emission holds $\dot{M}_t = \dot{M}^*_t$ Apply function $\Gamma(\cdot)$ to CS_t:

$$\Gamma(CS_t) = N\Gamma(\dot{a}_t) + \Gamma\left(\dot{M}^*_t\right) = Na_t\varepsilon_t - Na_t\varepsilon_t = 0$$

It is clear that $\dot{M}_t = \dot{M}^*_t \to \Gamma(CS_t) = 0$.

Therefore, credit supply does not change during shocks if $\dot{M}_t = \dot{M}^*_t$.

A.3.6 Proof of Lemma 2.6

Lemma 2.6 If $\dot{M}_t = \dot{M}^*_t$, then $r_t = r^*$.

Proof. Assume that the economy starts at the BGP, so $r_t = r^*$. Recall from Lemma 2.5 that under $\dot{M}_t = \dot{M}^*_t$, credit supply does not change with shocks. Reconvene that r_t is the price that clears the credit market. As credit supply does not change, the credit market equilibrium price r_t maintains on its pre-shock state. Therefore, $\dot{M}_t = \dot{M}^*_t$ implies that the interest rate holds $r_t = r^*$.

A.3.7 Proof of Lemma 2.7

***Lemma* 2.7** If $\dot{M}_t = \dot{M}^*_t$, then $\frac{\dot{y}_t}{y_t} = \frac{\dot{\tilde{y}}_t}{\tilde{y}_t}$. Therefore, $\phi_t = 0$ and $u_t = 0$.

Proof. Assume that the economy starts at the BGP, so production y_t equals potential output $\tilde{y}_t$. From Lemma 2.6 $r_t = r^*$. Thus, from Lemma 2.2, innovation effort μ_t does not change. Therefore, growth maintains on its BGP level $\frac{\dot{y}_t}{y_t} = \frac{\dot{\tilde{y}}_t}{\tilde{y}_t}$. Then, from Lemma 2.1 and Lemma 2.3, it is clear that $\phi_t = 0$ and $u_t = 0$.

A.3.8 Proof of Lemma 2.8

***Lemma* 2.8** If $\dot{M}_t = \dot{M}^*_t$, then Var $(\frac{\dot{\tilde{y}}_t}{\tilde{y}_t}) = 0$, Var $(\frac{\dot{A}_t}{A_t}) = 0$, and Var$(\mu_t) = 0$.

Proof. Assume that the economy starts at the BGP. From Lemma 2.6 $r_t = r^*$. Thus, from Lemmas 2.2, 2.3, and 2.1, μ_t does not change with shocks and Var$(\mu_t) = 0$. From Definition 2.2, a constant μ_t implies Var$(\frac{\dot{A}_t}{A_t}) = 0$. Therefore, it is straightforward from equation 2.25 that Var $(\frac{\dot{\tilde{y}}_t}{\tilde{y}_t}) = 0$.

A.3.9 Proof of Lemma 2.9

***Lemma* 2.9** If $\dot{M}_t = \dot{M}^*_t$, then $\pi^*_t = \bar{\pi}_t = 0$.

Proof. Start from equation 2.28. In the long-run $\varepsilon_t = 0$. Then, long-run emission is $\frac{\dot{M}_t^*}{M_t^*} = \frac{\dot{y}_t}{y_t}$. From equation 2.21, long-run inflation is $\pi^*_t = \frac{\dot{M}_t^*}{M_t^*} - \frac{\dot{y}_t}{y_t}$. Thus, $\pi^*_t = \bar{\pi}_t = 0$.

A.3.10 Proof of Lemma 2.10

***Lemma* 2.10** If $\dot{M}_t = \dot{M}^E_t$, and $\varepsilon_t < 0$, then $\pi_t = \bar{\pi}_t = 0$ but $\dot{r}_t > 0$ and $\dot{\mu}_t < 0$.

Proof. Start from $\dot{M}_t = \dot{M}^E_t$ Proving that inflation remains on target during recessions is straightforward from equations 2.21 and 2.30, as $\pi_t = \frac{\dot{M}_t}{M_t}^E - \frac{\dot{y}_t}{y_t}$ and $\frac{\dot{M}_t}{M_t}^E = \frac{\dot{y}_t}{y_t}$, so $\pi_t = 0$. Now, from equation 2.2 a negative shock $\varepsilon_t < 0$ reduces the flow of assets to the credit market, as savings fall. Reconvene that rt is the price that clears the credit market. Therefore, following Lemma 2.5, the fall of credit supply implies a rise in the interest rate $\dot{r}_t > 0$. Consequently, from Lemma 2.2, innovation effort falls $\dot{\mu}_t < 0$

A.3.11 Proof of Proposition 3.1

***Proposition* 3.1** There exists a unique constant natural long-run interest rate r^* consistent with the BGP conditions of Definition 3.1.

Proof. Starting from the balanced growth condition $\frac{\dot{c}_t}{c_t} = \frac{\dot{A}_t}{A_t}$ using equations 2.2, 2.17, and 2.15 with full employment $u_t = 0$, the natural interest rate r^* satisfies:

$$(r^* - \rho)(1 + r^*)^{\frac{\sigma}{1-\sigma}} = \zeta\gamma\lambda^{\frac{1}{1-\sigma}} \left(\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}} L\right)^{\frac{\sigma}{1-\sigma}}.$$

Define: $ls \equiv (r^* - \rho)(1 + r^*)^{\frac{\sigma}{1-\sigma}}$ and $rs \equiv \zeta\gamma\lambda^{\frac{1}{1-\sigma}} \left(\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}} L\right)^{\frac{\sigma}{1-\sigma}}$.

The proof relies on the existence of an r^* compatible with the equality $ls = rs$. Note that under the parametric condition $1 > \sigma(1 + \rho)$:

$$\frac{\partial ls}{\partial r^*} = (1 + r^*)^{\frac{\sigma}{1-\sigma}} \left(\frac{r^* + (1 - \sigma - \sigma\rho)}{(1-\sigma)(1+r^*)}\right) > 0$$

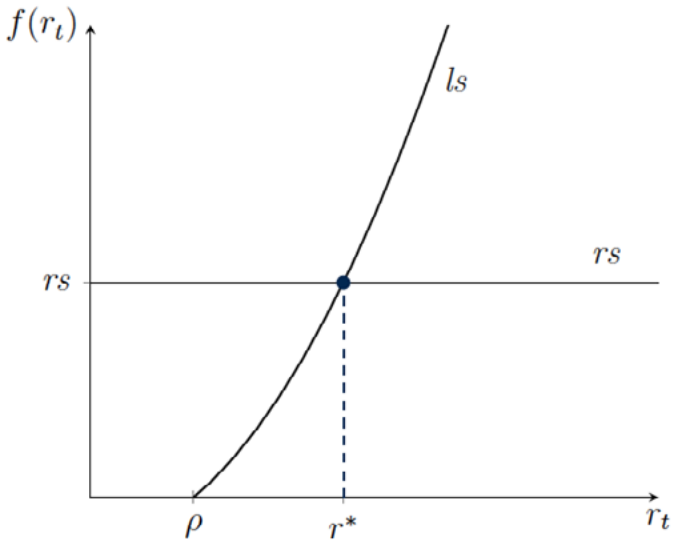
$$\frac{\partial^2 ls}{\partial r^{*2}} = (1 + r^*)^{\frac{\sigma}{1-\sigma}} \left(\frac{\sigma r^* + \sigma(1-\sigma)(1+\rho) + \sigma(1-\sigma-\sigma\rho)}{((1-\sigma)(1+r^*))^2}\right) > 0.$$

Clearly $\frac{\partial rs}{\partial r^*} = 0$. Thus, ls is a strictly increasing convex function in r^*, while rs is constant in r^*. Therefore, there exists a unique value for r^* such that $ls = rs$.

There exists a unique interest rate r^* consistent with the BGP conditions. Figure A.3 is a graphical representation of the argument in the proof. Note that the natural interest rate r^* results from the only point where the functions cross.

Figure A.3.

Graphical Representation of the Proof for Proposition 3.1



Note: The Figure shows a graphical representation of the arguments in the proof for proposition 3.1. *ls* is the left side of the equality given by $\frac{\dot{c}_t}{c_t} = \frac{y_t^*}{y_t^*}$, which functional form complies $\frac{\partial ls}{\partial r_t} > 0$ and $\frac{\partial^2 ls}{\partial r_t^2} > 0$. rs is the right side of the equality, with $\frac{\partial rs}{\partial r_t} = 0$. The natural interest rate r^* is the fulfilled equilibrium when ls and rs intersect. Note that when $r_t = \rho$, $ls = 0$.

A.3.12 Proof of Proposition 3.2

***Proposition* 3.2** If the central bank follows the optimal rule $\dot{M}^*_t$, expansionary monetary policy during recessions $\varepsilon_t < 0$ offsets the resource losses of the credit market. Thus, the interest rate stays at the natural interest rate r^* consistent with the BGP, keeping growth stable at the pre-shock rate g^*. Therefore, the economy does not suffer hysteresis, and the long-run trend prevails recessions.

Proof. The proof relies in showing that the long-run trend slope $\frac{\dot{y}_t}{y_t} = g^*$ does not change with the recession $\varepsilon_t < 0$ if monetary policy follows the optimal rule $\dot{M}^*_t$ Assume the pre-shock scenario is the BGP. Starting from Lemma 2.5, redit supply $Na_t + M_t$ grows at the same BGP pace $\frac{\dot{A}_t}{A_t}$ during recessions. From equation 2.23 in equilibrium with full employment $u_t = 0$ and full firms survival $\phi_t = 0$:

$$r^* = \left(\lambda\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}L\right)\left[\frac{A_t}{Na_t + M_t}\right]^{1-\sigma} - 1,$$

it is clear that Var(r_t) = 0. In particular, as in Lemma 2.6 r_t stays at its pre-shock level $r_t = r^*$. From Lemma 2.2 and equation 2.15, is clear that $r_t = r^* \rightarrow \mu_t = \mu^*$. In accordance to Lemmas 2.1, 2.7 and 2.8, and Proposition 3.1, from equations 2.17, 2.18 and 2.10, the equilibrium level of innovation effort maps an equilibrium growth rate $\mu^t = \mu^* \rightarrow \frac{\dot{y}_t}{y_t} = g^*$. The slope of the long-run trend is constant Var ($\frac{\dot{y}_t}{y_t}$) = 0 and equal to the pre-shock rate g^*. Thus, $\varepsilon_t < 0 \wedge \dot{M}_t = \dot{M}^*_t \rightarrow \frac{\dot{y}_t}{y_t} = g^*$.

The optimal expansionary monetary policy response $\dot{M}^*_t$ assures that the slope of the production function yt does not change during recessions Var ($\frac{\dot{y}_t}{y_t}$) = 0 and stays at g^*. Then, the long-run trend follows the same path, and the economy does not suffer hysteresis.

A.3.13 Proof of Proposition 3.3

***Proposition* 3.3** The monetary policy mistake $\dot{M}^E_t$ drives hysteresis due to a monetary grip during a recession $\varepsilon_t < 0$.

Proof. The proof relies in showing that the long-run trend slope decreases during recession $\varepsilon_t < 0$ if monetary policy follows the deepest error rule $\dot{M}^E_t$, so $\frac{\dot{y}_t}{y_t} < g^*$. Assume the preshock scenario is the BGP. Equation 2.30 and Lemma 2.10 imply that emission does not change directly with the shock $\varepsilon_t < 0$. However, when the fall in assets at of credit supply propagates to growth $\frac{\dot{y}_t}{y_t}$, emission becomes contractionary $\dot{M}^E_t < 0$. From equation 2.23 in equilibrium with full employment $u_t = 0$ and full firms survival $\phi_t = 0$:

$$r_t = \left(\lambda\sigma(1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}L\right)\left[\frac{A_t}{Na_t + M_t}\right]^{1-\sigma} - 1,$$

it is clear that $\frac{\partial r_t}{\partial M_t} < 0$, so the mistake $\dot{M}^E_t < 0$ leads to a rise in interest rates $r_t > r^*$, as in Lemma 2.10. From Lemma 2.2 and equation 2.15, it is clear that $r_t > r^* \rightarrow \mu_t < \mu^*$. In accordance to Lemmas 2.1 and 2.10, and Proposition 3.1, from equations 2.17, 2.18 and 2.10, the lower innovation effort indicates a fall in the growth rate $\mu_t < \mu^* \rightarrow \frac{\dot{y}_t}{y_t} < g^*$. The slope of the long-run trend decreases during the shock and diminishes average productivity $(1 - \phi_t)A_t$, increases unemployment ut and hurts production y_t. After the shock fades away $\varepsilon_t = 0$, the economy converges to the BGP and growth recovers its pre-shock rate g^*, but at an inferior long-run trend level. Therefore, $\varepsilon_t < 0 \wedge \dot{M}_t = \dot{M}^E_t \rightarrow \frac{\dot{y}_t}{y_t} < g^*$.

The monetary policy mistake $\dot{M}^E_t$ causes that the slope of the production function y_t falls during recessions $\frac{\dot{y}_t}{y_t} < g^*$. After the shock, growth recovers but the trend is inferior because of productivity costs. Then, the economy suffers hysteresis.

Defeating Colombia´s Cash Dependency: A Systematic Analysis for Promoting Electronic Payments

November 2020

César Pabón and Juan Andrés Paez[1]

[1] Cesar Pabon (cesar_pabon@student.hks.harvard.edu) is student at the Harvard Kennedy School and Juan Andrés Páez (juanpaez@hks.harvard.edu) is a Research Fellow at the Center for International Development at the Harvard Kennedy School. The authors would like to thank Kenneth Rogoff and Arvind Subramanian for their support and generous feedback. We thank Carlos Arango, Andres Velasco, Jose David Torres, and Manuela Restrepo, for their time, comments, and access to valuable information. We would also like to thank other HKS faculty members like Matt Andrews and Ricardo Hausmann for their advice and help to consolidate our policy proposals. Finally, we are also grateful for the financial support received from the Mossavar-Rahmani Center for Business & Government. An earlier version of this document was submitted as a Master's thesis in Public Administration and International Development at the Harvard Kennedy School

Abstract

During the last decades, the government of Colombia has embarked on a comprehensive agenda to promote access to financial products across the country. Paradoxically, the increase in the access to financial products has not yielded a higher usage of electronic payments -cash payments represent 98% in volume and 88% in value of total transactions in the economy. This paper undertakes three exercises to get a better understanding of the drivers behind Colombia's cash-dependency and the most effective governmental actions to enhance electronics payments. First, we systematically evaluate the binding constraints behind the high levels of cash usage in the economy, in which we highlight the low adoption of electronic payments by businesses and a deficient regulatory framework. Second, we report results of a survey-based qualitative analysis that studies 16 central banks in Latin America, which demonstrates that there is a generalized implementation of policies to enhance the use of electronic payments targeted on financial inclusion, reducing informality, and increasing transparency. Third, we methodically analyze five policy recommendations to improve the dynamism of electronic payments in the country considering the technical correctness as well as the political and administrative challenges. Two potential solutions emerge: promoting a strategy to incentivize the acceptance of electronic payments by firms and generating regulatory adjustments to spur financial service providers to develop efficient payment products.

JEL Classification Numbers: E42; J33

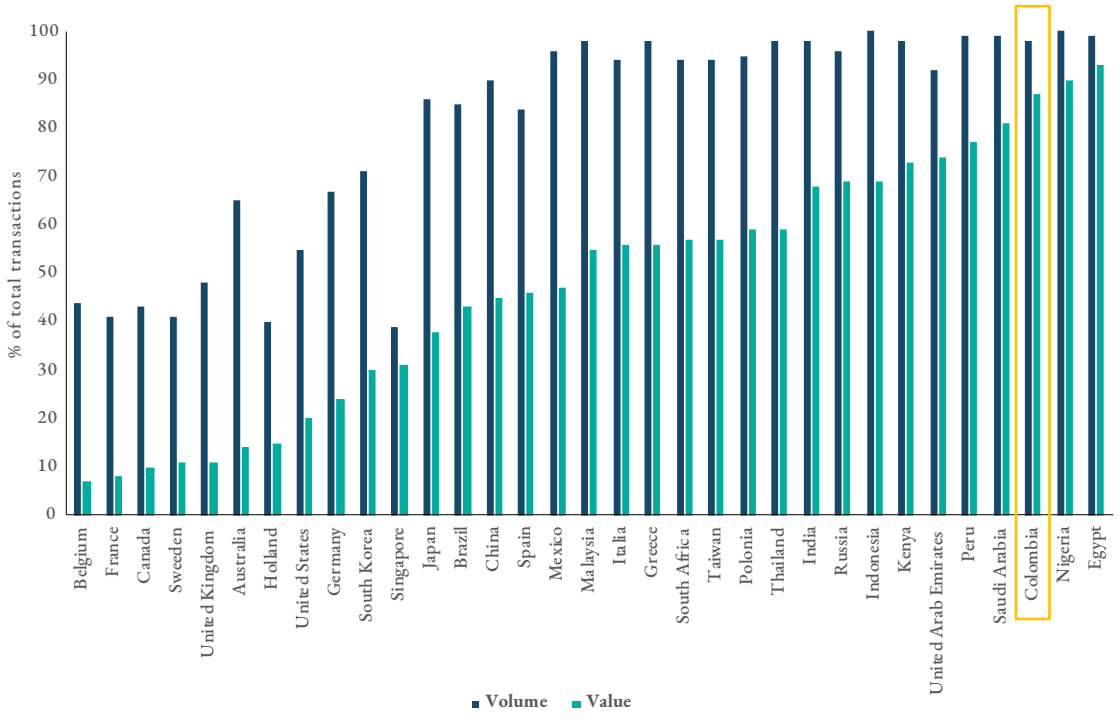
Keywords: Noncash Payments, Payment System

I. Introduction

Despite the development of electronic payments over the last decade, cash remains the most widely used means of payment worldwide, especially for low-value transactions. Colombia has not been a stranger to this phenomenon: the country is one of the most cash-intensive societies, with this means of payment comprising almost 98% of the total volume of consumer transactions and 88% of its value (Arango et al., 2017a). Colombia's value figure ranks as third among 35 economies – only below Egypt and Nigeria- while the volume level is comparable to other developing peers (Figure 1). It is noteworthy that information for Latin American countries is scarce, with only Brazil, Mexico, and Peru revealing similar data availability.

Figure 1.

Transactions carried in cash (% of total transactions)



Source: Mastercard (2013). For Colombia we used the Survey of the Central Bank of Colombia (2016).

Paradoxically, financial inclusion in the country has increased substantially in the last decade, with a higher proportion of people having access to financial products. According to the Financial Superintendence of

Colombia, 81% of the adult population in Colombia had at least one financial product in 2018 – 25 percentage points more than in 2008. Furthermore, the Economist Intelligence Unit (EIU, 2019) ranked Colombia as the country with the best environment for financial inclusion across 55 developing countries. It thus appears that in the extensive margin Colombia has experienced an increase in the penetration of financial products, but in the intensive margin most people are not employing them in their day to day transactions. For instance, only 40% of people with access to debit or credit cards report using it at least once in the last year[2].

The use of cash is not bad *per se* and it has many advantages that explain why it is still the preferred payment instrument worldwide: (i) is fast, easy to use, and does not require any third party or infrastructure for validation; (ii) is highly accepted by the majority of businesses, and in a higher proportion than electronic payments; (iii) has low direct costs as there are no additional fees for using it in transactions; and (iv) is a decentralized means of payment and guarantees the anonymity of transactions (Chekley, 1999).

Despite its well-known benefits, the excessive use of cash by societies is also a matter of concern. Rogoff (2016) demonstrates that, although paper money has become increasingly marginalized in the legal activities of most developed countries, there is a record amount in circulation and most of it is used to finance the underground economy (i.e., tax evasion, corruption, terrorism, the drug trade, human trafficking, among others). The author highlights that anonymity and untraceability of cash is associated with the higher preference for the underground economy, and high denomination bills facilitate their use as a store of value.

The mentioned issues gain relevance for Colombia´s context since they are related to some of its most deep-rooted challenges. For instance, as cash is an effective mechanism to evade the control of the tax authority, the informal sector is prone to use it frequently. This is something to bear in mind in a country where one of its main features is its high rates of informality, which are estimated at 50% in jobs and 60% in firms (Fernandez and Villar, 2019). Also, as cash guarantees anonymity, it enables the proliferation of informal and illegal activities, another main concern for the country considering that the underground economy -which includes drug-related businesses- has been estimated at close to 45% of GDP (Schneider, 2013). It is therefore not surprising that regions with higher rates of organized crime and coca crops exhibit elevated cash withdrawals and low levels of deposits (Information and Financial Analysis Unit, 2015).

In addition, while there is limited direct costs for consumers at the point-of-sale (POS), cash exhibits other costs related to transportation, storage, and time. Notably, there are also higher risks of carrying cash in countries affected by violence and theft. These costs are borne not only by consumers and businesses, but also by the financial and banking sector, which incorporates them into bank fees, affecting the levels of financial intermediation (Mastercard, 2013). According to Mastercard (2013), the burden of cash usage is as much as 1.5% of GDP worldwide.[3]

There is a growing recognition that enabling electronic payments not only offsets some of these negative effects but also enhances economic development, macroeconomic stability and social equity. Specifically, electronic payment penetration can hinder tax evasion and increase tax revenues. Fedesarrollo and Credibanco (2016) found evidence that expanding electronic payments by 10% increase fiscal revenues by half a billion pesos (~0.05% of GDP). Furthermore, lower fiscal evasion could translate into increased government spending and investment toward the provision of public goods and sectors with positive effects for economic development (Myles, 2000).

[2] Central Bank of Colombia (2016).

[3] Also, Humphrey et al. (1996) estimate that complete migration to electronic payment instruments in a given country would result in annual cost reduction equivalent to 1 percent to 3 percent of its GDP.

Electronic payments are highly associated with access and use of financial services, with important positive externalities for the economy and consumers.[4] As shown by Murdoch (1995), financial inclusion is an efficient vehicle to smooth income and consumption cycles. This is critical for households as they protect themselves from economic downturns while channeling internal savings towards efficient investment projects. In addition, access to financial products can help tackle inequality and poverty. Chiba (2009) finds a positive impact of financial inclusion on poverty reduction and the well-being of low-income people in developing countries[5].

The COVID-19 pandemic of 2020 emerges as an ideal opportunity to boost a strategy of greater financial inclusion and digital payments. Amid the pandemic, a significant number of governments have deployed cash transfers and subsidies to protect vulnerable households, workers, and businesses. Even though there is no evidence that the probability of transmission of COVID-19 via banknotes is higher than for other frequently-touched objects, electronic payments and digital technologies are essential to avoid the agglomeration of people and the spread of the disease when they receive these benefits (Auer et al., 2020). Digital payments are also an efficient way of ensuring transparent reporting and adequate allocation of this resources (Muralidharan et al., 2016). Between May and June 2020, Colombia has transferred unconditional cash transfers to more than 2.5 million households from which 57% of them already had a financial product and 39% are newly bancarized (National Planning Department, 2020).

The COVID-19 pandemic of 2020 emerges as an ideal opportunity to boost a strategy of greater financial inclusion and digital payments.

This paper undertakes three exercises to get a better understanding of the drivers behind Colombian cash-dependency and the most pertinent governmental actions to enhance electronics payments. First, following the Growth Diagnostics methodology of Hausmann et al. (2005), this paper evaluates the binding constraints behind the high levels of cash usage. We find that the most important constraints for achieving this goal are the low adoption of electronic payments by businesses and the regulatory framework that limits the provision of new payments products, and reaching sectors of the population that did not previously have access to payment services. On average, the firm's cost-benefit perception of adopting electronic payments is negative, since the country exhibits one of the highest tax rates in the world and high fixed costs of accessing this means of payment. Besides, the current regulatory framework inhibits the access of new players and innovations in the payment system, coupled with a suboptimal oversight coordination.

Our second analytical exercise consists of a survey of 16 central banks in Latin America regarding the databases available in the region and policies implemented for encouraging less cash-dependent societies. We confirmed that there exists a shortage of data availability in the region. Nevertheless, there is a generalized implementation of policies to increase electronic payments - mainly focused on obtaining financial inclusion, reducing informality, and increasing transparency. Direct interventions, such as price regulation or fiscal incentives, are unusual in the region. There has been a growing interest in these countries to promote policies that enhance competition in the payment system and allow non-financial firms to enter the market.

[4] Electronic payments also increase transparency and security of transactions (Muralidharan, 2014), and eliminate some physical and informational barriers that hamper financial inclusion, penetration, and access to capital (Masino, 2014).

[5] With financial inclusion people can make better money management, access financing at a reasonable cost, and access the advantages of having more options than those available in the informal sector (Chiba, 2009).

Finally, we systematically evaluate five policy solutions to enable electronic payments in Colombia and reduce the current levels of cash usage, assessing the technical correctness, political support, and administrative feasibility of the reforms. The government faces administrative and political limitations, and their policy-making capital should be deployed in alleviating binding constraints rather than in going after too many targets at once. Our findings suggest that two policies could improve the dynamism of electronic payments. A first approach relies on promoting a strategy to incentivize the acceptance of electronic payments by firms. This strategy should consist of providing tax incentives to firms to use financial products, combined with a component of financial education for firms to understand how to use electronic payment tools, and recognize the advantages of managing them. The second proposal aims to generate four regulatory adjustments by: (i) strengthening competition between the two automated clearing houses platforms; (ii) promoting corporative governance practices; (iii) defining and delimiting responsibilities of the new actors and processes consistent with the technological progress; and (iv) strengthening oversight arrangements.

It is noteworthy that some of these issues and solutions have already been highlighted by either the academic, private sector and governmental offices. However, the pace and determination to kickstart these policies have been languid. This paper contributes to the literature by using a systematic approach to understand which is the critical constraint that is holding the country back from transitioning into electronic payments, and which are the policies that will yield the highest result towards this objective. We underscore that including management methods and building multi-method capabilities are warranted to obtain an effective public implementation. We also highlight the need for conducting a countrywide communications campaign to raise awareness and build confidence among the business community and the public about the benefits of embracing electronic payment facilities.

Section 2 provides the potential drivers that explain the high usage of cash in the country exploiting a framework on the demand-side and supply-side characteristics of electronic payments. In section 3, we report the results of a survey-based qualitative analysis done to 16 central banks in Latin America and the Caribbean on this topic. In section 4 we analyze the range of policies that could be used in Colombia to promote electronic payments, recognizing potential benefits and risks while identifying their administrative, political, and technical challenges. Section 5 presents a conclusion of the analysis.

II. Cash Usage in Colombia: Diagnostic

We consider an analytical framework, following the Growth Diagnostics methodology[6] of Hausmann et al. (2005), to identify the most binding constraints for the low adoption of electronic payments in Colombia. Although numerous drivers may explain it, we seek to understand which is the critical ceiling that is holding the country back from transitioning into electronic payments. The rationale is that if this constraint is lifted, it will yield the result towards this objective.

[6] Growth diagnostics approach provides a framework for formulating hypotheses on what may be constraining a country's growth; this approach views economic growth as the result of an optimization process under constraints and seeks to identify the factors that are the most binding, in the sense that their removal would allow a growth spurt (Hausmann et al, 2005).

In this framework, however, our attention focuses on Colombian's idiosyncratic and institutional factors

Payment systems are defined as a set of instruments, banking procedures, and, typically, interbank funds transfer systems that ensure the circulation of money, which includes the institutions, instruments, people, rules, procedures, standards, and technologies that make exchanges possible ((BIS, 2003)[7]. We center our analysis on the low value payment systems, which includes most of the daily transactions carried out by consumers, and an important proportion of the payments undertaken by the government and firms[8]. In a similar vein, our analysis focuses on understanding the high levels of cash usage in payment transactions rather than the demand for cash in the economy (related with both exchange and storage values). We also recognize that international evidence has demonstrated that demand for cash is mainly determined by traditional factors such as income, interest rates, and cash provisioning cost (Tobin, 1956). In this framework, however, our attention focuses on Colombian's idiosyncratic and institutional factors, which have been highlighted as critical barriers for the shift to a less cash intensive society (Humphrey et al, 1996).

For simplicity, our analytical framework considers a two-sided market for the low-value payment system that facilitates the exchange of money in the economy. In the demand-side, we consider consumers, businesses, and the government that carry out daily transactions either by cash or other type of payments (e.g. electronic transfers), while in the supply-side, we consider Payment Service Providers (PSP), such as financial institutions and payment infrastructure providers, which process and clear payments between multiple banks, card, and networks.

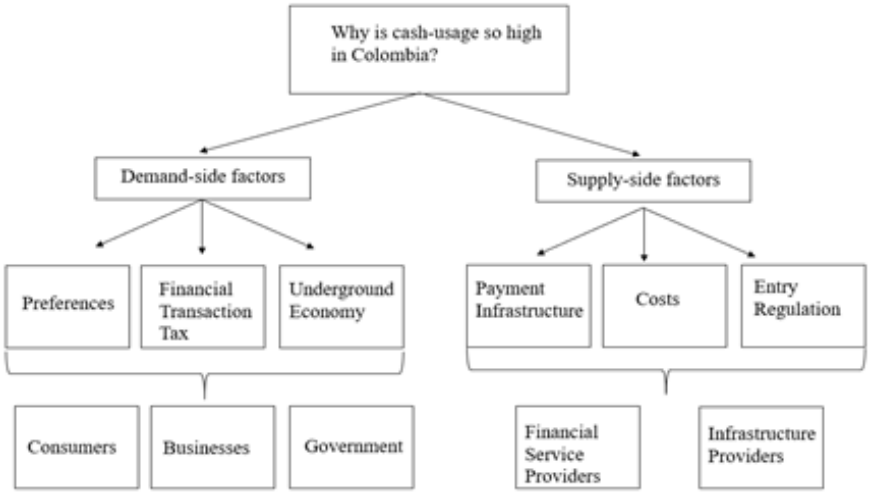
Figure 2 describes our analytical framework. On the left side of the tree, we test for three potential drivers that affect the demand-side of financial transactions from consumers, businesses, and the government: i) general preferences of firms and consumers, ii) the tax on financial transactions, and iii) the underground economy - including both the informal and illegal sectors. On the right side of the tree, we explore potential drivers that affect the supply-side of electronic payments by PSP. We separate these into i) payment infrastructure constraints - i.e. number of point-of-sales (POS) terminals and connectivity access, ii) the costs of financial products affecting financial access and electronic payments usage, and iii) the market structure associated with the regulatory framework and competition practices. It important to underscore that POS terminals are also affected by demand-side factors, but we include them in the right side of the tree considering that payment service providers play the key role of supporting and helping small businesses to reach the "last mile" in a considerably heterogeneous country.

[7] According to the BIS (2014), the payment process includes five stages: pre-transaction, authorization, clearing, settlement, and post transaction. Appendix I summarizes a traditional payment system,

[8] According to the Central Bank (2013) almost 85% of all payments carried out monthly are low value associated.

Figure 2.

Analytical Framework



Source: Own elaboration

II.A. Consumers and Firms Preferences and Characteristics: Cost-Benefit Analysis

Consumer behavior has been at the center of understanding the demand for money. Baumol (1952) and Tobin (1956) based their models upon rational behavior of cash management, which depends on the amount of money of expenditure to be financed, the cost of transaction between money and an alternative interest-bearing asset (bond), and the interest return on the alternative asset. Still, a main concern of these models is that the shift away from cash has progressed at different paces, with behavioral patterns among users partially explaining these trends (Humphrey et al, 1996). Hence, there has been a growing concern that those models were largely abstract, and that understanding consumer payment choice on cash usage is warranted. In this

vein, there has been increasing implementation of consumer payment diaries, that featured detailed information on individual payments and characteristics of individual consumers.[9]

In Colombia, Arango-Arango et al. (2020) use socioeconomic and perception indicators from a consumer payment diary survey to understand the mechanisms these agents choose between different payment instruments. On the consumer side, cash is the payment instrument that has the most perceived net benefits[10], being ten times larger than those seen for debit or credit cards. Individual characteristics such as income, education, and formal employment exhibit a positive relationship with the net recognized benefits of electronic payments. High value of the purchase items and higher probability of acceptance of electronic payments (measured by the density of POS terminals in the place of residence of the individua) are other important determinants. The former point displays the crossed network effects of payment systems - higher penetration of POS terminals by firms can lead to higher use of electronic payments by consumers.

Cash also exhibits a relative perceived advantage compared to other means of payments in terms of access, ease of use, and budget control. According to a survey conducted by Asobancaria, Redeban, and the Central Bank of Colombia (2017), 96% of surveyed people claimed that cash has the advantage of being highly accepted, while 56% claim this is true for debit cards, and 45% for credit cards. Cash is perceived as easier to use by 97% of respondents. In contrast, only 61% and 57% of consumers consider the same for debit and credit cards, respectively. Likewise, consumers show a preference for cash as they believe it allows them to control their budget effortlessly: 76% of people claimed that cash allowed for better budget control. This figure for debit (41%) or credit cards (31%) was significantly lower (Arango et al., 2017b).

Besides, the perceived fixed access costs[11] to debit and credit cards are negatively associated with income, the number of hours worked, and formal employment. These fixed costs are also lower among those individuals who see the adoption of cards as a gateway to other financial products such as credit. Arango-Arango et al. (2020) estimate that the perceived fixed costs of cash are on average half of those observed for debit and credit cards.

On the side of small firms, which comprise 82.1% of the total number of transactions on firms, Arango et al. (2017a) show that the low adoption of electronic payments is explained by an average negative cost-benefit analysis.[12] According to consumer payment booklets, only 15% of small firms in the country accept payment instruments different from cash. While businesses that have monthly sales of more than 20 million COP (~6,000 USD) exhibit an electronic payment acceptance rate of 70%, those with monthly sales below 2 million COP (~600 USD) only exhibit an 11% acceptance rate of this method of payment.

A main determinant that increases the potential advantages of adopting electronic payments by small firms is the expectation of attracting clients and increasing efficiencies. Firms that believe electronic payments increase their competitiveness show a higher likelihood of using electronic payments. Likewise, firms that think they can increase their access to credit through electronic payments have a higher chance of adopting them. Those who exhibit greater average transaction values are also more prone to incorporate electronic payments (Arango et al., 2017). Overall, most businesses that do not accept electronic payments claim that: i) their clients prefer paying in cash, ii) their average transaction values are too low, and iii) their income and expenses are so small that they do not need to have a bank account. These results indicate that some of the small businesses decide to avoid accessing financial products and electronic payments

[9] Bagnal et al (2014) conducted a cross-country comparison with payment diary survey data on seven countries, finding that the use of cash is strongly correlated with transaction size, demographics, and POS characteristics.
[10] Net benefits in this model are measured in utiles rather than in monetary terms.
[11] Fixed costs include financial education level, number of hours worked, and the easiness of use, among others.
[12] For this purpose, the authors use a small firm-level survey, which includes socioeconomic characteristics of small firm owners and the firm, as well as their perception on different benefits and costs of electronic payments.

because of their size, while others perceive low benefits from budget control or additional clients.

In contrast, the acceptance of electronic payments by small firms has a strong negative relationship with their perceived fixed costs in terms of installation, monthly fees, and implementation's easiness, among others. According to Arango et al. (2017a) a reduction of 10% in the perceived costs by firms is associated with an 8.8% increase in the average proportion of firms who accept electronic payments. The authors also find that a higher expectation of paying taxes is correlated with a lower adoption of electronic payments (section II.B.). Finally, the financial education of business owners and the trust for the financial system exhibit a positive association with electronic payments adoption.

In short, our findings suggest that both consumers and businesses pose strong preferences for using cash instead of electronic payments. On the side of consumers, there is a perception that cash is easily accessible, highly accepted, and better for budget control than electronic payments. Additionally, consumers only tend to use electronic payments when the value of the transaction is high and, most importantly, when they perceive there is a high acceptance rate by businesses in the areas where they carry their transactions. For firms, the perceived benefits of electronic payments are on average lower than the perceived costs of adopting them. Even though businesses perceive electronic payments as a potential driver for competitiveness and client attraction, the high perceived costs of adoption and the expectation of future taxes partially offset these benefits. Therefore, there is a problem in this two-sided market with consumers avoiding electronic payments because they are burdensome and are not accepted ubiquitously, and firms not adopting them because, on average, they have a negative cost-benefit analysis. We conclude that reducing the perceived costs for firms, which are the weakest link in this two-sided market, is an essential solution to enable the positive crossed network effects of payments systems.

II.B. Financial Transaction Tax

The second factor potentially constraining the adoption of electronic payments is the financial transaction tax. Like many other Latin American countries, the government of Colombia introduced a financial transaction tax (*Gravamen a los Movimientos Financieros*, GMF) during the late 1990's[13]. The objective of this tax was supporting the mortgage sector, reinforcing the deposit insurance scheme, and avoiding the collapse of the financial sector (BTC, 2015). Currently, the GMF has become one of the most convenient taxes to collect and represents a non-negligible source of government funding (5% of the central government's tax revenues or 0.7% of the GDP according to the Medium-term Fiscal Framework of the Ministry of Finance of 2019).

[13] This tax has undergone several changes since its creation: (i) it changed from temporary to permanent in the year 2000; (ii) its rates have been unified and increased from 0.2% to 0.4% in 2003; (iii) its gradual removal has been rescheduled three times; and (iv) the elimination of its specific destination and the consequent allocation to the National Budget.

Recent literature discovers that this financial tax generated intermediation inefficiencies. For instance, the tax on financial movements was borne primarily on depositors, but since 2002 the government established that intermediaries also had to pay for the liquidation or renewal of fixed-term certificates of deposit. Lozano-Espitia et al. (2017) found that this tax became costly for banks, and this cost was partially or fully transferred to the lending or deposit interest rate. As a result, this meant an increased intermediation margin of around 70 basis points on average. The authors also confirmed a double mechanism employed by banks to compensate for the cost generated by the GMF through higher interest rates on loans and/or lower interest rates on deposits.

Another main concern relates to the fact that this tax represents an additional transaction cost for customers, therefore discouraging the use of bank services. As already stated, the GMF is charged to transactions involving the use of funds deposited in checking or savings accounts by individuals or businesses, including those in the central bank. Consequently, this tax could affect the usage level of different payment instruments by incentivizing people to avoid the financial system and hold cash for their day to day transactions. Giraldo & Bucklet (2011) show that the tax caused a decrease in checking account balances, but adjustments produced in cash total balances after 2001 cannot be attributed to it. In other words, there is evidence of disintermediation, but no indication of a substitution effect from bank accounts towards cash holdings.

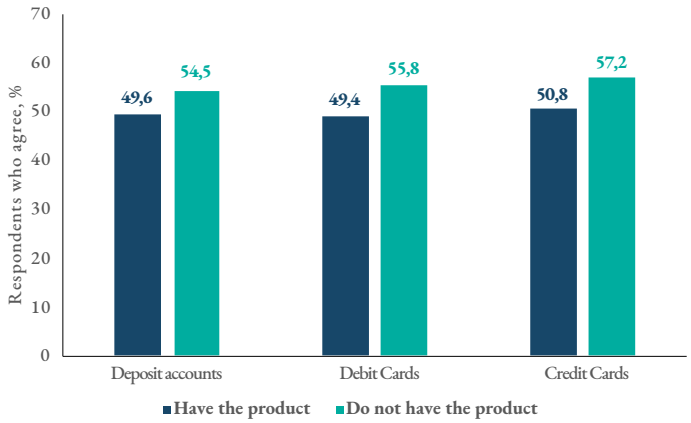
Arango et al. (2018) supports the fact that the GMF does not appear to be the main driving force behind the demand for cash. The authors suggest that the GMF had an impact over the demand for cash in 1998 when it was introduced, but there were no additional effects in the years 2000 and 2003, when it was gradually increased. The authors claim that this can be explained by a one-time change in agents' behaviors, which were not affected by the subsequent changes.

Also, the GMF does not seem to be the most relevant determinant for the cash preference of consumers in Colombia. The Central Bank survey, reported by Arango et al. (2017b), inquires consumers if the GMF is the most important reason in their decision to manage most of their expenses with cash. Figure 3 shows that a higher proportion of consumers who do not have access to financial products believes that the GMF is the most important reason to use cash, compared to those consumers who have access to these products. However, the difference for all categories is considerably small and never exceeds 6.4 percentage points (Figure 3). The authors found a similar response when this question was asked to small firms, confirming that GMF does not appear to be the main determinant for the higher preference of cash usage (Arango et al, 2017a).

We can conclude that the financial tax, one of the usual suspects for the high cash usage in the country, only had a clear effect on the demand for cash when it was implemented in 1999, but no substantial influence when it was gradually increased in later years. If this constraint is binding, we would expect consumers and firms to state that the tax is a critical reason to avoid financial products and transactions. We are cognizant that the GMF naturally creates distortions in financial transactions, but evidence suggests that neither costumers nor firms recognize it as the primary driver behind cash preference. In fact, there are some laws that exempt users from paying the GMF or allow them to receive fiscal benefits.[14]

Figure 3.

Is the GMF the most important reason why you use cash in your expenditures?



Source: Arango et al. (2017a)

II.C. Underground economy: informality, tax evasion and illegal activities

One of Rogoff's (2016) main conclusions is that cash is becoming increasingly marginalized in the legal economy and that, due to its anonymity and untraceability, most of it is used in the underground economy[15] – comprising tax evasion, corruption, terrorism, and drug trade, among others. These issues gain relevance in the context of Colombia as they are

[14] People in Colombia can choose one of their savings accounts to be exempt from the GMF for up to 12.5 million COP (~3,500 USD) during the year. Considering that close to half of Colombians earn less than 1 million COP a year (~300 USD), this exemption covers a considerable amount of transactions. Similarly, 50% of the GMF tax is deductible from the personal income and corporate income tax statements of people and businesses (BTC, 2015).

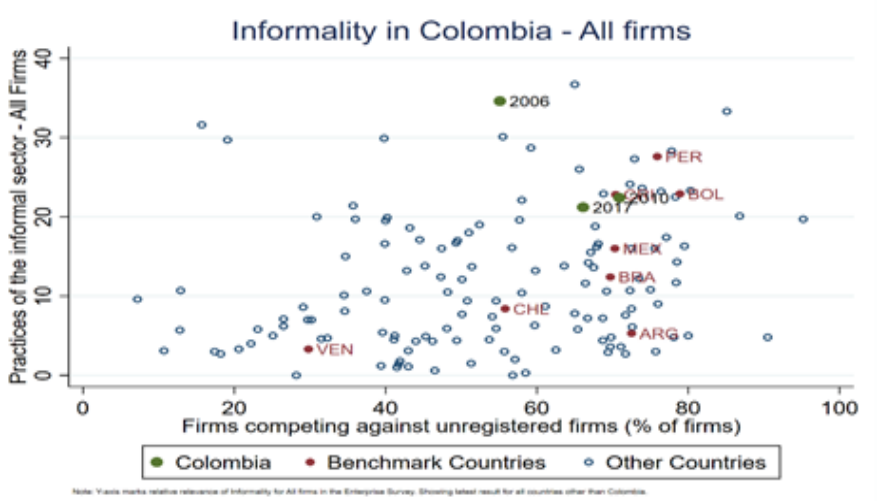
[15] For underground economy we consider the definition of Feige (1989) that refers those economic activities and the income derived from them that circumvent or otherwise avoid government regulation, taxation or observation.

related to some of the most deep-rooted challenges that the country faces. Schneider (2013) estimates that the size of the underground economy represented close to 45% of GDP between 1980-2012.

Colombia's informality levels, understood as the non-compliance of firms with legal requirements (e.g., taxes and social security contributions), reflect the considerable size of this issue. Labor and firms' rates of informality are estimated at 50% and 60%, respectively (Fernandez and Villar, 2019). High informality is not a unique feature of Colombia within Latin America. In fact, almost 70% of the firms in most countries of the region report competing against unregistered firms (Figure 4). However, the topic generates more concerns in Colombia considering that 21.2% of firms claim that informality is the biggest obstacle for growth, significantly above the levels seen in Latin America and the Caribbean (LAC) countries (12.4%) and the world average (11.8%). Only Peru and Bolivia exhibit similar levels of concern for this issue.

Figure 4.

Informality for businesses: Colombia and benchmark countries



Source: World Bank Enterprise Surveys (2017). Note: Y-axis marks relative relevance of practices of the informal sector, (i.e.) competing against informal firms, for all firm. Showing the latest result for all countries.

Informality is usually associated with higher levels of cash usage as firms use this payment instrument to avoid the oversight from the government. This includes firms that are broadly engaged in legal activities but who are avoiding taxes and regulations. By contrast, there is evidence that informality has a negative relationship with the adoption of electronic payments[16]. Intuitively, electronic payments increase the traceability of transactions impeding tax evasion by both formal and informal firms. In Colombia, informality represents a relevant determinant of electronic payment adoption: the proportion of formal businesses accepting electronic payments is 23 percentage points higher compared to the informal sector (Arango et al., 2017a).[17]

[16] Kearney (2015) finds a strong negative correlation between the size of the informal economy and electronic payments penetration.
[17] The percentage of informal and formal businesses that accept electronic payments is 10% and 33%, respectively.

An interrelated source for the intensive cash usage in the underground economy is tax evasion. The literature demonstrates that an increasing burden of taxes and social security contributions is one of the main triggers for the development of informality (Schneider, 2005). According to the World Bank Enterprise Survey (2017), 19.5% of Colombian firms perceive high tax rates as a critical constraint for their development, above LAC (12.8%) and world (11.9%) averages. Furthermore, the World Bank Doing Business Survey shows that Colombia had a total tax rate equivalent to 71.2% of profits in 2019[18], ranking as the 10th highest total tax rate on firms in the sample - only below regional peers of Argentina (106.3%) and Bolivia (83.7%).

A third major vehicle of the use of cash in the underground economy is through illicit activities, mainly in the drug business. Colombia has been a key player in illegal drug markets during the last 30 years. According to UNODC (2012), between 60% and 70% of the cocaine consumed worldwide is produced in Colombia. The income generated by illegal businesses reached a peak of 12% of GDP in 2001 and then decreased to 3% of GDP in 2013 (Villa et al., 2016). These resources come from the trafficking of cocaine and are also the main source of funding for illegal armed groups, such as paramilitary groups and criminal bands. The traditional way to utilize these funds is via money laundering, mainly by cash[19], which takes proceeds from illicit activities and filters them through seemingly legitimate enterprises to produce licit money[20].

Noteworthy, the drug business relates to other illicit activities in the country. In fact, Rettberg & Ortiz-Riomalo (2016) found links between armed conflict, crime, drugs, and illegal gold mining in Colombia underscoring practices and incentives associated with the development of illicit markets. There are also salient cases of corruption that have been recently processed by the Colombian justice, which usually involved the exchange of large amounts of cash notes[21].

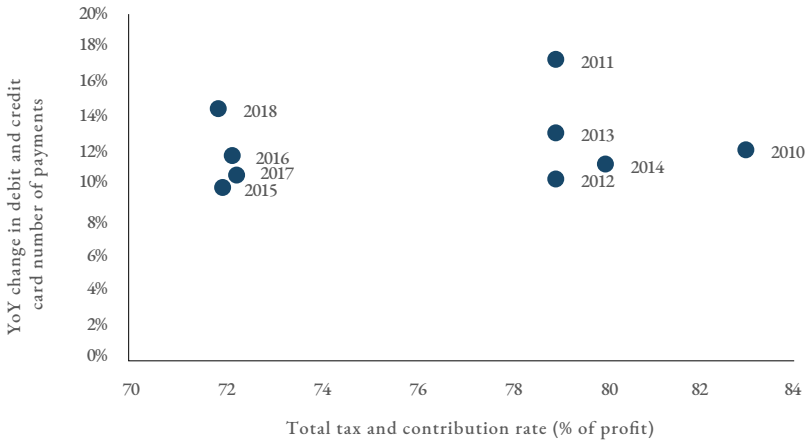
We now test if these factors could be considered as a major binding constraint for enhancing electronic payments. For the case of informality, it is important to note that the perception of informality as the biggest obstacle to growth has decreased with respect to the 2006. This lower negative perception is consistent with the higher formalization in the country, especially in the labor market (IMF, 2019). By definition, if a binding constraint is relaxed, then this will increase the value of the objective function (Hausmann, 2005). However, both the perception and the level of informality have dropped in the country, but the percentage of electronic payments has not experienced a significant upward trend. This is partial evidence that informality is not the most binding constraint behind the high levels of cash usage.

With respect to taxes, we find that despite the reduction of corporate tax rates in recent years[22], Colombian firms still perceive taxes as a major obstacle for their development; meanwhile, the growth rate of electronic payments has not increased sharply. As mentioned previously, if lifting this taxing restriction is not associated with a higher adoption of electronic payments, this is partial evidence that this factor is not the most binding constraint. Figure 5 shows the relationship between total taxes and the annual growth in the number of debit card and credit card payments. Similar tests are not possible for the case of illicit activities due to the lack of information. However, Rogoff (2016) recognizes, reducing the levels of cash in the economy should, to some extent, harm illicit businesses as increases the traceability of transactions. Criminals and tax evaders will find different ways of carrying out their transactions; however, these methods, however, are unlikely to be as safe and liquid as cash.

[18] Rincon and Delgado (2018) show that effective tax rates on firms are significantly lower once tax benefits, elusion, and evasion are considered.
[19] Caballero & Amaya (2011) estimates that cash represent 89.6 % of the total money laundry
[20] In fact, the money laundering from narcotics traffickers based in Colombia and Mexico has been coined as the "the Black-Market Peso Exchange (BMPE)", which involves taking black money to legitimate cash-intensive businesses, and then cooking the books to make it appear that the money was earned legitimately.
[21] According to Asobancaria (2017), while the cost of corruption worldwide is estimated to be represent 2%, in Colombia this figure is between 4-5%.
[22] Law 1819 of 2016 and Law 2010 of 2019 also called Ley de Crecimiento (Growth Law).

Figure 5.

Taxes for businesses and Electronic Payments growth



Source: World Bank Enterprise Surveys and Central Bank of Colombia

To sum up, cash lies at the heart of the underground economy, which is one of the most deep-rooted challenges that Colombia faces. For more than half a century, informality, tax evasion, and illicit markets have become key features of the country's economic and social structure. All these characteristics are well-known to be cash intensive due to its advantage for anonymity and untraceability. Nonetheless, we find no evidence that this is one of the most binding constraints behind expanding electronic payments in the country: when informality and taxes decline, electronic payments usage has not and, in turn, experienced a significant upward trend. Noteworthy, these three issues respond to structural factors that are probably not going to be resolved exclusively by diminishing the use of cash. As expressed by Rogoff (2016), however, reducing cash usage will provide a blow to most of the underground economy as it will allow for greater oversight from the government.

II.D. Financial Inclusion and Payment Infrastructure

Financial inclusion and payment infrastructure are relevant drivers for electronic payment penetration. If consumers lack access to financial products (i.e., bank accounts, debit, or credit cards), cash will be their default payment instrument. Similarly, businesses cannot adopt electronic payments if they suffer from inadequate connectivity access and reliable payment devices.

According to the Economist Intelligence Unit (EIU, 2019), Colombia was the country with the best environment for financial inclusion across 55 developing countries. The country ranks high in government support and consumer protection but lags other peers in payment infrastructure. In particular, the country ranked 18 among 55 developing countries in terms of point-of-sales (POS) terminals. With 66 POS terminals per 10,000 inhabitants, Colombia is behind regional peers such as Costa Rica, Panama, Uruguay, Ecuador, and Paraguay (Figure 6). Brazil, which has roughly a similar GDP per capita than Colombia, has almost four times the level of POS per 10,000 inhabitants. This result is consistent with the informality and tax rate discussion in the previous section, which is associated with low demand for electronic payments from businesses.[23]

Likewise, the country is poorly positioned in the connectivity indicators measured by the EIU, ranking 18 out of 55 countries. For instance, the percentage of households with internet access in Colombia is 55%, significantly below the level of Chile (88%), Argentina (81%), Uruguay (64%), and Brazil (61%). Without the proper complementary connectivity access, firms face significant problems incorporating electronic payments into their day to day transactions.

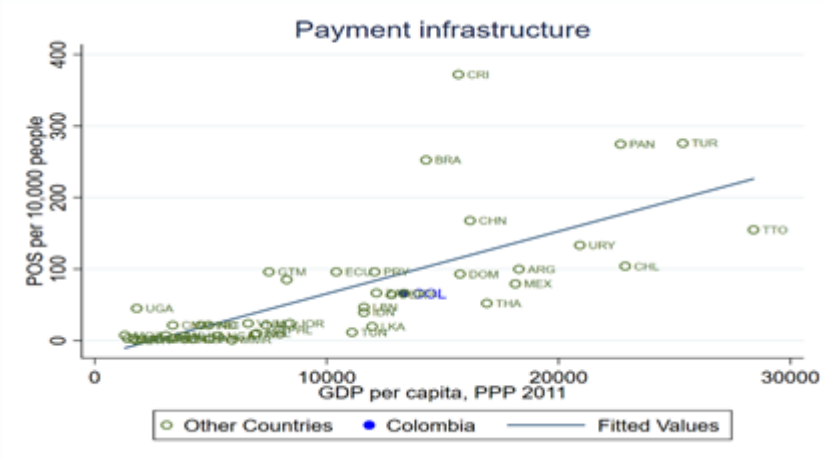
It is important to underscore that POS terminals are also greatly affected by demand-side factors. Colombia exhibits a lower penetration of financial products compared to LAC region, as well as other upper-middle-income countries. The World Bank's Global Findex database for 2017 shows that 46% of people older than 15 have a bank account, 26% debit card, and 14% credit card. While access to deposit accounts and credit cards are close to the LAC level (54% and 18%, respectively), access to debit cards is particularly low in Colombia (45%).

According to the Economist Intelligence Unit (EIU, 2019), Colombia was the country with the best environment for financial inclusion across 55 developing countries.

[23] A possible hypothesis for this result is that firms believe that having POS terminals gives information to the government about their business and thus increases their probability of paying taxes (Arango et al. 2017a).

Figure 6.

Payment Infrastructure: Number of POS (2016) and income levels



Source: World Bank Enterprise Surveys and Economist Intelligence Unit

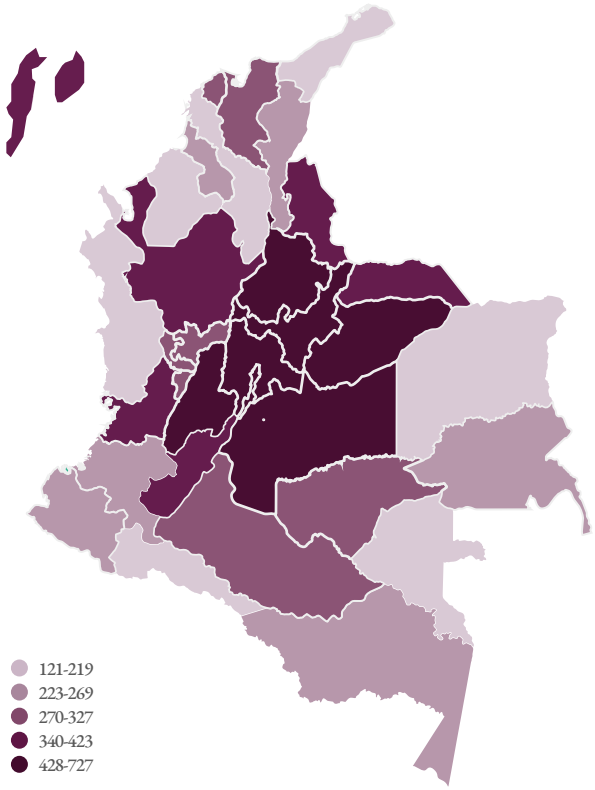
Notably, there is great heterogeneity in access to financial products in the country: in urban areas 89% of the adult population has at least one financial product, while this figure is only 55% in rural areas. Likewise, the coverage of the banking system, measured by the number of offices and banking correspondents, is concentrated in Bogota and the Andean region, whilst peripheral areas face a lower exposure to this critical infrastructure (Figure 7). The lower population density, the lack of provision of public goods by the state, and the high costs of reaching these areas can partly explain the differential access to financial products (Asobancaria, 2017).

Amid the COVID-19 pandemic, payment infrastructure is essential to guarantee the safe deployment of cash transfers and subsidies to protect vulnerable households, workers, and businesses. Even though there is no evidence that the probability of transmission of COVID-19 via banknotes is higher than for other frequently-touched objects, electronic payments and digital technologies are essential to avoid the agglomeration of people and the spread of the disease when they receive these benefits (Auer et al., 2020).

To conclude, Colombia's electronic payment infrastructure is significantly below countries with a similar income level and vis-à-vis regional peers. Both access to POS terminals and internet connectivity are important infrastructure drivers that explain the underdevelopment of electronic payments in the country. On the other hand, except for debit cards, financial product penetration ranks relatively well compared to the Latin American region, but on the intensive margin, there is no evidence that consumers are increasingly using them. Reaching the last mile of consumers in the rural areas of Colombia is a significant challenge for electronic payments development. These regions suffer disproportionately lower access to financial products, payment infrastructure, and connectivity. We believe these are factors where the country exhibits an important lag compared to other countries and reflect an important symptom of the low adoption of electronic payments from businesses.

Figure 7.

Offices and banking correspondents per 100,000 inhabitants



Source: Financial Superintendence of Colombia

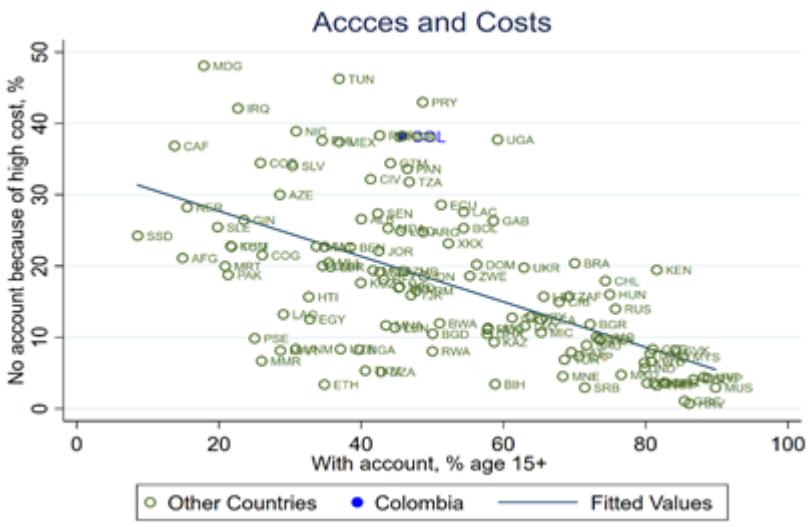
II.E. Costs of Financial Products

High costs of financial products can exclude consumers and businesses from adopting electronic payments in their daily transactions. Figure 8 exhibits how higher bank account penetration is negatively related with the perceived costs of banking across countries. Given its level of bank account penetration, Colombia has a significantly high perceived cost by respondents. The country has the 7th highest financial cost perception across the 123 countries measured in this survey.

Also, the Global Findex database (2017) shows that insufficient funds (41%), and high costs (38%) are the main reasons why people report not having a financial institution account in Colombia (Figure 8). Other reasons such as documentations (18%), lack of trust for the financial system (18%), and distance (14%), appear to be less relevant. If people lack access to a bank account, then their default payment instrument will be cash.

Figure 8.

Access to bank accounts and cost perception



Source: Global Findex Database

High adoption costs are also a potential factor that affect businesses' demand for electronic payments. As stated by the World Bank (2016), having the necessary economies of scale and network externalities to operate on an efficient scale is one of the main problems for payment service providers around the globe; with such a low usage and with high average costs per business, Colombia is currently at a suboptimal equilibrium level. Additionally, small businesses are disproportionately affected by these costs as they have a higher proportion of small-value payments and often lack the minimum monthly balances needed to avoid paying fixed fees.

According to the Economic Commission for Latin America and the Caribbean (ECLAC) Colombia has the highest maximum Merchant Discount Rates (MDR)[24] for both credit and debit cards compared to regional peers such as

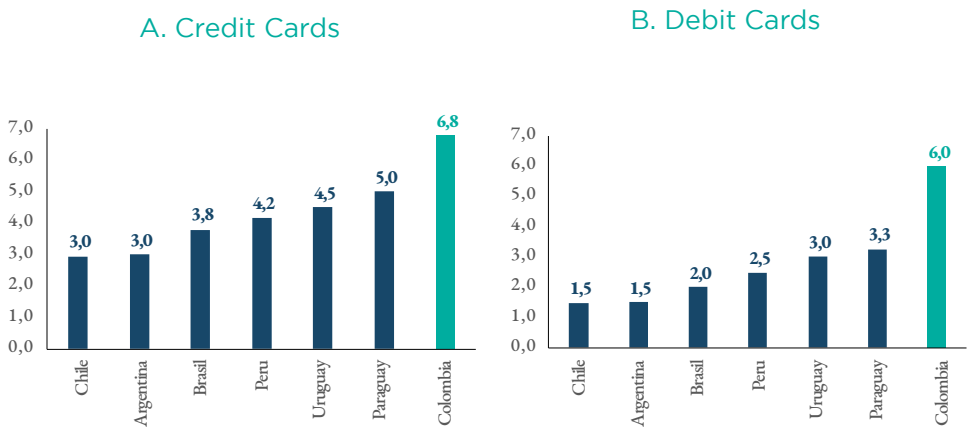
[24] Merchant Discount Rates (MDR) are a variable cost that businesses pay to access electronic payments. This fee is an ad valorem commission paid to acquirers, such as Visa or Mastercard, who process credit or debit card payments.

Argentina, Chile or Brazil (Figure 9). Even though they refer to maximum rates, it embodies an indication that acquirers have a significantly larger space to determine these fees in Colombia.

Furthermore, the low adoption of POS terminals by businesses mentioned in the previous section can be explained by the low competition and high prices in the provision of payment service infrastructure. As discussed in section II.F, Colombia has a network of credit cards and debit cards that is controlled by Redeban, Credibanco, and Visionamos. These organizations also control the technology and system interoperability of the debit and credit card market and can thus exclude new actors from entering with innovative, disrupting, and cheaper solutions. In addition, as recognized by the Daude and Pascal (2015), concentration in the Colombian banking sector has followed an upward trend and is dominated by a few large conglomerates.

Figure 9.

Maximum Merchant Discount Rates (2017)



Source: Economic Commission for Latin America and the Caribbean

Overall, Colombia exhibits a lower penetration in bank accounts and a higher perceived cost of access by consumers compared to most countries around the world. Similarly, businesses in the country face excessive maximum interchange fees that can deter them from adopting electronic payments. The high concentration

in the banking sector and the regulatory barriers to entry which will be discussed in section II.F. can partially explain this phenomenon. We believe that the costs of accessing financial products and payment instruments in Colombia are considerably above other regional and developing country peers and are thus an indication that this is an important binding constraint to tackle.

II.F. Regulatory environment

As previously stated, Colombia has the best environment for financial inclusion across 55 developing countries (EIU,2019). The country ranks first in the dimensions of government and policy support as well as consumer protection, but lags peers in the dimensions of stability and integrity (6th) and products and outlets (8th), which are more regulatory intensive.

As stated by the BIS (2014), historically, the regulatory framework on payment systems was focused on macro-prudential objectives addressing relevant risks and protecting consumers. However, in recent years several countries implemented initiatives focused on the development of the payment industry, through the modernization of the standards of operation, the strengthening of competition, and the promotion of digitalization and immediacy of transactions. In Colombia, this shift in regulatory vision happened in recent years. This point has also been recognized by the Unit of Financial Regulation (URF, for its Spanish acronym) of the Ministry of Finance, which produced a study about the regulatory framework on the low-value payment systems (URF, 2018). We recognize four salient challenges from this report.

First, the limited participation of new players in the payment system. In Colombia, the infrastructure for the processing of electronic transfers consists of two automated clearing houses (ACH for its acronym in Spanish). In one hand, ACH Cenit, owned and administrated by the Central Bank, whose operations represent almost 20% of GDP, from which almost forth fifths comprise treasury payments (Central Bank, 2019). On the other hand, ACH Colombia, owned and administered by the banks, whose credit operations represent 93% of GDP and almost all of them (97%) are made in the real sector (Central Bank, 2019). It is currently composed by 27 active participants - 24 banks, one financial cooperative, one pension fund and a digital platform owned by bank institutions- and the five most relevant establishments concentrate three fourths of the total credit transactions (Central Bank, 2019). This contrasts with the case of Mexico, where 44 non-bank financial institutions participated directly in the automated clearing house, since the authorities directly forced access expansion (see Box 2).

Overall, we observe that ACH Colombia serves most of the low value payments and is used for the day to day transactions, whilst ACH Cenit is mainly used for governmental payment purposes. There exists a concentration of the processing of payments in the largest banks - owners of the private clearing house - and non-financial actors are still inexistent. We can conclude there exist some constrains for new players to access[25] the clearing house systems, core business in this system. In this vein, the EIU (2019) report highlights that Colombia exhibits stricter initial and ongoing requirements to non-bank financial institutions compared to other countries since these institutions face distortions in prices arising from the existence of interest rate caps (usury rate) and taxation of operation. Notably, non-banks can improve the efficiency of the retail payments system by increasing competition, and providing new or improved payment options (BIS, 2014).[26]

[25] For accessing ACH Cenit all institutions must comply with the operating regulations established by the Central Bank, while for accessing ACH Colombia institutions must meet the regulatory access conditions and be accepted by an Admissions Committee.

[26] BIS (2014) highlights that some risk may arise between the initiation of a transaction and its final settlement, including fraud, operational, legal, settlement and systemic risks. However, the report surfaces that potential regulatory differences between banks and non-banks may lead to differences in risk mitigation measures.

In the credit card network, we observe a similar pattern: seven payment system administrators are responsible for processing and routing transactions, as well as clearing and settling card payments. The main administrators are Redeban and Credibanco, owned by bank establishments. Notably, Visa and Mastercard decided to also become low-value payment system administrators in 2018 and thus be able to process, compensate, and liquidate the operations carried out with the cards of their franchises. The penetration of these companies in these activities, however, is still almost null, according to the information of the Central Bank (2019).

Second, the weak implementation of good corporative governance practices that inhibit the best interest among actors of the system and the development of the payment industry. Based on URF (2019), at least five concerns arise in this regard: (i) the conflict of interest within the capital structure of payment system, which is dominated by the banking establishments; (ii) the absence of non-discretionary rules regarding the government-bodies' conformation and decision-making in the financial system; (iii) the limited mechanisms to boost transparency regarding the market characteristics and prices in the payment system[27]; (iv) the shortage of rules regarding the security and stability of the system, as well as explicit policies for the protection of personal data privacy and money laundering; and (v) the absence of mandatory interoperability among all actors in the system.

The penetration of these companies in these activities, however, is still almost null, according to the information of the Central Bank (2019).

Third, as technology innovations generate new ways of providing payment methods, the current regulation becomes obsolete and ignores the new processes that are gaining relevance. Therefore, it is of utmost importance to provide them legal recognition and responsibilities. Likewise, technological innovation triggered complexity along the payment chain, which also requires legal recognition. Some examples are gateways and the "aggregators", new actors in the activities of providing technological tools, supplying equipment, route information, and even taking responsibility for linking businesses[28].

Fourth, the multiplicity of relevant policymakers and the absence of adequate institutional coordination mechanisms hampered the development of an effective supervision function and has led to a shortage of trust and transparency among market participants (Stephanou & Guadamillas, 2008). Specifically, the main governmental actors of the system, such as Central Bank, Superintendence of Finance, the URF, the Superintendence of Industry and Commerce, seek for

[27] While ACH Cenit's pricing policy is publicly available, ACH Colombia does not disclose its prices.

[28] According to the studies of the Colombian Chamber of Electronic Commerce (2017) there are at least 27 online payment processors, which have had significant growth in recent years

their own objectives[29] and there are limited coordination actions to align them.

Noteworthy, some of these concerns have been subject to changes. For example, in the electronic transfer market, ACH Colombia developed the PSE platform, an online payment service that contributed to the expansion of the payment ecosystem. Also, the government created a new type of financial institution, named as Specialized Companies in Electronic Deposits and Payments, which can raise funds from the general public through "electronic deposits". However, as highlighted by the URF (2018), in most cases companies still exhibit difficulties in accessing these systems considering there exist scarce information regarding the access conditions or the requirements defined by the systems correspond to the nature of traditional entities. Regarding corporative practices, private companies have promoted modification in the composition of their governing bodies and transparency upon their pricing and accession policies. Nonetheless, the mandatory execution of these policies is still very limited and depends largely on the willingness of each company for doing it (URF, 2018).

It is also important to underscore the recent efforts of the Ministry Finance, with the support of the main actors of the payment system, to enhance good corporative practices in the payment system and facilitate the identification of new financial developments for supervisory and regulatory authorities. For instance, in December 2019, the Ministry published for public comments a Decree proposal, which included the main new definitions of actors and processes, as well as good corporative practices in the payment system, such as transparency in access and prices, mandatory interoperability, clearer rules for governing bodies and procedures to deal with conflict interest. Furthermore, the entity published in February 2020 a Decree proposal for comments to create a regulatory sandbox - a controlled testing space to experiment with technological development, service providers, and activities of the financial sector.

We recognize that, if approved, these proposals embody a major step forward for financial service providers to develop effective and efficient payment products. We acknowledge at least five potential improvements: (i) the coordination mechanism to enhance oversight could be more clear; (ii) inclusion of economic and competition analysis of the proposed measures; (iii) the complexity of the payments system requires going beyond the definition and think the holistic regulatory framework for all the new actors and processes in the technological progress; (iv) provide clear mechanisms for the two ACH mechanism to compete directly; (v) decrees are subordinated legal tool so is recommended to upgrade it to a law.

Altogether, we can conclude that, despite recent improvements, the regulatory framework still hinders competition and access of new players to the payments system. Although some new actors arriving into the system, Colombia exhibits a high concentration in the processing of payments. We also observe a challenge regarding corporative governance practices that could hinder fair competition between actors of the market and non-arbitrary rulemaking. A third concern that emerges is the capability to define the new actors and processes involved in

[29] The Superintendence of Finance is responsible of the vigilance of the entities that administer and compensation systems (Law 795, 2003); the Financial Regulatory Unit is responsible of the preparation of the regulations for the exercise of the faculty of exchange regulations, monetary and credit (Decree 4172, 2011); the Superintendence of Industry and Commerce ensures the proper functioning of the markets through the surveillance and protection of free economic competition and consumer rights (Decree 4172, 2011); and the Central Bank monitors and participates in the payments system as part of its role in preserving financial stability (Law 964, 2005).

the technological progress. Finally, another issue that persists is the lack of coordination among all the oversight bodies that creates an inefficient provision of the service and spur lack of confidence between stakeholders.

III. Survey on the Use of Cash and Electronic Payments in Latin America and the Caribbean

We surveyed 16 Central Banks in Latin America and the Caribbean (LAC) about the use of cash and electronic payments in their countries.[30] The survey[31] seeks to understand the evolution of the adoption of electronic payments in Latin America and the Caribbean and the predominance of the use of cash in economic transactions.

The results of the questionnaire suggest that data in Latin American is still very limited. Only four of the 16 central banks report the existence of Surveys on Consumer Payment Choice and/or Firm Payment Choice. Regarding the information available, we find that Costa Rica reports that around 73.2% of consumers' payment transaction volumes are made through cash, while in value this figure halves to 34.5%. In Peru, firms recognize that almost three-fourths of the payment transactions are made by cash, while this number in value is 67%. In Chile, this same figure equals 68% for volume.

Notwithstanding, the survey reveals a shortage of data in the region. These results are supported by Diniz et al. (2011), which reviewed a total of 196 papers[32] of mobile means of payment around the world, finding that only one of the peer-reviewed papers was targeted at Latin America. Hence, the authors emphasize the geographic gap in the knowledge of mobile payments derived by this phenomenon. The authors underscore that the minor importance in the region might be related to the reduced ability of banks and regulators to understand regional peculiarities.

Despite the lack of data availability, 12 of 16 central banks acknowledge that during the last ten years, their respective countries have implemented policies to discourage the use of cash. But we do not find measures targeted directly to limit cash transactions (e.g. imposing higher taxes to cash transactions or cash payment limits). We then find a gap between the intentions of policymakers and the vehicle for doing so. It is essential to underscore that just three of the 16 countries reported that they have any target on cash[33] levels on the economy. Thus, it is very challenging to measure the effectiveness of the different policies implemented.

[30] Includes: Argentina, Bahamas, Bolivia, Colombia, Chile, Costa Rica, Dominican Republic, Ecuador, El Salvador, Jamaica, Honduras, Guatemala, Mexico, Paraguay, Peru and Uruguay.

[31] The questionnaire was divided into three parts: (i) databases indicating the levels of cash usage and penetration of electronic payment instruments; (ii) policies or programs implemented to reduce the use of cash and encourage the use of electronic payment; and (iii) evidence that estimate the effect of these policies on the use of different payment instruments.

[32] 94 peer-reviewed and 92 non-peer-reviewed published in the period 2001-2011

[33] Colombia, Jamaica and Guatemala.

We also find that 12 of 16 central banks report that their countries have implemented regulatory licenses for payment service providers

We also find evidence that all surveyed countries, except for Honduras, report having implemented policies to promote electronic payments in the country during the last ten years. All of them recognize that these policies have been aimed to obtain financial inclusion. Reducing informality and increasing transparency are the following most frequent objectives for the implementation of these policies. We also find that 12 of 16 central banks report that their countries have implemented regulatory licenses for payment service providers, while 10 acknowledge direct support in the initial provision or improvement of the payment system. No country surveyed recognizes an existent target on electronic payments. This contrasts with other experiences worldwide where concrete electronic payments goals have been imposed[34]. Notably, 12 of the 16 countries are members of the Alliance for Financial Inclusion (AFI)[35] and commit to the Maya Declaration[36], which is a world platform to make concrete financial inclusion targets, implement in-country policy changes, and regularly share progress updates.

Central banks surveyed reported 55 policies implemented in the region. All of them summarized in Table 1. Although there exist heterogeneous results upon the strategies and policies revealed by each of them, we highlight some commonalities. First, 10 of 16 central banks recognize the implementation of national financial inclusion strategies as mechanisms to increase digital payments in their countries. While in El Salvador and Uruguay, this strategy has been converted into law, most of the other countries - such as Colombia, Paraguay, Guatemala, or Jamaica- use these strategies as guidelines to implement principles, prioritize topics, and propose action plans. Furthermore, 9 of 16 central banks recognize that government offices have moved to electronic payment mechanisms. For instance, Dominican Republic, Uruguay, and Peru have carried out programs for payments to state providers through bank account payment. Another way in which governments have encouraged the use of digital payments is by the distribution of social programs through accounts with the financial system. As examples, Colombia, Mexico, and Costa Rica nowadays distribute most social subsidies via financial accounts.

[34] For example, the Rwandan government, with the National Payment System Framework and Strategy 2018 – 2024, already established the goal of becoming a cashless economy – targeting for 2024 that at least 80% of adult population will have access to digital payment methods and that the value of electronic payments to GDP will be over 80%.

[35] Alliance for Financial Inclusion is a policy leadership alliance, including members from more than 90 countries, working together to accelerate the adoption of proven and innovative financial inclusion policy solutions with the goal of making financial services more accessible to the world's unbanked (AFI, 2020).

[36] The Maya Declaration is a global initiative for financial inclusion that aims to reduce poverty and ensure financial stability, while determines measurable set of financial inclusion commitments by developing and emerging economies (AFI, 2020).

Additionally, 7 of 16 respondents report having made recent reforms to modernize the electronic retail payment systems (e.g. clearing houses system) and instruments to provide more efficient payment services, alongside the development of guidelines for using it[37]. In this same vein, almost all surveyed countries have taken steps for the introduction of electronic money (e-money) by enacting technical norms and guidelines for the constitution of the companies offering these services.[38] Roa et al. (2017) highlight the growing importance of the service of e-money in the region and its possible role as a propeller of financial inclusion. The author found 43 services of electronic money on 26 LAC countries. Most of these services allow transfers between users, balance inquiry, deposits, withdrawals, purchase credit, and public service payments, among others. More innovative systems allow paying for public transportation, paying for taxis, receiving international remittances, and receiving program subsidies of conditional cash transfers.

Table 1.

Overview of digital payment measures per country[39]

Country	Cutting of red tape	Demone-tization	Enhance competition	Fees Abolishment or Reduction	Financial Inclusion Strategy	Government digitalization	Improve Infrastructure	Mandatory Electronic Payment	Promoting Electronic Money	Tax incentives
ARGENTINA			■					■		■
BAHAMAS		■					■		■	
BOLIVIA				■					■	
CHILE			■							
COLOMBIA	■		■		■	■	■		■	■
COSTA RICA	■	■			■	■		■	■	■
DOMINICAN REPUBLIC						■	■	■	■	
ECUADOR				■			■		■	
EL SALVADOR					■				■	
JAMAICA					■				■	
HONDURAS									■	
GUATEMALA					■					
MEXICO	■		■	■	■	■	■	■	■	■
PARAGUAY					■		■		■	
PERU			■		■	■	■		■	
URUGUAY					■			■		■

Source: Own elaboration based on Survey, AT Kearney (2015) & Roa et al. (2017).

[37] For example, in 2016 Honduras updated the technological platform of the clearinghouse for electronic payment transactions that allow the accreditation of funds in real-time as well as the confirmation of the reception of the funds immediately.
[38] The European Central Bank defines e-money as an electronic store of monetary value on a technological device that may be widely used for making payments to entities other than the e-money issuer. For more information see Appendix II.
[39] Appendix II provide the definitions for the different policies reported.

Direct interventions, such as tax incentives or refunds for digital payments, have been scarcer in the countries analyzed. For instance, Colombia imposed a return of 2 percentage points of value-added tax (VAT) for purchases made with debit or credit cards, but this measure was eliminated in the year 2014. Similarly, Honduras has 8% refund of sales tax to purchases made by debit card or credit card. Also, Costa Rica used enforcing measures by requiring by law that all businesses should offer the possibility of electronic payments and not only cash. On the other hand, Mexico established public subsidies for promoting the acceptance of cards and other electronic payment instruments by merchants. As these measures were recently implemented, estimations of these impacts are still very limited, yet it is recognized to be a non-negligible cost for the government budget. Argentina and Uruguay, by contrast, used a carrot and stick system: all businesses offering providing consumption services are obliged to accept digital payments, but the cost of implementing the system may be discounted from the VAT.

Price regulations by reducing or eliminating digital transactional fees are another uncommon mechanism of intervention. In 2018, Ecuador directly reduced the cost of transactions of almost 23 financial services by an average of 21%. Similarly, the government of Bolivia required free electronic services for a limit amount of transfer funds (to Bs50,000 or its equivalent) between accounts of the same entity, online service payments, and tax payments. In Argentina, the Central Bank scheduled withdrawal of maximum interchange fees applicable to transactions with debit, credit and purchase cards since 2017 (from 1% in 2017 to 0.6% in 2021 in debit exchanges and from 2% to 1.3% in credit exchanges).

One of the most recent measures has been the incorporation of a quick response code (QR by its acronym in English), which allows card payments without POS terminals. Peru and Mexico are currently implementing a standard QR code, irrespective of the type of transaction, open to any electronic wallet, and with lower membership costs than those of the POS terminals. Another innovation has been the publication of data for the general public. For instance, Ecuador created a statistics website on electronic payment methods available for all the citizens and people in the industry. Notably, Costa Rica has presented a large-scale reform which put together most of the programs mentioned above (See Box 1).

Another policy that governments in the region have tried is enhancing competition in the low-value payment system and including non-banking actors to operate. Box 2 provides the example of Mexico in this direction. Also, Chile focused on increasing competition in the credit card market. The Central Bank of Chile followed recommendations of the Court of Defense of Free Competition, and in 2016 authorized the issuance and operation of means of payment with the provision of funds or any other similar system by non-bank companies.

To sum up, the survey revealed a shortage of information in Latin American countries regarding the cash and electronic payments usage. This issue could also be related to the reduced ability or interest of banks and regulators to understand regional peculiarities concerning the system. Despite the shortage of data availability, there is a generalized implementation of policies for enhancing the use of electronic payments in the region, mainly focused on obtaining financial inclusion, reducing informality, and increasing transparency. Specific targets for these policies are also absent in the analyzed sample. Deeming into specific interventions, most of the countries acknowledge to have implemented a financial inclusion strategy and digital payments in the governmental transactions -mainly for the distribution of social programs-, alongside with the modernization of the electronic retail payment systems and the introduction of e-money in the economy. Although most of the Central Banks recognize they seek to discourage the use of cash, policies are primarily focused on promoting electronic payment systems. Direct intervention, such as price regulation or fiscal incentives, seem to be unusual in the region. Recent measures consist of the implementation of QR, which allows card payments without the need for POS terminals. Finally, we observe a growing interest in these countries to promote policies that enhance competition in the payment system and allow non-financial firms to enter the market.

Box 1. Costa Rica efforts to become a less cash society

Costa Rica implemented several reforms to promote electronic payment and reduce cash usage. The Central Bank, in coordination with the financial sector, has played a key-role in this agenda. First, in the year 2015 they created a free electronic payment services ("*Sinpe Mobile*"), which allows transfers of low amounts from any mobile phone that is linked to a bank account. Second, they created the Simplified File Account (*"Cuenta de Expendiente Simplificada"*) which simplifies the access of fund account at any bank agency, office or correspondent for unbanked customers with low risk profile. Thirdly, for 2020 they are implementing a program of demonetization of coins of lowest denomination. They also implemented the program *"Libre de Efectivo"* which certifies the private and public entities that, among other conditions, demonstrate that 90% of their payments are made completely electronically. This has been accompanied by governmental initiatives from the executive and legislative branches. For instance, the government ordered creating bank accounts to all the beneficiaries of social programs. Furthermore, the congress approved a Law in 2018 to enhance public finances which stipulates the refund of the sales tax to medical services that are paid by cards (debit and credit) and require all POS to offer the possibility of electronic payments.

Box 2. Mexico's reform in the payment system

Competition on retail payment system has become an utmost topic of the lasts governments of Mexico through the inclusion of non-bank entrance in the low payment system. To open access to new players in the electronic transfer payment system authorities decided to directly force access expansion. In 2004, the Central Bank replaced the Extended Use Electronic Payment System (SPEUA) with the Interbank Electronic Payment System (SPEI), with the aim of facilitating interbank payments and enabling low value payment services. Then, in 2006, the Central Bank allowed a large group of non-bank financial entities as direct participants to the system, supervised by the Mexican financial authorities, who previously had no access. In 2011 the government allowed banks to establish schemes facilitating financial inclusion via cooperation with non-banks. This regulation also allowed users to open low-risk banking accounts remotely on the provision of basic identification information. Also, banks were authorized to establish business relations with a banking agent (such as a retailer or store), so that the latter is authorized to offer financial services to its customers on behalf of the bank. In addition, some mobile payment services in Mexico are offered by banks in cooperation with mobile phone operators, which play a fundamental role in providing the necessary communication services between account holders and their banks.

Source: Own elaboration based on Survey, BIS (2014) and URF (2018).

IV. Assessment of Policy Alternatives

In 2006, the Government of Colombia implemented a financial inclusion strategy and investment program named "Banca de las Oportunidades" aimed at promoting access to credit and other financial services to the population - mainly low-income families, Small and Medium Enterprises (SMEs) and entrepreneurs. Since then, these strategies have been at the center of the policymaking arena evidenced by the fact that all the Presidential Development Plans have included actions towards financial inclusion and education. In addition, the government created two intersectoral high-level committees and endorsed the Maya Declaration from the Alliance for Financial Inclusion - AFI[40] (Appendix III summarizes all main national governmental initiatives related to electronic payments).

Despite good intentions, Colombia's economy still exhibits major binding constrains that maintain the country's cash dependency. In this context, we acknowledge that rather than going for many targets at once the government would be more effective by prioritizing reforms that yield the largest benefit. Thus, there is a preference for measures targeted on the most binding constraints for the cash dependency in the economy. Moreover, we are cognizant that these general principles require considerable knowledge of local specificities, considering the political and administrative limitations that the government usually faces. Therefore, we use Pritchett et al. framework (2005), which argues that the most achievable gains for a policy design consist of the intersection of the technical correctness[41], administrative feasibility[42] and political supportability[43].

We assess five governmental initiatives as possible solutions for enabling electronic payments in Colombia. These policies were highlighted regularly in our survey for Central Banks in Latin America, the policy debates within the country, and are closely associated to the revealed binding constraints. We provide the summary of each technical assessment (see the detailed assessment for each proposal in the Appendix IV).

Option 1: Eliminating the tax on financial transactions (GMF)

A possible option would be eliminating the tax on financial transactions (GMF), but we consider this is an implausible proposal. First, we underscore that seeking for a new tax substitute is warranted as the GMF nowadays represents one of the primary sources of public funding and the government is legally forced to accomplish public balance targets based on the fiscal rule[44]. Despite the fact that this fiscal rule has been eliminated for 2020 and 2021, the government will require

In 2006, the Government of Colombia implemented a financial inclusion strategy and investment program named "Banca de las Oportunidades"

[40] The Central Bank of Colombia and Banca de las Oportunidades became members of this Alliance.
[41] Technical correctness analyzes the expected effectiveness of the measure removing the binding constraints that enhance cash dependency on the economy, as well as the cost-benefit assessment of the proposal.
[42] Political support asses the approval level among main stakeholders regarding the policy proposal, including government agencies, central bank, bank guilds, businesses and civil society representatives.
[43] Administrative feasibility evaluates the institutional and human capacity (mainly in government) to implement the policy.
[44] See Appendix IV.

this fiscal resources in the medium term. The Ministry of Finance (2015), for instance, estimated that the eventual abolition of the GMF would require around a three-percentage point increase of the VAT. Currently, it is not clear which tax could replace it. For example, the VAT was already increased in the year 2016 and the effective tax rate on firms is estimated to be one of the highest in the region (section II.C).

The technical correctness of this measure is considered as moderate. Although the financial transaction tax displayed undesired results on banking disintermediation and hinders the willingness of people and firms to access the financial system, it does not characterize the most binding constraint for higher preference of cash nowadays (see Section II). Furthermore, the political support of this proposal seems very limited. This tax represents a significant revenue (0.7% of the GDP for 2019) and ranks as the third most important source of funding for the Central Government. Notably, in 2015 the Ministry of Finance conformed an Expert Committee for Equity and Tax Competitiveness (EXET) to seek out for the redesign of the tax system of the country. In this regard, the Committee concluded, *"the Commission understands that the GMF has negative effects on financial intermediation. However, it is an easy collection tax that generates significant resources to the Treasury"* (Bonilla et al., 2015). Therefore, eliminating the tax becomes an undesirable topic for the government as it creates the additional challenge of replacing it. This becomes even harder in the actual situation of fiscal austerity after the drop in oil prices of 2014[45].

At first glance, eliminating the financial transaction tax could be an administrative relief for the tax office as the government could stop its collection. However, this proposal reveals muted administrative feasibility as the GMF, well-known for its relatively easy collection[46], would be replaced by a new tax that with high likelihood will require more laborious administrative procedures. This gains relevance in the context of Colombia, where tax administration is relatively weak and where the market exhibits high levels of informality. Overall, this is a policy we would not prioritize or undertake.

Option 2: Gradually phasing-out paper currency by removing the high-value notes first

Following Rogoff's (2016) recommendation for advanced economies, this proposal aims to create a strategy to phase out paper currency economy for curving the underground economy and fighting against some of the deep-rooted problems of Colombia. Cash is an important enabler of the informal economy and tax evasion as it facilitates transactions that sidestep government oversight. Also, money laundering, mainly done with cash, is estimated to equal between 2 and 3 percent of GDP (Caballero & Amaya, 2011). This is mainly done for drug trafficking purposes, but corruption, crime, and other illegal business are also relevant features for this black money market. As highlighted by Rogoff (2016), this phasing out process should be made in a slow and deliberate pace and with ample anticipation to allow households to adopt other means of payment. This strategy should begin by removing the high-value notes, which in the case of Colombia, are the 100.000 COP bill notes. In this regard, the author highlights the example of the multipronged approach of the Nordic countries, in which the expirations of the bills are announced with years in advance, tandem with a significant promotion of electronic payments.

[45] Before the shock, hydrocarbons represented almost 17% of total fiscal income (Ministry of Finance, 2015)
[46] Matheson (2011) analyzed financial transaction taxes in the G-20 and concluded that the ease of collecting such a tax is frequently cited as a reason to adopt it.

From a technical perspective, this strategy appears to be appropriate considering that the underground economy represents a great share of the economy. We acknowledge that limiting cash will not end these activities, but it will definitely disturb their businesses model. However, the highest bill notes of the economy, 100.000 COP ($32) bill, is way below average of countries with similar income level ($58) and is aligned with the evolution of wages and the general level of prices (Fedesarrollo, 2016)[47]. Moreover, the penetration of this note bill in the economy is still considerably limited[48]. The recent experience of India[49] shows that, although it appears to exhibit a positive outcome for the long run, the trade off in the short term is a deceleration in the economic activity - particularly for informal and rural households[50]. Some of these undesired outcomes may be mitigated by a gradual implementation with slow and deliberate pace.

Regarding political support, former experiences reveal clear discrepancies among main stakeholders. When the 100.000 COP bill was introduced in 2016, the Central Bank underscored the benefits and pertinence of issuing this bill note, while Asobancaria, the Colombian Banking Association, strongly argued against it. It is also important to consider the vast discontent that spurred as a consequence of this policy in the case of India; if Colombia follows suit, civil discontent could emerge and the political cost for the government of promoting this policy could thereby be non-negligible.

Finally, this proposal implies a relevant expense for the central bank associated with the cost of destroying the high bill notes and issuing low denominations that are required to offset the demand for currency. Also, the administrative difficulties could stem from civil society: citizens should change their bills rapidly, while banks should quickly change the currency.[51] To sum up, we do not believe this is an option to prioritize as the marginal costs of social unrest and political opposition are likely to be greater than the marginal benefits of its implementation. On the other hand, Colombia's 100.000 COP bill lies significantly below the highest bill value of other countries with similar income per capita.

[47] According to Villar et al (2016), in the absence of the 100,000 notes, around 13 notes would be required for the minimum wage, and 18 for the basic basket, amounts well above those observed when changes were made to the highest denomination notes in the last quarter century.
[48] According to the Central Bank Statistics, it corresponds to 5.6% of the total value of banknotes in circulation and 0.5% of GDP (Central Bank, 2019).
[49] On November 8, 2016, the government unexpectedly declared the highest notes bills - which represented 86% of the existing currency in circulation - illegal tender, effective at midnight. For more information see Appendix IV.
[50] Chodrow-Reich et al. (2020) showed that districts experiencing more severe demonetization exhibited lower economic activity and lower bank credit growth.
[51] In India, traumatic administrative procedures were observed at the beginning of the demonetization. People seeking to exchange their banknotes had to stand in lengthy queues, and several deaths were linked to exchange cash.

Option 3: Creating a strategy promoting firms' appropriability of financial services

The government should promote a strategy to incentivize firm's acceptance and appropriability of electronic payments. This strategy should consist of two components. On one hand, the government should provide tax incentives to firms to use financial service products by increasing the tax on financial transactions deductibility on VAT from 50% to 100% to all firms. Ideally, following the carrot and stick model applied in other LAC countries (Costa Rica, Uruguay and Argentina), this should be coupled with a gradual requirement that all businesses offer the possibility of electronic payments, yet is complicated with actual legal framework which limits the inherence of the state in any private initiatives.

On the other hand, the government should create a financial education program to defeat the firms' lack of knowledge and low confidence in the financial system. We agree that this proposal should be focused on Small and Medium Enterprises (SMEs) and informal firms, by producing massive diffusion of the campaign, creating a holistic strategy within the government, and involving representatives of different sectors in its implementation. This initiative is aligned to a recent proposal published by the government[52], focused on the interinstitutional coordination and guidelines of the phenomenon. We should also bear in mind the experience of Netherlands, which included a public campaign in 2007, and consisted of several nationwide interventions and some regional ones clustered in time. Jonker et al (2017) demonstrated that these interventions aimed at increasing debit card acceptance by retailers were effective, specifically providing information to retailers about the benefits of cards acceptance led to higher card acceptance.

This proposal appears to be technically sound as it is a very cost-effective measure to face one of the most binding constraints for the intensive cash usage

[52] In February 2020, the Government, led by the Ministry of Planning, published for comments a draft of a white paper for financial inclusion that also highlights the importance of implementing financial education programs for firms (DNP, 2020).

in the economy. The fiscal cost for the first proposal is estimated at 0.09% of GDP (EXET, 2015), while the benefits for society appear substantial: 92% of firms are considered SME and it is estimated that only 11% accept these payment methods[53]. The first component showcase to be easily implementable as the VAT discount from the tax on financial transactions is already in place. Regarding the second component, during the last decade, the government has heavily invested in a strategy for implementing financial education in the country but mainly for citizens. The challenges escalate as the government needs to change the scope of this program to firms, and more importantly include informal firms which are harder to reach. Independently of the timing, reaching the last-mile user implies a very tough task in the social and geographic context of Colombia.

It is worth noting that meaningful conflict among main stakeholders is far-sighted with this proposal. Although the government and bank guild have shown agreement in the need of improving financial inclusion strategies, firms exhibit a lack of confidence in the financial system and governmental intervention. Lack of interest or even resistance to participation is another obstacle that is costly for financial education to overcome[54]. Thus, a massive awareness campaign is warranted to obtain the support of the civil society and restore their confidence.

Option 4: Regulatory interventions enhancing competition and access of new players

The proposal aims to generate four regulatory adjustments to foster competition and access of new players on low-value payment infrastructure. First, strengthening competition between the two ACH platforms by guaranteeing liberalized interchange fees in both markets and public disclosure of prices so that end-users can choose between them. Thus, authorities should take a stronger pro-competitive stance and rigorously monitor enforcement of non-discriminatory business practices and adequate public disclosure. Second, promote good corporative governance practices in all the payment system by creating clear public disclosures and arrangement of the operating procedures (e.g., shareholder structure, decision-making mechanisms, pricing, access policies, and interoperability). Thirdly, considering the technological progress and innovation in the financial services sector, it is essential to define and delimit the regulation for new actors and processes involved in the payment system. One important step in this direction was the government's proposal of creating a regulatory sandbox and promoting new definitions and good corporative practices, which, if approved, should facilitate a new agenda of technological transformation in the payment system (See Section II). All these measures must be accompanied with institutional coordination mechanisms among the institutions that supervise and oversee the payment system (i.e. Central Bank, URF, and Superintendence of Finance, etc.). It is therefore critical to define a legal framework on payment systems provision – independent of the type of entity – that describes the objectives, participants, and coordination mechanisms in this field.

There exists extensive literature studying the determinants of payment systems and the market structure. This market is generally analyzed as a two-side market payment with two heterogenous groups of consumers: cardholders and merchants/businesses. Therefore, theoretically, interchange fees[55] could play a critical role producing efficient prices to both sides of the market (Rochet & Tirole, 2003; Baxter, 1983). It is assumed that the substitution between cash and payment cards can only occur if consumers and merchants are provided with the appropriate incentives, through the interchange fee, to use and accept cards, respectively. The question is then whether competition in this market can

[53] Ideally, this should be coupled with a gradual enforce of firms for implementing this infrastructure, yet this appear to be misaligned with the actual legal framework. In case this restriction is maintained, the government should consider the public provision of electronic infrastructure to firms.

[54] Microenterprises have a negative perception of surveillance of banks by the authorities: 43.4% of microenterprises consider that the authorities' surveillance of banks is bad or very bad (Banco de las Oportunidades, 2017). Also, 30.8% of the population have very low confidence in the bank's operations to improve the quality life of the citizens (Latinobarimetro, 2018).

[55] Fees that banks that service merchants (acquirers) pay to banks that issue cards (issuers) with respect to transactions between their respective customers.

create the conditions for an efficient interchange fee. In this regard, theorical and empirical evidence confirms that there is no standard answer on the most effective competitive policy that would ensure dynamic efficiency. Much of it depends on initial conditions and the broader institutional setup (Rochet and Tirole, 2003; Weiner and Wright, 2005; Guthrie and Wright, 2005; Chakravorti and Roson, 2006).

As a result of the uniqueness of the market structure and institutional environments of Colombia, it is difficult to establish an appropriate benchmark to assess the country's competition levels. However, Section II.F. demonstrated that the current conditions hinder competition and access of new players to the payments system through the discriminatory or arbitrary application of access requirements, and lack of transparency in access fees and the methodology of their calculation. Also, Stephanou & Guadamillas (2008) underscore two main concerns that hindered Colombia's payment market structure and are still present to date. First, the presence of two ACH platforms increased contestability, but some discriminatory business practices inhibited direct competition. Notably, the authors acknowledge that the coexistence of two ACHs produce adverse effects on the economies of scale and scope of this market[56]. Second, the multiplicity of relevant policymakers and the absence of adequate institutional coordination mechanisms have hampered the development of an effective supervision function; the complex oversight framework has led to a lack of trust and transparency among market participants.

As a result of the uniqueness of the market structure and institutional environments of Colombia, it is difficult to establish an appropriate benchmark to assess the country's competition levels.

In this context, these measures would generate lower operational costs and thus better pricing for end users as a result of stronger incentives to become more efficient as well as greater product innovation and access (including from non-bank financial institutions) stemming from increased competition (URF, 2018). Also, enhancing transparency in the functioning of the payment market would be useful to overcome mistrust and further promote competition. Additionally, a stronger supervision framework would prevent potential regulatory gaps and promote a comprehensive approach to developing a more efficient and accessible electronic payments systems infrastructure.

The political landscape shows a superficial agreement among main stakeholders. Despite the broad superficial consensus, the outcome could negatively affect the network profits of the banking system, which could in turn generate rejection from them[57]. This effect could be exacerbated by the prevalent lack of trust between them. It is also worth noting that the complex oversight framework

[56] According to Stephanou & Guadamillas (2008). given the significant fixed costs reported by both ACHs, there are inefficiencies from having a duplication of infrastructures.Also, ACH Colombia has greater economies of scale due to its size, yet differences in ownership structure across payments systems have prevented it from clearing other payments instruments (scope).

[57] Chakravorti and Roson (2006) demonstrate increasing the competition necessarily reduces the profits of payment networks.

has led to a lack of trust and transparency among market participants. In this context, a major element of this framework is the establishment of institutional mechanisms to promote coordination and information sharing between relevant public and private sector participants[58]. Without it, political constraints could raise up quickly and obstruct the implementation of this policy.

In contrast, the administrative costs seem to be marginal. This policy proposal expects market competition to be the main driver that enables the use of electronic payments in the country. Therefore, government interventions should be limited in defining the rules and actors in the payments system. The administrative cost from the government would rely primarily on the supervision and revision of the payment system guaranteeing its stability. However, if the regulation is clear enough, this cost should be relatively negligible.

A relevant feature about this topic is the need of clearly defining the regulatory framework since it involves many different governmental offices working at the same time[59].

Option 5: Implementing a cap level to the interchange fees

Following the examples of different countries (e.g., Australia, Mexico, Spain, and Europe), the government may intervene by capping the level of interchange fees that are collected by the issuers in payment systems. In this regard, they should determine an appropriate regulated value for the interchange fee equal to the efficient level (i.e., the level that internalizes externalities between the parties to a transaction). Technically speaking, this proposal seems reasonable as this reduction in the interchange rate may in turn decrease merchant discount rates. As a result, adoption of card payments by businesses would likely increase, and this effect would be higher in a context of low electronic payments adoption such as Colombia.

However, determining an appropriate technical value for the interchange fee is challenging. As recognized by the URF (2018), setting limits on these rates could lead to costs being transferred to other customers of the system or charged to other products offered to the same users. This phenomenon occurred in the European Union with the implementation of the Interchange fees for Consumer Card-Based Payments (IFR). On the card user side of the market, card user fees (rewards) would be expected to rise (fall), thereby making the card less attractive to consumers, compared with other payment methods (Praguer et al, 2009). Consequently, the final effect would depend on the balance between the consumer substitution effect away from cards and the impact on the increased business acceptance. Notably, calculation of that fee requires knowledge of social costs and benefits that are difficult to measure accurately. Also, the determination

[58] The case of Mexico has shown the effectiveness of having this coordination tables on this regard.

[59] The Superintendence of Finance is responsible of the vigilance and inspection of the entities that administer credit or debit card systems as well as those that administer payment and compensation systems. The Financial Regulatory Unit is responsible of the preparation of the regulations for the exercise of the faculty of exchange regulations, monetary and credit and regulatory and intervention powers in financial activities, stock market, and insurance. The Superintendence of Industry and Commerce ensures the proper functioning of the markets through the surveillance and protection of free economic competition and consumer rights. Furthermore, Central Bank monitors and participates in the payments system as part of its role in preserving financial stability.

of which costs should be included in a cost-based fee is necessarily arbitrary, and measuring those costs is nontrivial, particularly if frequent re-estimation is necessary.

The political environment appears to be more complicated for this policy. Imposing tariffs will affect the business of bank establishments. Thus, a clear rejection of this proposal by this group is sighted. In fact, Asobancaria (2016) showed strong opposition to this type of proposal advocating for the promotion of competition as the best regulator of financial costs. The government does not appear to have strong support for this type of policies either. In fact, the URF (2018) also agrees that the benefits of this measure are not clear still and this change could be transferred to other customers or products of the system. Besides, the legal framework determined that financial institutions enjoy autonomy to establish at their discretion the services they offer to their clients and the costs they charge. In any case, the law establishes that the rates must be objective and non-discriminatory.

One well-recognized advantage of this policy approach is that once the regulated values for the interchange fee have been chosen, implementation of the policy is straightforward and transparent. It appears to be an administratively feasible measure from the government's perspective. Also, adjusting to a flat price could facilitate companies' operations management. If they decide to include different rules to determine the price seeking for a less arbitrary level, the degree of administrative requirements would be increasing. Nonetheless, in all scenarios it appears to represent negligible changes in this regard.

Summary:

After carrying out this assessment, two policy alternatives emerge as the most suitable in terms of being technical correct, politically supportable, and administratively feasible. These options are encouraging firms to appropriate electronic payments and a regulatory intervention to enhance competition. The following evaluations supported the decision for the two policies. Table 2 summarizes the results of this analysis.

Table 2.

Summary Assessment Analysis Policy Proposals

Solution	Technically Correct	Politically Supportable	Administratively Feasible
1.Eliminate the tax on the financial transactions	Medium	Low	Medium
2.Gradually phasing-out cash currency	Medium	Low	Medium
3.Encourage firms to appropriate electronic payments	High	Medium	Medium
4.Regulatory intervention to enhance competition	High	Medium	High
5.Impose some tariffs for financial intermediation	Medium	Low	High

V. Conclusion and Policy Implementation

Despite the development of electronic payments over the last decade and the numerous governmental strategies to promote financial inclusion and education, Colombia stands as one of the most cash-intensive societies worldwide. We find two main binding constraints for Colombia's high cash dependency: an obsolete and outdated regulatory framework and firms' negative cost-benefit evaluation of accepting electronic payments. Our survey, conducted with 16 central banks in Latin America, demonstrates that there exists a shortage of data availability in the region. However, there is a generalized implementation of policies to increase the use of electronic payments targeted on obtaining financial inclusion, reducing informality, and increasing transparency. There has been a growing interest in these countries to promote policies that enhance competition in the payment system and allow non-financial firms to enter the market.

In this context, we evaluate five policy solutions to enable electronic payments in Colombia and reduce the current levels of cash usage, assessing the technical correctness, political supportability, and administrative feasibility of the reforms. Our findings suggest that two central policies could improve the dynamism of electronic payments in the country. A first approach relies on promoting a strategy to incentivize the acceptance of electronic payments by firms. This strategy should consist of providing tax incentives to firms to use financial service products. At the same time, they also need to create a financial education program to defeat the firm's lack of knowledge and low confidence in the financial system and its potential benefits.

The second proposal aims to generate four regulatory adjustments to promote competition and access of new players: (i) strengthen competition between the two ACH platforms; (ii) promote corporative governance practices in all the payment system by creating public disclosures and arrangement of the operating procedures; (iii) define and delimit responsibilities of the new actors and processes consistent with the technological progress of the payment system; (iv) strengthening oversight arrangements, mainly via the establishment of robust institutional coordination mechanisms.

These proposals must be of utmost importance for the government as they will reduce the high level of cash usage in Colombia, which is closely associated with the deep-rooted challenges such as informality, illegality, and tax-evasion. Besides, boosting electronic payments will improve the security and transparency of transactions, as well as enhancing financial inclusion and poverty reduction.

These proposals must be of utmost importance for the government as they will reduce the high level of cash usage in Colombia, which is closely associated with the deep-rooted challenges such as informality, illegality, and tax-evasion.

As described by Prittchet, Woolcock & Andrews (2013), despite the enormous effort and engagement in ‘reforms’, implementation failure is one of the critical flaws of developing countries, given the complexity of each context and lack of administrative capabilities. Academics and even governmental offices have mentioned and advocated for some of these proposals. Nonetheless, the pace and

determination to kickstart these policies have been languid, reducing their likelihood of being implemented. Therefore, we shed some light on key features that these two reforms should include in their management method to build multi-method capabilities and obtain effective public policy implementation.

The first proposal exhibits low conflict, as most stakeholders seem to agree on this issue. Nonetheless, it displays a high level of ambiguity since it requires to reach last-mile users in remote areas of the country. Also, providing incentives and proper financial education campaign implies an important learning process for the different stakeholders involved. In contrast, our second policy recommendation, which focuses on regulatory reforms, exhibits lower ambiguity. In fact, the URF (2019) already established some of the essential components that this reform should include, and our Central Banks survey provides some insights into the first steps taken in some Latin American countries. The challenges of this second proposal lies in the high-pitched conflict nature of the regulatory reforms. There is more uncertainty regarding if all stakeholders, and mainly bank establishments, will support the implementation of this policy, bearing in mind that they will probably lose market power.

According to Andrews (2019) different challenges require different approaches. For instance, for the cases where there exist high ambiguity and low conflict the author recommends an experimental implementation, where contextual conditions are critical. By contrast, for the cases where there exists low ambiguity but high conflict a political approach is warranted – prioritizing power dynamics. Also, the type of problem should also be connected to the management method applied to implement each policy[60].

To guarantee the effectiveness of these policies, the government should consider conducting a countrywide communications campaign to raise awareness among the business community and the general public about the benefits of embracing electronic payment facilities. It is also critical to bring the banking establishments and the government together to organize the regulatory landscape of the payment system. If these proposals are unsuccessful, informality, tax evasion, and illegality will continue to be a rampant feature of the Colombian society, while economic development and financial inclusion will remain stagnated. Finally, in a highly unequal country like Colombia, these proposals are critical as they will guarantee access to the last mile user –specifically in the regions abandoned by the state.

[60] For the first solution where the technical solution is somehow uncertain an adaptive approach is more appropriate. Some items that could be include in doing so is flexibly fusing design, stages multiple short iterations, learning by doing, among others. As for the second proposal where the problem is well known but there is political uncertainty plan and control strategy could best fit. In this sense, the government should allow learning mostly about product and performance.

References

ANIF. (2019). Gran Encuesta a las Microempresas. Bogotá

Amromin, G. y S. Chakravorti (2009). "Whither Loose Change? The Diminishing Demand for Small Denomination Currency," Journal of Money, Credit and Banking, Blackwell Publishing, vol. 41(2-3), pages 315-335, 03.

Angus, M., Jain, A., & Thomas, H. (2013). Mastercard Advisors' Cashless Journey.

Alliance, B. T. C. (2015). Country diagnostic: Colombia. Bankable Frontier Associates (BFA) bajo la supervisión de Beatriz Marulanda.

Arango, C., S. Garrido, N. Suárez (2017a), "Cómo pagan los colombianos y porque?". Borradores de Economía: No. 991.

Arango-Arango, C. A., Zárate-Solano, H. M., Suárez-Ariza, N. F., & Arango-Arango, C. A. (2017b). Determinantes del acceso, uso y aceptación de pagos electrónicos en Colombia. Borradores de Economía; No. 999.

Arango-Arango, C. A., Arias-Rodríguez, F., Rodríguez-Niño, N., Suárez-Ariza, N. F., Zárate-Solano, H. M., Arango-Arango, C. A., ... & Zárate-Solano, H. M. (2020). Efectivo y Pagos Electrónicos. Revista Ensayos Sobre Política Económica; No. 93, febrero 2020. Pág.: 1-76.

Asobancaria. (2017). Avances y Desafíos para los Intrumentos Electrónicos en Colombia. ASOBANCARIA Semana Económica.

Asobancaria (2017). Proyecto F: diagnóstico del uso del efectivo en Colombia (S. Castro, ed.), Bogotá: Comunicación Gráfica y Legis S. A.

Auer, Raphael, Giulio Cornelli, and Jon Frost. Covid-19, cash, and the future of payments. No. 3. Bank for International Settlements, 2020.

Babatz, G. (2013). Sustained effort, saving billions: Lessons from the Mexican government's shift to electronic payments. Better Than Cash Alliance Case Study. http://betterthancash. org.

Banca de las oportunidades y Superintendencia Financiera de Colombia (2019). Reporte de Inclusión financiera 2018.

Baumol, W. J. (1952). The transactions demand for cash: An inventory theoretic approach. The Quarterly Journal of Economics, 545-556.

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

Bech, Morten Linnemann, Faruqui, U., Ougaard, F., & Picillo, C. (2018). Payments are a-changin' but cash still rules.

BIS. (2003). Payment and settlement systems in selected countries.

BIS (2014) Non-bank in retail payments. CPMI.

BIS (2016) Payment aspects of financial inclusion. CPMI. World Bank.

Boeschoten, W. C., & Fase, M. M. G. (1989). The way we pay with money. Journal of Business & Economic Statistics, 7(3), 319-326.Hausmann, R., & Rodrik, D. A. Velasco, 2005. Growth diagnostics". Mimeo, Harvard University.

Boeschoten, W. C., & Fase, M. M. (1992). The demand for large bank notes. Journal of Money, Credit and Banking, 319-337.

Bonilla, R., Córdoba, R., Lewin, A., Morales, O. D., Montoya, S., Perry, G., ... & Villar, L. (2015). Comisión de Expertos para la Equidad y la Competitividad Tributaria.

Caballero, C., & Amaya, A. (2011). La lucha contra el lavado de activos: instituciones, resultados y desincentivos. Política antidrogas en Colombia: éxitos, fracasos y extravíos, 301-328.

Central Bank. (2019). Reporte de Sistema de Pagos, Bogotá, June 2019.

Central Bank. (2016). Reporte de Sistema de Pagos, Bogotá, June 2016.

Chibba, M. (2009). Financial inclusion, poverty reduction and the millennium development goals. The European Journal of Development Research, 21(2), 213-230.

Cricco, M. (2017). Estructura tarifaria del mercado de pagos electrónicos en el Uruguay y otros países de América del Sur.

Chakravorti, S., & Roson, R. (2006). Platform competition in two-sided markets: The case of payment networks. Review of Network Economics, 5(1).

Chakravorti, S. (2010). Externalities in payment card networks: Theory and evidence. Review of Network Economics, 9(2), 3.

Checkley, K. (1999). Cash is still King. Taylor & Francis US.

Chodorow-Reich, G., Gopinath, G., Mishra, P., & Narayanan, A. (2018). Cash and the Economy: Evidence from India's Demonetization (No. w25370). National Bureau of Economic Research.

Daude, C., & Pascal, J. (2015). Efficiency and contestability in the Colombian banking system.

Diniz, E., Porto de Albuquerque, J., & Cernev, A. (2011). Mobile Money and Payment: a literature review based on academic and practitioner-oriented publications (2001-2011). Oriented Publications (2001-2011). (December 3, 2011). Proceedings of SIG GlobDev Fourth Annual Workshop, Shanghai, China December.

EIU (Economist Intelligence Unit), 2019; Global Microscope 2019: The enabling environment for financial inclusion; New York, NY.

Fernandez, C., & Villar, L. (2016). Informality and Inclusive Growth in Latin America: The Case of Colombia.

Galindo, A., & Majnoni, G. (2006). Represión financiera y el costo del financiamiento en Colombia. Finance, Private Sector and Infrastructure Group, World Bank, June.

Garcia, P., Gregory, C., William C., & McKay. C (2016). Interoperabilidad de las finanzas digitales e inclusión financiera: Un análisis de 20 países. Working paper. Washington DC: CGAP.

Giraldo, M., & Buckles, B. W. (2011). The impact of financial transactions taxes on money demand in Colombia. Latin american journal of economics, 48(1), 65-88.

Jonker, N., Plooij, M., & Verburg, J. (2017). Did a public campaign influence debit card usage? Evidence from the Netherlands. Journal of Financial Services Research, 52(1-2), 89-121.

Humphrey, D. B., Pulley, L. B., & Vesala, J. M. (1996). Cash, paper, and electronic payments: a cross-country analysis. Journal of Money, Credit and Banking, 28(4), 914-939.

Lozano-Espitia, I., Vargas-Herrera, H., & Rodríguez-Niño, N. (2017). Financial transaction tax and banking margins: An empirical note for Colombia. Ensayos sobre Política Económica, 35(83), 154-160.

Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2016). Building state capacity: Evidence from biometric smartcards in India. American Economic Review, 106(10), 2895-2929.

Morduch, J. (1995). Income Smoothing and Consumption Smoothing. The Journal of Economic Perspectives, 103-114.

Myles, G. D. (2000). Wasteful government, tax evasion, and the provision of public goods. European Journal of Political Economy, 16(1), 51-74.

Rettberg, A., & Ortiz-Riomalo, J. F. (2016). Golden opportunity, or a new twist on the resource–conflict relationship: Links between the drug trade and illegal gold mining in Colombia. World Development, 84, 82-96.

Rincón-Castro, H., Delgado-Rojas, M. E., & Rincón-Castro, H. (2018). ¿ Cuánto tributan efectivamente el consumo, el trabajo y el capital en Colombia?. Borradores de Economía; No. 1041.

Roa, M., García, N., Frías, A., & Correa, L. (2017). Panorama del dinero móvil en América Latina y el Caribe. Centro de Estudios Monetarios Latinoamericanos, 30-40.

Rogoff, K. S. (2016). The curse of cash: How large-denomination bills aid crime and tax evasion and constrain monetary policy. Princeton University Press

Schneider, F. (2013). The Shadow Economy in Colombia: Size and Effects on Economic Growth. Department of Economics, Johannes Kepler University of Linz. Working Paper N. 1319. October, 2013.

Stephanou, C., Guadamillas, M. (2008). "Cooperation versus Competition in Colombia's Automated Clearing House (ACH) Market", Financial Infrastructure Policy and Research Series, The World Bank.

Tamayo, C. E., & Malagón, J. (2017). Inclusión Financiera en Colombia. Banco Interamericano de Desarrollo.

Thomas, H., Jain, A., & Angus, M. (2013). The global journey from cash to cashless. MasterCard Advisors' Cashless Journey, 1-15.

Tobin, J. (1956). The interest-elasticity of transactions demand for cash. The review of Economics and Statistics, 241-247.

UNODC.(2012). Colombia:Monitoreo de Cultivos de Coca 2012.

URF (2018) Estudio sobre los sistemas de pago de bajo valor y su regulacion. Documento de consulta

URF (2019) Actualizacion normativa de los sistemas de pago de bajo valor. Documento de consulta

Valverde, S. C., Chakravorti, S., & Fernández, F. R. (2016). The role of interchange fees in two-sided markets: An empirical investigation on payment cards. Review of Economics and Statistics, 98(2), 367-381.

Verdier, M. (2011). Interchange fees in payment card systems: A survey of the literature. Journal of Economic Surveys, 25(2), 273-297.

Villar, L., Salazar, N., & Pérez, C. (2016). Tendencia Económica. Informe Mensual de Fedesarrollo. No. 164. 2016.

Weiner, S. E., & Wright, J. (2005). Interchange fees in various countries: Developments and determinants. Review of Network Economics, 4(4).

Wright, J. (2004). The determinants of optimal interchange fees in payment systems. The Journal of Industrial Economics, 52(1), 1-26.

Villa, E., Misas, M. A., & Loayza, N. V. (2016). Illicit activity and money laundering from an economic growth perspective: a model and an application to Colombia. The World Bank.

UAF. (2015). The economic dimension of money laundering. UIAF Documents.

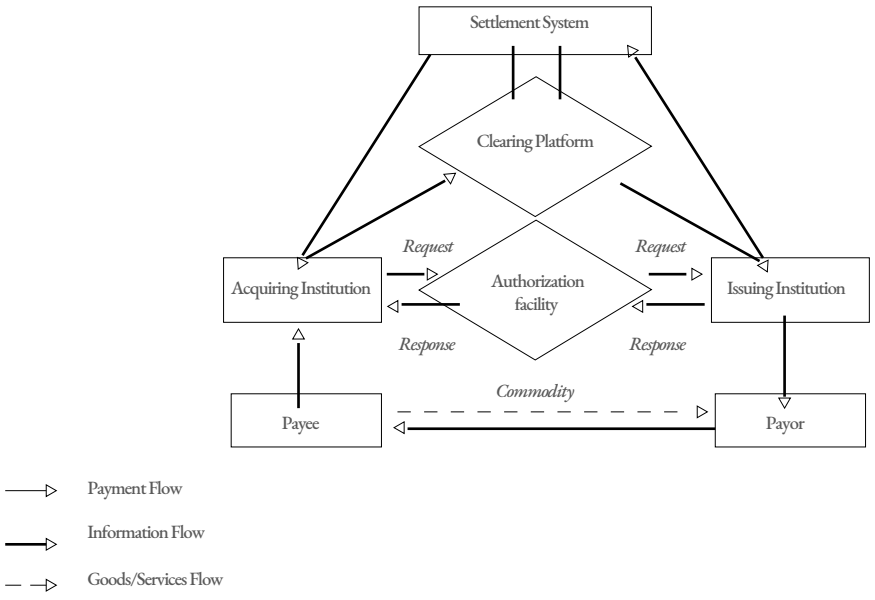
World Bank. (2018). Global Payment Systems Survey (GPSS). Washington, DC: World Bank

Appendix I. Card Payments Systems.

Open card payment systems, such as VISA or Mastercard, are made up of five key actors: the system operator, the card issuing financial institutions, the card acquiring institution, the payor and the payee. Since the tasks of issuing cards and acquiring businesses are separated, the acquirers obtain their income from a fee of discount applied on the price of the final good normally known as merchant discount rate. On the other hand, issuers receive their income based on fixed and variable rates from cardholders, the collection of interest and penalties for non-compliance, for and the collection of a rate to the acquirers, called interbank exchange rate. In particular, the higher the interbank exchange rate, the higher are the discount rates that the acquirers charge to the merchants. According to the BIS (2014), the payment process includes five stages: pre-transaction: creates the initial infrastructure required for payments; (ii) authorization: enable a payment transaction to be authorized and approved before it can be completed; (iii) clearing: enable the submission of claims by members in the payment system against each other; (iv) settlement: posting of credits and debits in the account of the bank/financial institution with the settlement bank; and (v) post transaction: provision of various types of value added services for statement generation, reconciliation, dispute resolution and ex-post compliance services.

Figure 10.

Card Payments System



Source: World Bank (2008)

Appendix II. Electronic payments policies definitions.

Name	Definition
Cutting red tape	Some countries opt to simplify procedures and lessen the administrative burden of the informal economy. These types of measures encourage compliance by providing some type of benefit
Demonetization	Involves reduction of the cash in circulation to reduce the perceived advantages of paying with cash as well as changing habits and other traditions that perpetuate its use
Electronic money	Introduction of electronic store of monetary value on a technical device that may be widely used for making payments to entities other than the e-money issuer.
Governmental electronic payments	Enabling digital transactions for all public matters or providing infrastructure for digital payments at its offices
Fees abolishment or reduction of fees	Authorities or the central bank intervene to cap the level or eliminate of interchange fees that is collected by the issuers in payment, considering that lowering rates significantly could increase acceptance by merchants and adoption and use by of consumers.
Financial Inclusion	Strategies to protect the financially disadvantaged people: building up crucial electricity, internet, and mobile infrastructure, while educating citizens about the benefits of financial products.
Improve Electronic Payment Structure	Several countries have established acceptance development funds to improve the acceptance of digital payments in geographic regions or market sectors with low penetration. In addition, technological advances are encouraging terminal adoption. Speed and convenience have also boosted the use of near-field communication (NFC) terminals.
Regulatory changes to enhance competition	Legal reforms that attempt to mitigate transaction costs through competition by private sector providers in the establishment and operation of financial infrastructure, taking advantage of economies of scale, scope and network externalities
Tax incentives	Incentives for using card payments have proven to be an effective way to increase the use of electronic payments (e.g. tax rebates for cad payments), especially when linked to lower taxes.

Source: Own elaboration based on Kearney et al. (2015) and Stephanou & Guadamillas (2008)

Appendix III. Main national governmental initiatives related to electronic payments

Name	Mechanism	Year	Objective
Financial Inclusion Policy and creation of the investment program "Banca de las Oportunidades"	National Plan 2006-2010	2006	Developing projects to raise the levels of financial inclusion in the country, and creating the necessary conditions to promote access to the formal financial system and the use of comprehensive financial services to the population with lower incomes, small and medium enterprises and entrepreneurs
Creation of the banking correspondent model	Decree 3078	2006	Credit institutions were authorized to provide their financial services through different commercial establishments, with low operational implementation costs and in terms of infrastructure
The financial inclusion policy was set as a long-term action	Law 1151	2007	Established a set of instruments to bring financial services to the underserved population, including the creation of low-cost savings accounts
Financial education is defined as a principle between financial consumers and supervised entities	Law 1328	2009	Established that as a legal rightof the financial consumer to receive an adequate education about the different products and services
Promoted measures to guarantee the permanence of the formal financial offer in those remote places	Law 1450 National Plan 2010-2015	2011	Strategy to promote access to financial services and financial education in the design of programs for the development of basic skills. Adopt a financial inclusion goal that consisted in increasing banking access of the adult population from 57% to 68% during the four-year period.
Colombia signed the Mayan Declaration of the Alliance for Inclusion Financial	AFI	2011	Some of the commitments derived from the declaration are: (i) the promotion of cost-efficient inclusion policies for SMEs, (ii) the generation of an impartial regulatory environment, (iii) the recognition of financial education and (iv) consumer protection, as the basis for financial inclusion.
SEDPES Specialized Deposit and Electronic Payment Companies are created	Law 1735	2014	Achieve greater competition in the provision of transactional financial services at a lower cost, and thus allowing a greater number of Colombians to access these financial services
Creation of the National Administrative System for Economic and Financial Education	Decree 457	2014	A set of policies, guidelines, guidelines, norms, Public and private activities, resources, programs and institutions related to the financial education.
Formed the Intersectoral Commission for Financial Inclusion	Decree 2338	2015	Instance of policy orientation and coordination of entities related to financial inclusion.
They included a series of goals to advance the access, use and deepening of financial services.	Law 1753. National Plan 2014-2018.	2015	This included an increase in the financial inclusion indicator from 72.6% in 2014 to 84% in 2018, reduction of cash from levels of 11.7% in 2014 to 8.5% in 2018. In order to reach the agreed goals for the four-year period, two strategies were designed, one for financial inclusion and another for economic and financial education.
National Plan 2018-2022	National Plan.	2018	Low value payment ecosystems from the promotion of SEDPES, which currently total 5 entities monitored by the Financial Superintendence, and the adoption of new technologies in means of payment in rural areas and urban of the country

Source: Own elaboration based on CONPES draft Financial Education and Inclusion (DNP, 2020).

Appendix IV. Detailed Technical Assessment of Policy Alternatives.

a) Eliminating the tax on financial transactions (GMF):

As shown in Section II.B., recent evidence demonstrates that the GMF produced an initial negative effect on the people's willingness to use financial products. Yet, twenty years after its implementation, this effect does not appear to be the most binding constraint. The evidence finds no substitution effect from any bank account toward cash holdings, and customers and firms do not recognize it to be a driving force for a higher cash preference. The authors claim that this can be explained by a one-time change in agents' behaviors, which were not affected by the subsequent changes. Also, there are some laws that exempt users from paying the GMF or allow them to receive fiscal benefits reducing the overall impact on the economy. It is also essential to underscore that many other Latin American countries adopted a tax of this kind, including Argentina, Bolivia, Brazil, Ecuador, Mexico, Peru, and Venezuela. Like Colombia, most of them were introduced as emergency and temporary taxes but have become permanent as a consequence of the ability to deliver relatively easy tax collection.

This tax represents a significant source of government funding (1.3% of the GDP for 2015) and ranks to be the second most important source of funding for the central government. It is important to underscore that in mid-2011, the congress of Colombia approved a fiscal rule that sets the structural fiscal deficit on a declining path, and the government is then legally forced to achieve balance goals. Thus, eliminating the tax becomes a sensitive topic for the central government as the following challenge would be how it would be able to replace it. This political dilemma is tangible, bearing in mind that this tax was supposed to be temporary, and it has been changed three times to a permanent tax. The Ministry of Finance (2015) estimated that the eventual abolition of the GMF would require around a three-percentage point increase of the VAT. Even more, the decision to eliminate it nowadays is harder in the context of fiscal austerity after the drop in oil prices of 2014. Notably, in 2015 the Ministry of Finance conformed an Expert Committee for Equity and Tax Competitiveness (EXET) to seek out for the redesign of the tax system of the country. In this regard, the Committee concluded, *"the Commission understands that the GMF has negative effects on financial intermediation. However, it is an easy collection tax that generates significant resources to the Treasury"* (Bonilla et al., 2015). It not clear which tax could replace it. For example, the VAT was already increased in the year 2016 by three percentage points, and the effective tax rate on firms is estimated to be one of the highest in the regions (section II.C).

b) Gradually phasing-out paper currency beginning by removing the high value notes

Rogoff (2016) explores the past, present, and future of currency, concluding that, contrary to conventional economic wisdom, paper money lies at the deep-rooted causes of some of the world's most challenging problems. The author provide evidence that for advanced economies, large-denomination bills aid crime, tax evasion, and constrain monetary policy. Therefore, he proposes a strategy for these countries to gradually phase out the use of paper currency – beginning with large notes. For instance, Jangid and Sohini (2017) analyzed at least 13 countries that have adopted demonetization policy for curbing black money, fighting against corruption tendencies, and holding inflation. The authors found evidence that except for minor exceptions, the majority of the countries could not achieve their goals through demonetization.

A clear example of this policy is the case of the government of India. On November 8, 2016, the government unexpectedly declared the highest notes bills - which represented 86% of the existing currency in circulation - illegal tender, effective at midnight[61]. In this regard, Chodrow-Reich et al. (2020) showed that districts experiencing more severe demonetization exhibited lower economic activity and lower bank credit growth[62]. In particular, they found evidence of a contraction in aggregate employment and night lights and bank credit of 2 percentage points in 2016Q4 relative to their counterfactual paths. Nevertheless, these adverse effects dissipated over the next few months. The experience of India then provides us mixed results. Although it appears to exhibit a positive outcome for the long run, the trade-off in the short term is a deceleration in the economic activity, particularly for informal and rural households. Notably, the Indian experiment was done in such a fast and unannounced way, which created undesirable outcomes. Against this measure, Rogoff (2016) has reaffirmed that any phase-out of cash should take place slowly and with ample anticipation to allow households to adopt other forms of payment.

In Colombia, the largest existent bill paper is the 100,000 pesos note introduced in 2016. At the moment of introduction, Asobancaria, the Colombian banking Association, was against the creation of this note arguing that it could promote and exacerbate informality and illegality. Against this reaction, the Central Bank argued that the quantity and volume of economic transactions had grown over the last years, while the denominations of bills in circulation have remained unchanged. In fact, Villar et al. (2016) found that the value of the high-note Colombian peso bill ($32) was way below the average ($ 58) of the group of countries with similar income levels to Colombia (approximately $USD 7,900 a year). Also, the average for higher-income countries (such as Japan, Canada, Sweden, and New Zealand) was $90, well above Colombian example[63]. The authors also argued that the introduction of the 100,000-peso bill was aligned with the evolution of wages and the general level of prices. In the absence of the 100,000 notes, around 13 notes would be required for the minimum wage, and 18 for the basic basket amounts well above those observed when changes were made to the highest denomination notes in the last quarter-century. Moreover, four years after its introduction, the penetration of this note bill in the economy is considerably limited: corresponds to 5.6% of the total value of banknotes in circulation and only 0.5% of GDP (Figure X). The following high note is the bill of 50,000 pesos, which represent almost 80% of total cash as percentage of GDP. Eliminating this will represent a vast impact to the economy that is not considered in this analysis.

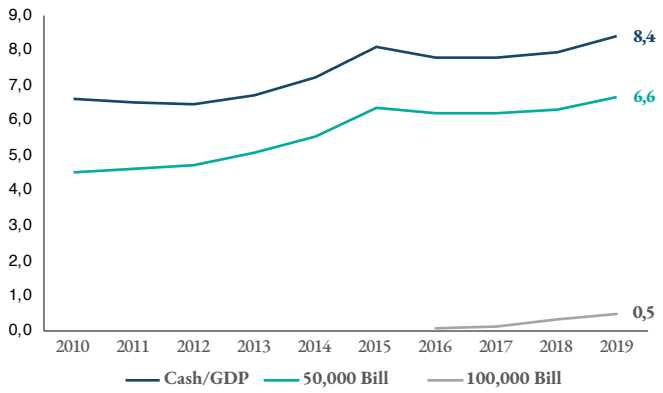
[61] On 8 November 2016, the Government of India announced the demonetization of all ₹500 and ₹1,000 banknotes of the Mahatma Gandhi Series. It also announced the issuance of new ₹500 and ₹2,000 banknotes in exchange for the demonetised banknotes

[62] Likewise, Hosain (2019) found that the demonetization shock had a significative negative effect on the economy (particularly for rural and informal sectors), yet in the post demonetization it appeared to increase non-cash transactions, tax returns and demand in rural and formal households.

[63] The 100.000 pesos bill is similar to the largest denomination in Brazil, Chile and Bolivia, while is almost half of the Peru and Mexico highest denomination, since the dollar equivalent for each of these are 61 and $58 respectively (Fedesarrollo, 2016).

Figure 11.

Cash and high-value denominations (% of GDP)



Source: Central Bank of Colombia (2019)

c) Create a strategy to promote firms' appropriability of financial services

As aforementioned, although we found that in the extensive margin Colombia has experienced an increase in the penetration of financial products, in the intensive margin people are not employing them on their day to day transactions. When we analyzed the demand side, we found that a predominant trigger behind this outcome was the absence of electronic infrastructure available in businesses: while more than 80% of the citizens appear to have access to at least one financial product, small firms (those with monthly sales below 600 USD) only exhibit an 11% acceptance of this methods of payment. Therefore, a strategy to obtain firm's acceptance of these technologies should be warranted.

This appears to be a widespread issue in the region due to the high level of informality and tax evasion. In fact, as already mentioned, governments have already implemented diverse solutions to face this challenge by requiring by law that all businesses should offer the possibility of electronic payments (Costa Rica), establishing public subsidies for promoting this acceptance (Mexico), or using a carrot and stick system (Uruguay and Argentina).

The authors estimate that this measure would produce a negative effect in the tax collection of $ 730 billion (0.09% of GDP).

These experiences shed light on technical desired ways to implement these measures for Colombia. There should be reduction of the high costs of the firms of adopting digital payments or providing incentives for firms. Aligned with this idea, in fact, the Expert Committee for Equity and Tax Competitiveness (2015) recommended that GMF to increase VAT deductibility of corporate income tax from 50% to 100%. The authors estimate that this measure would produce a negative effect in the tax collection of $ 730 billion (0.09% of GDP). This would improve the acceptance of firms towards these products and eliminate some of the tax distortions.

As in a "carrot and stick" system, requiring all businesses to offer the possibility of electronic payments seem to be ideal but is complicated with actual legal framework which limits the inherence of the state in any private initiatives[64]. Also, the supervision and accountability of these processes are still very limited. This would be an additional barrier in the context of Colombia, with regional disparities in terms of economic development and financial access. If this legal constrain holds, the government could consider providing this infrastructure directly.

A second component that the government should consider in this strategy is financial education. As highlighted by the Ministry of Planning (2020) the country does not count with standardized guidelines to monitor the offer of Economic and Financial Education programs. Even though some entities offer programs of non-formal financial education, there is no information on the results of these strategies or the impact they generate. The Ministry of Planning (2020) also highlights that, in addition to informality, lack of information emerges as a main barrier for SMEs accessing financial– relevant considering SMEs represent on average 92% of total firms (Confecamaras, 2018). In fact, according to the National Association of Financial Institutions - ANIF (2019) entrepreneurs misinterpret the rates they pay for their loans in the financial system hindering their willingness to acquire financial services[65]. Moreover, the perceived insecurity and lack of knowledge on using virtual services are considerable barriers to this type of transactions. The entrepreneurs survey made by Superintendencia Financiera and Banca de la Oportunidades (2016) regarding financial inclusion found that entrepreneurs identified as barriers to access financial products: high costs (15%), insufficient income (14%), distrust (6%) and aversion to procedures and requirements (4%).

[64] The Article 333 of the Constitution highlights that private initiative are free, within the limits of the common goods. For its exercise, no one may demand prior permits or requirements, without authorization from the law.

[65] On average 54.3% of the entrepreneurs indicated that the interest rate on your credit was less than 5%, value much lower than interest rates traditional banking for these companies, alongside on average 11.3% of Entrepreneurs surveyed report that they do not know what the interest rate of their credit is (ANIF, 2019)

d) Focus on regulatory interventions to enhance competition and access on low value payment infrastructure

The economic importance of payment systems has spurred a lot of research during the two decades. The literature has converged to analyze it as a "two-sided market", where payment cards are demanded by two heterogenous groups of consumers – cardholders and merchants/businesses (Rochet & Tirole, 2003; Baxter, 1983). The key implication is that the price allocation on the two sides of the market matters, or the share that each type of end-user pays the platform, affects the total volume of transactions. Another main implication is the existence of two potential externalities – usage and network[66]. Thus, the standard economic prescription to deal with externalities involves adjusting prices so that agents' private incentives reflect the true social costs and benefits of their decisions. In the payment system then interchange fee may improve efficiency by internalizing externalities - appropriate pricing of card services for all the parties.

In this context, the determinants of interchange fees in payment card systems has generated a number of public policy debates. One prominent debate is the level in which private marketplace (competition) yield interchange fees that are efficient. In traditional markets, competition is generally viewed as an effective mechanism for providing incentives that lead private agents to make choices that produce efficient outcomes. Thus, the core issue is whether a card network, in competition with other card networks and payment methods, has incentives to choose an interchange fee that produce efficient prices to both sides of the market. Notably, the issue of the substitution between cash and electronic payment instruments lies at the heart of this debate. According Verdier (2009), the substitution between cash and payment cards can only occur if consumers and merchants are provided with the appropriate incentives, through the interchange fee, to use and accept cards, respectively, instead of other payment instruments.

Theoretically, Rochet and Tirole (2003) developed a model to show that the interchange fee set by a single monopoly card association can be identical to that set by two competing schemes. Guthrie and Wright (2005), an addition, demonstrate that when merchants are strategic, this result no longer holds, and inter-system competition will generally affect the interchange fee in one direction or another. Furthemore, Chakravorti and Roson (2006), considering the effects of competition on both price level and structure, show that competition may generate a fall in the total price that is charged to end-users, which generally enables consumers' and merchants' welfare to be increased, while reducing the profits of payment networks. The conclusions of the theoretical literature are then that network competition has effects on interchange fees, depending crucially on the behavior of the two side consumers and country-specific factors. Notably, authors highlight that changes in external factors such as greater awareness of the benefits of payment cards or reductions in processing and credit intermediation costs may result in greater adoption and usage by consumers.

Empirically, Weiner & Wright (2005) revealed an up-trend in the US for both the top-five issuer concentration measure and average interchange fees. However, when analyzing the experience of more than 20 countries the authors confirmed that there is no obvious relationship between issuer market concentration and interchange fees, and that a number of complex and interrelated factors, many country-specific, played a role in interchange developments.

We can conclude then that theorical and empirical literature on two-sided markets for payments system provides no standard answer on the most effective competitive policy that would ensure dynamic efficiency; much depends on initial conditions and the broader institutional setup. For Colombia, Section II demonstrated that the actual market structure reveals to hinder competition and access of new players to the payments system by discriminatory or arbitrary application of access requirements to the system, and/or lack of transparency in access fees or the methodology of their calculation.

---

[66] Usage externality refereed as each party in a given transaction evaluates his or her own costs and benefits associated with a particular payment method but does not consider the costs and benefits of the other party. The network externality reflects the fact that the value of a payment card network increases for both merchants and consumers as the card becomes more ubiquitous (Praguer et al, 2009).

Furthermore, Stephanou & Guadamillas (2008) revealed two main concerns that hinder Colombia's payment market structure and are still present to date. First, the presence of two ACH platforms that increased contestability, but direct competition was inhibited by some discriminatory business practices. Although it may be argued that the coexistence of a public and a private ACH can be justified to avoid a market monopoly, this theory is easily disregarded as both clearing houses have traditionally catered to mostly different market segments[67]. Despite the difficulty to find an appropriate benchmark to assess the market structure, the authors emphasize that the existence of some barriers to direct competition between the 2 ACHs is not a positive sign. Finally, the authors expect adverse effects on economies of scale and scope from the coexistence of 2 ACHs. For instance, given the significant fixed costs reported by both ACHs, there are 'macro level' inefficiencies from having a duplication of infrastructures. Besides, ACH Colombia has greater economies of scale due to its size, yet differences in ownership structure across retail payments systems in Colombia have prevented it from clearing other payments instruments and thereby experiencing economies of scope[68]. Second, the multiplicity of relevant policymakers and the absence of adequate institutional coordination mechanisms have hindered the development of an effective oversight function. The complex oversight framework has led to a lack of trust and transparency among market participants.

Second, the multiplicity of relevant policymakers and the absence of adequate institutional coordination mechanisms have hindered the development of an effective oversight function.

Stephanou & Guadamillas (2008) proposed four different policies to overcome these issues: (i) maintain two separate ACH platforms but strengthens competition between them so both platforms would need to modify and publicly disclose their pricing policies, combined with a stronger pro-competitive stance taken from the authorities; (ii) merge the two infrastructures into a unique ACH platform to potentially lower operational costs (by leveraging economies of scale) and overall pricing, coupled with stronger governance arrangements and a robust oversight and antitrust framework; (iii) enhance transparency in the functioning of the ACH market to dispel mistrust and further promote competition; mainly by greater public disclosure of the operating arrangements s (i.e. shareholder structure, decision-making mechanisms, pricing and access policies etc.); (iv) strengthening oversight arrangement by the establishment of robust institutional coordination mechanisms.

e) Implementing a cap level to the interchange fees

As already established, the issue of the substitution between cash and electronic payment instruments lies at the heart of interchange fees debate. According Verdier (2009), the substitution between cash and payment cards can only occur if consumers and merchants are provided with the appropriate incentives, through the interchange fee, to use and accept cards, respectively, instead of other payment instruments. Valverde et al (2016) suggest that a 10 percent reduction in the rate of decline per quarter in the average interchange fee by an acquirer resulted in a rate of increase in merchant acceptance per quarter of up to 1.4 percent. Notwithstanding, in the context of immature payment card markets, such as Colombia, there is evidence that lowering rates significantly increases acceptance by merchants and adoption and use by of consumers (Carbó et al., 2009). In this context, in several countries (e.g. Australia, Mexico, Spain, Europe.),

[67] With ACH and ACH CENIT serving commercial banks and the central government, respectively.
[68] This has been the case in several European countries (Belgium, Finland, France, Italy, Netherlands, Portugal, Spain) whereby a single electronic platform handles the processing of various payments instruments (Stephanou & Guadamillas, 2008)

the competition authorities or the central bank intervened to cap the level of interchange fees that is collected by the issuers in payment systems.

Extensive academic and policy documents have been published with inconclusive views on how regulate prices in the payment card industry (Chakravorti, 2010; Bolt and Chakravorti, 2011). Economic theory provides guidance about some of the effects of such a reduction in this two-side market, although the empirical magnitudes of these effects are currently unknown (Praguer et al, 2009). On the merchant side of the market merchant discounts would be expected to fall. Thus, merchant acceptance of the card would likely increase; and this effect would be higher for a card that already enjoys low merchant acceptance (e.g. Colombia). On the card user side of the market, card user fees (rewards) would be expected to rise (fall), thereby making the card less attractive to consumers, compared with other payment methods. Therefore, the final effect would hinge on the balance between the effect consumer substitution away from card system and the impact on the increased merchant acceptance[69].

In this context, determining an appropriate regulated value for the interchange fee can be quite challenging. Ideally, the regulator would want to set the interchange fee equal to the efficient level (i.e., the level that internalizes externalities between the parties to a transaction). Calculation of that fee requires knowledge of social costs and benefits that are difficult, if not impossible, to measure accurately. Also, the determination of which costs should be included in a cost-based fee is necessarily arbitrary, and measuring those costs is nontrivial, particularly if frequent re-estimation of costs is necessary. Furthermore, as recognized by the URF (2018), it is also important to underscore that setting limits on these rates could possibly lead to costs being transferred to other customers of the system or charged to other products offered to the same users, as occurred in the European Union with the implementation of the IFR.

[69] There is also some interrelation between agents. For instance, the he extent to which any cost savings are passed through to customers would depend on the merchants' market power in the final product market.

Are banks going with the flows? An empirical assessment of the international credit channel in the Pacific Alliance countries

July 12, 2021

María Andrea Figueroa Suden*

* Presented by María Andrea Figueroa Suden (ma.figueroa10@uniandes.edu.co) and advised by David Pérez-Reyna (d.perez-reyna@uniandes.edu.co).

Abstract

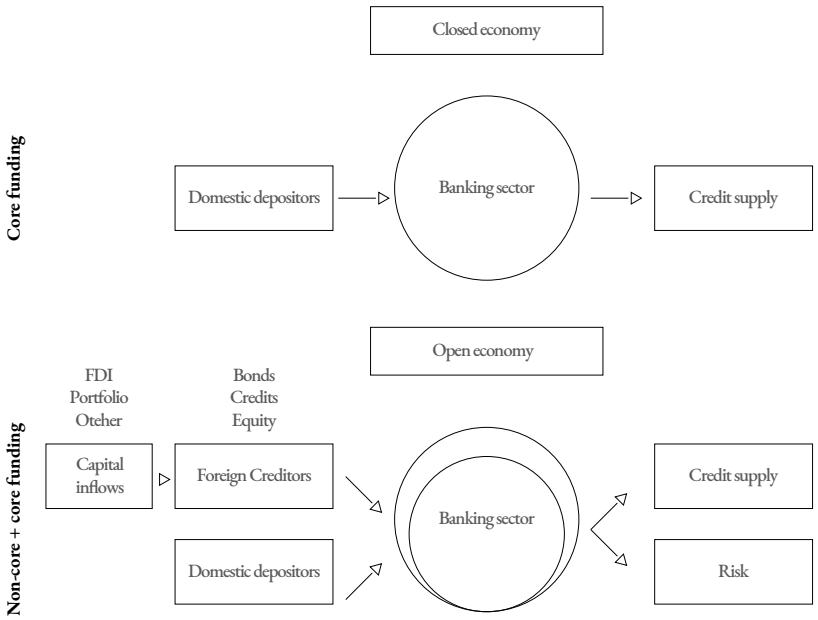
Based on a cross-country bank-level analysis within four emerging market countries in Latin America, this document addresses how international capital flows from 2016 to 2019 impact banks' loan supply and risk, and if these effects are contingent upon the banks' funding strategies. Our identification strategy relies on the VIX index as a measure of the global financial cycle to instrument debt capital inflows in a two-stage least squares framework. Our main findings are that as debt portfolio inflows increase 10%, the net loan portfolio expands 0.7% and the solvency risk exposure increases 1.0%, *ceteris paribus*. Testing whether these results are subject to the type of funding, we stress out that a higher concentration of non-core funds in the liabilities side of the balance sheet amplifies–rather than mitigates- the former results. As extensions to our baseline model, we provide evidence that capital inflows impact is differential across types of non-core funding (bonds vs. credits) and bank's ownership (foreign-owned vs. domestic).

I. Introduction

Capital flows' volatility raises the need to understand the complex underlying mechanisms of the international credit channel –how capital flows run through domestic financial sectors-. This is particularly relevant for emerging markets that have been recipients of external flows to fill domestic savings shortfalls. Since capital flows' impact on financial systems has several edges, we narrow the scope of study to its implications on bank-level credit supply and risk. Furthermore, by conditioning on the type of funding strategy, we propose funding sources to be the underlying connecting thread of the international credit channel.

Figure 1.

The international credit channel



*Based on Hahm et al (2012).

As shown in Figure 1, unlike closed economies, in an open economy the banking sector has access to alternative funding sources to deposits, i.e. non-core liabilities, from foreign creditors. Among them, for example, credits with overseas entities or bonds acquired by foreign investors that are reflected in the

balance of payments as gross capital inflows, more specifically in the debt inflows account.

Since these external inflows enable banks to act as financial intermediaries of a larger pool of funding sources, it is worth finding out if the increased quantity and reduced price of foreign funds modifies banking institutions' lending dynamics and if it has further consequences on banks' risk levels. In the same vein, it is relevant to point out if capital flows' impact is heterogeneous across financial entities depending on their funding structure because first, as institutions reliant on noncore sources of funding benefit disproportionately from the savings shortfall of the domestic economy, capital inflows should impact to a larger extent lending from this type of banks. Second, the literature has recently deemed as more vulnerable and unstable a composition of liabilities tilting towards non-core funds rather than deposits (Hahm, J., et. al., 2011b). Thereafter, we find it is worth testing these risk differentials across types of funding.

Based on a cross-country analysis within four emerging market economies in Latin America (Pacific Alliance members), this document addresses how international capital flows affect banks' loan supply and risk, and if this effect is contingent upon their funding strategy. To answer this question, we exploit the variability of the balance sheet's composition across 101 banks in a four- year time horizon -from 2016 to 2019- distributed in the following countries: Mexico, Colombia, Chile, and Peru.

In relation with the two bank-level outcomes of credit supply and risk we must highlight that credit supply is captured by the natural logarithm of the net loan portfolio, while risk is measured with the Z score. This indicator is used to assess insolvency risk following the existing literature (Altunbas et al. (2011), Demirgüc-Kunt and Huizinga (2010), among others). Taking into account ROA stands for Return on Assets, CAR for Capital Assets Ratio, and SD(ROA) for the standard deviation of Return on Assets, this ratio is mathematically defined as follows:

$$Z - Score = \frac{ROA + CAR}{SD(ROA)}$$

Intuitively, the Z-score determines "the number of standard deviations that a bank's return on assets has to fall for the bank to become insolvent" (Demirgüc-Kunt, 1993). Therefore, it must be interpreted as the inverse likelihood of insolvency, which means that a lower Z-Score represents high risk and lower bank stability, in comparison to a high Z-score reading.

Our findings yield that, as debt portfolio inflows increase by 10%, the net loan portfolio expands on average by 0.7% keeping everything else constant. On the other hand, while there is not enough evidence to conclude about debt capital flows' impact on solvency risk, it appears to move banks closer to their insolvency threshold. To size this effect, the Z score decreases 1.0% in response to a 10% increase in debt portfolio inflows. Decomposing this ratio into its two additive components -portfolio and leverage risk-, we find out that leverage risk is the main driver behind Z score results. It decreases 0.5%, in contrast to leverage risk which actually increases 0.3%, in response to capital inflows rising 10%.

Testing whether these effects are amplified or mitigated by banks' funding structure choice, we find that a higher dependence on non-core liabilities amplifies -rather than mitigates - the former results. In other words, insofar as the bank's balance sheet has a larger share of non-core funds, the net loan portfolio increases by a higher magnitude, while the Z score falls even more. To quantify these results, for each 10% increase in gross portfolio inflows, a bank with an average share of non- core liabilities to total funding sources increases its lending by 0.9% and decreases its Z score by 1.0%, *ceteris paribus*. Meanwhile, a fully-funded bank by non-core sources expands its net loan book by 2.1% (120 basis points more) and decreases its Z score by 1.1% (an additional 1 basis point).[1] Therefore, we provide evidence that non-core liabilities in the balance sheet drive the net loan book's procyclicality to capital inflows. It might happen that the increased quantity of loans granted brings along an additional layer of risk to the ongoing operation, nonetheless there is no conclusive evidence of seriously risking financial stability.

Therefore, we provide evidence that non-core liabilities in the balance sheet drive the net loan book's procyclicality to capital inflows.

As extensions to our baseline model, we also test differential impacts across banks' type of non- core funding (bonds versus credits) and ownership type (foreign-owned versus domestic). Our findings point out that among non-core sources of funding, banks having bonds as their preferred funding source are more procyclical to capital flows. This kind of "unmonitored lending", where investors free ride others' monitoring activities, might lead to excess lending and increased exposure to insolvency risk. On the other hand, we encounter that foreign-owned banks play an essential role in the international credit channel, also driving the procyclicality of credit to capital flows, as these banks have a more direct channel through which funds flow: *the internal capital* markets within the holding company.[2]

Identifying the causal effect of capital flows on financial systems raises endogeneity concerns. Current and future economic prospects affect domestic economic conditions (interest rates, inflation, and credit growth, among others), while at the same time might induce more or less foreign lending. Besides this, to assess the impact of capital flows on credit supply, it is crucial to disentangle confounding factors, such as demand-driven loan expansions or contractions, and particular macroeconomic conditions of the countries where banks operate.

We use an instrumental variables approach to address these endogeneity issues associated with reverse causality and omitted variable bias. In particular, we rely on a two-stage minimum least- squares framework instrumenting portfolio flows with a measure of the global financial cycle: the VIX index (Chicago Board Options Exchange Volatility Index). This index tracks the implicit volatility in the US stock market by averaging the weighted prices of the S&P 500 out-of-the- money puts and calls. Investors know it as the "fear gauge" because it reflects the time-varying degree of market wide risk aversion and general sentiment.

The VIX Index is of paramount importance explaining the underlying forces of financial flows across the world as it leverages investment decisions from the financial community as a whole (investors, risk-rating agencies, research analysts,

[1] Although fully-funded banks by non-core sources are a minority in our sample, they do exist and we highlight this case for the sake of comparison with an average bank in terms of funding.

[2] Evidence of banks having active internal capital markets -as many other business organizations- has been provided by many authors. Among them Cetorelli and Goldberg (2011) which will be addressed in the literature review. Conceptually, these internal capital markets are viewed as a managerial tool of liquidity and a source of efficiency since resources are allocated following optimization principles.

portfolio managers, etc.). Moreover, several authors in the literature such as Calvo G. et. al (1993), Calvo & Reinhart (2000), Fratzscher (2012), Ghosh et al. (2014), and Passari and Rey (2015), point out external factors ("push factors") –instead of internal ("pull factors") as the main determinants of capital flows. They have not only provided empirical evidence of the VIX index's explanatory power but have also reported that the sensitivity of portfolio flows to it has increased over time for the Latin American region.

Our identification strategy relies on the exclusion assumption in the **short-term.** Given the highly liquid positions represented in debt portfolio flows and their shorter investment horizon in comparison to foreign direct investment, these flows are more elastic in terms of speed and magnitude of adjustment to global risk perception changes than any other variable that might end up affecting domestic financial conditions. In this regard, and considering the quarterly frequency of our dataset, the exclusion restriction holds since the VIX index is not likely to affect the supply of lending nor risk by any other channel distinct to higher or lower capital inflows in a short time window as we will discuss it later.

This document contributes to two distinct but interconnected literature branches: the study of capital flows' impact on financial systems and the financial implications of different bank-level funding strategies. Our findings contribute to this existing literature in four ways. First, we examine in tandem two phenomena identified as sources of macroeconomic imbalances: the external position and financial instability. Second, we employ a cross-country analysis in the context of emerging economies, which are far more sensitive to external shocks deriving from global liquidity conditions. In the third place, we exploit the dataset's granularity to disentangle the micro effects of capital flows and contributing to a better understanding of bank-level decisions. Lastly, we provide empirical evidence on which funding strategies are likely to increase the financial system's resilience to external flows. While funding strategies have been assessed in response to a particular event, mainly a monetary or financial shock, evidence in the light of the external position is limited and has hardly been identified along a complete cycle. Overall, our purpose is to elucidate the trade-offs regarding banks' funding choices in open economies.

a. Capital flows impact on financial systems

Various lines of research detach from our first study field of capital flows. This literature's origin goes back to the analysis of push (external) and pull (internal) factors to explain cross-border flows. In the context of Latin America, Calvo, et. al. (1993) attribute these inflows to external conditions, such as lower international interest rates and recessions in the United States. In this work, we distance ourselves from this standpoint by exploring capital flows' consequences rather than its causes. Turning to the empirical evidence, several studies have highlighted current account deficits in emerging markets as an early warning of financial crises. Mendoza and Terrones (2008) find that emerging markets crises usually involve domestic credit booms preceded by capital inflows. In the same fashion, Ostry et al. (2011) conclude for a panel of emerging markets that 60% of credit booms, which led to crises from 1995 to 2008, were related to surges of capital inflows.

Huang and Ratnovski (2008) raised a theoretical debate by coming across another model that underlined "the dark side" of non-core sources of funding.

In this work we will focus on the growing body of evidence in recent years postulating financial entities' liabilities as the second and most direct channel through which capital flows permeate the financial sector. This strand of literature gained relevance with Hahm et al. (2011b). The authors highlighted the crucial role international capital flows play in emerging market economies as sources of banks' non-core funds. Since capital markets are less developed and deep across these emerging latitudes, banks are forced to resort to foreign creditors when their funding needs rise and domestic deposits are not enough to support more lending. That is how, according to Hahm et al. (2011b), shifts from core to non-core funding sources might spur lending booms.

b. Financial implications of different bank-level funding strategies

This non-traditional form of financing leads us to the second branch of literature in which this work is based: financial implications of funding strategies. Its origin traces back to models of agency problems by Diamond (1983), and Calomiris and Khan (1991). These authors found non-deposit funding beneficial for banks as they encourage discipline and represent third-party monitoring, apart from the depositors and the regulator. They also argued that banks raising capital in international financial markets, and having stable access to them, provided creditworthiness signals mitigating problems related to information asymmetries. Notwithstanding, Huang and Ratnovski (2008) raised a theoretical debate by coming across another model that underlined "the dark side" of non-core sources of funding. Given that some of these funds are instantly demandable and the majority of them are excluded from deposit insurance schemes, they might withdraw at the first negative public signal, even triggering solvent banks' to be liquidated.

Empirical developments in this research line were mainly motivated by the Global Financial Crisis (GFC) in 2009. Several authors blamed the increased financial fragility on the system's reliance on non-core liabilities. For example, Demirguc-Kunt et al. (2011), using an international sample of 1334 banks in 101 countries prior to the financial crisis, demonstrated that overall non-deposit funding strategies were risky. Along these lines, other authors have explored how non-core liabilities from the banking sector can reflect the degree of a systems' financial vulnerability. Based on a sample of emerging and developing economies, Hahm et al. (2011a) found out that non-core liabilities predicted both credit and currency crises. Focusing more on the link between non-core liabilities and financial procyclicality, we encounter Kim et al. (2012) study for the Korean case, Damar et al. (2012) for the Canadian banking system, and Hamann (2014) in the context of the Colombian financial sector. These three authors stress the relative importance of non-core liabilities in enhancing financial risks at a systemic level.

As mentioned above, in this document we differentiate the effects of funding across banks that rely primarily on bonds (public debt) versus credits (private debt), which is practically underexplored in banking institutions' extant literature.[3] Many papers have focused on analyzing the decision of financing at a firm-level belonging to the real sector. However, we apply many of these findings in the banking sector's case, as they revolve around the notion of moral hazard.

According to Diamond (1984), the monitoring exercise is better executed by a financial intermediary (private debt) than multiple individual investors (public debt), which might end up free riding the task done by other agents. In the same vein, Sharpe (1990) and Rajan (1992) build on the idea that in the private debt case, the information asymmetry is partially mitigated. Lender banks have access to more private information than investors and are able to impose covenants on the borrower banks reducing the leeway for self-interested actions, and fostering compliance.

Furthermore, we also test in an extension to our baseline model how effects diverge between domestic and foreign-owned banks. This strand of literature, inaugurated by Cetorelli and Goldberg (2011, 2012a and 2012b), points out that international banking linkages due to global banks' presence have played a significant role in transmitting financial shocks. Foreign-owned banks belonging to a broader organization can rely on *internal capital markets* (credit and funding lines within the holding company), in contrast to domestic-owned banks which depend exclusively on *external capital markets*, making it easier for them to tap international resources.

In synthesis, this second branch of literature relating to liabilities composition provides conflicting theoretical approaches. At the same time, empirical evidence is limited to assessing funding strategies under particular conditions or in response to a specific event, such as a financial crisis or monetary stance change. Our goal is to expand the evidence of how funding strategies perform in response to a complete and standard capital flow cycle. The value-added factor comes from a cross- country analysis within emerging market countries that differentiates across the type of funding (bonds versus credits) and bank (domestic versus foreign-owned).

[3] In this work, public debt refers to bonds (not Government debt) as they are fixed income securities made available to the public when companies such as banks need to raise capital.

Lastly, to the best of our knowledge, the closest papers to our investigation are Dinger and Te Kaat (2017) and Baskaya et al. (2017) in the sense that both explore the impact of capital flows in bank- level outcomes conditional on the funding strategy. Nonetheless, our work differs from the former by analyzing the international credit channel in the context of emerging –not advanced- economies and by measuring the impact in solvency risk rather than in credit risk. Although both risks are part of banks' core risk management, solvency risks have been more closely linked to several systemic problems. On the other hand, we distance ourselves from Baskaya et al. (2017) by doing a cross- country analysis, instead of a country-specific investigation.

This paper is structured as follows: The data and its descriptive statistics are the focus of Section **2.** In Section 3, we present our empirical strategy. Section 4 exploits our dataset's bank-level dimension by examining the resulting effects of capital flows. Section 5 provides robustness tests and Section 6 the study's limitations. In Section 7 we conclude and in Section 8 we compile the Annexes.

2. Data

Our panel data set comprises quarterly bank-level observations from the four Pacific Alliance countries between 2016Q1 and 2019Q4. Balance sheet data were drawn from the regulatory entities responsible for supervising the banking sector in each country.[4] Data is provided on a regular and standardized basis for banks as local authorities follow closely financial sector evolution. We employed unconsolidated balance sheet data as our source of analysis because consolidated financial statements encompass foreign subsidiaries' operations. Since capital flows vary in intensity from one country to another, using consolidated information would bias these flows' effect on bank-level outcomes. Lastly, we excluded from our sample only those banks that did not appear at least two quarters, ending up with a sample of 101 banks.

We combine banks' observations with a country-level dataset by matching over a set of macroeconomic variables. The foremost are the capital flows recorded in the financial account obtained from the Balance of Payments and International Investment Position database from the International Monetary Fund (IMF). Other macro variables drawn from the International Financial Statistics database of the International Monetary Fund (IMF) too, include GDP growth, monetary policy rate, and inflation. These variables are used as control regressors in our identification strategy, which is explained in Section 3.

[4] In the case of Colombia: Superintendencia Financiera de Colombia, Peru: Superintendencia de Banca, Seguros y AFP, Chile: Superintendencia de Bancos e Instituciones Financieras Chile, Mexico: Comisión Nacional Bancaria y de Valores.

a. Description of the data

Table 1 displays summary statistics for the average financial account among the Pacific Alliance members. Between 2016Q1 and 2019Q4, countries in the sample recorded current account deficits, which were financed largely by FDI and portfolio investment. The former with an average of net inflows around 3.5 billion USD and the latter with an average 1.4 billion USD. Discriminating by type of flow stands out the fact that FDI –as it is widely recognized in the literature- has been the most stable form of financing with a standard deviation around 3.2 billion USD due to its long- term investment horizon. Meanwhile, other investment has been the least stable with a 4.5 billion USD standard deviation.

Table 1.

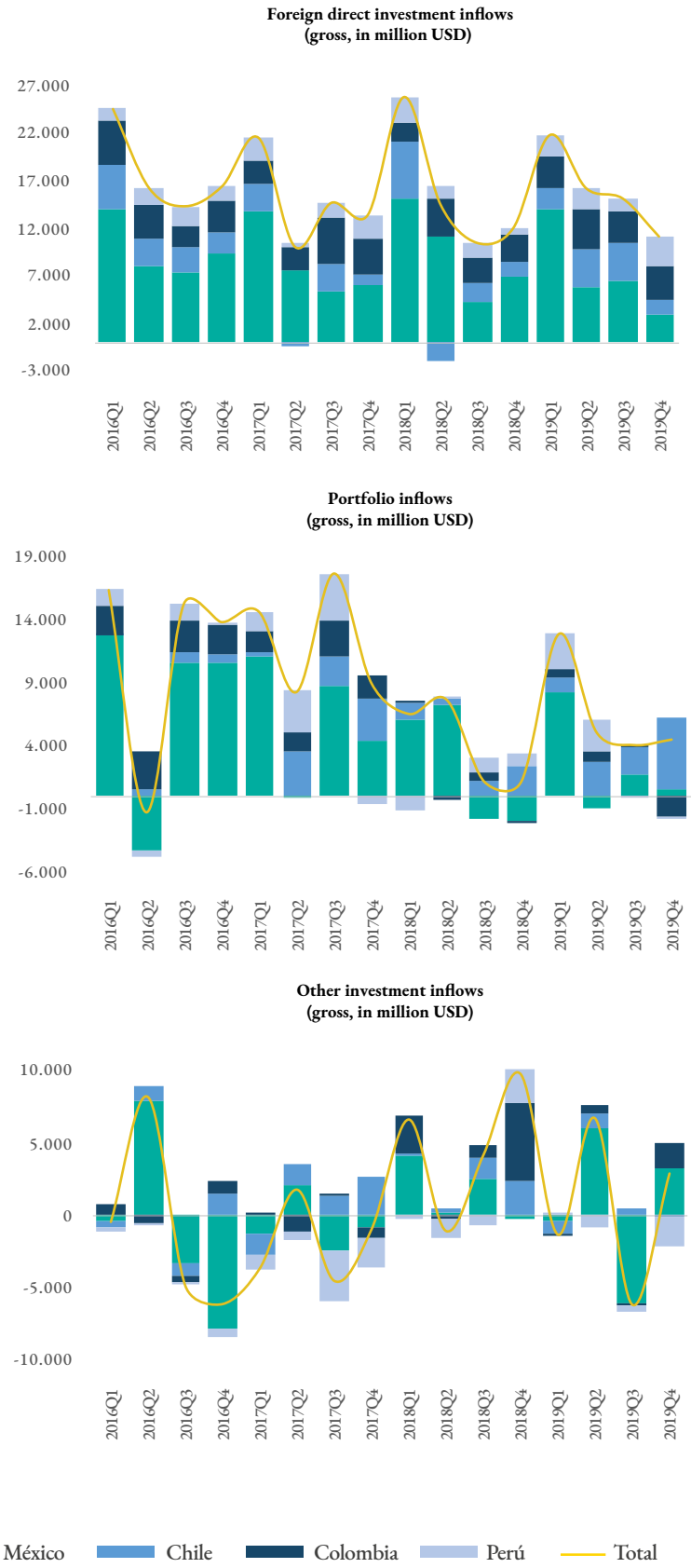
Summary Statistics - Financial Account BoP (Million USD)

	Obs.	Mean	Std. Dev.	Percentile 25	Median	Percentile 75
Foreign direct investment, net	1,616	-3,518	3,243	-5,246	-2,469	-1,265
Gross inflows	1,616	4,793	3,827	2,346	3,472	6,122
Gross outflows	1,616	1,276	1,484	631	1,228	2,183
Portfolio investment, net	1,616	-1,397	3,607	-2,255	-1,092	536
Gross inflows	1,616	2,526	3,728	271	1,434	3,452
Gross outflows	1,616	1,129	2,173	-352	916	2,121
Other investment, net	1,616	882	4,500	-1,597	-347	1,674
Gross inflows	1,616	304	2,656	-740	-24	1,428
Gross outflows	1,616	1,186	3,760	-945	575	2,118

As shown in Figure 2, Mexico was the largest inflows recipient, considering the sum of its three components: FDI, portfolio, and other investment inflows. Meanwhile, Chile, Peru, and Colombia had lower but relatively similar levels. The volatility proper from each type of flow is also reflected in Figure 2, where FDI has had a steady behavior during the time window considered, while other investment has recorded large swings.

Figure 2.

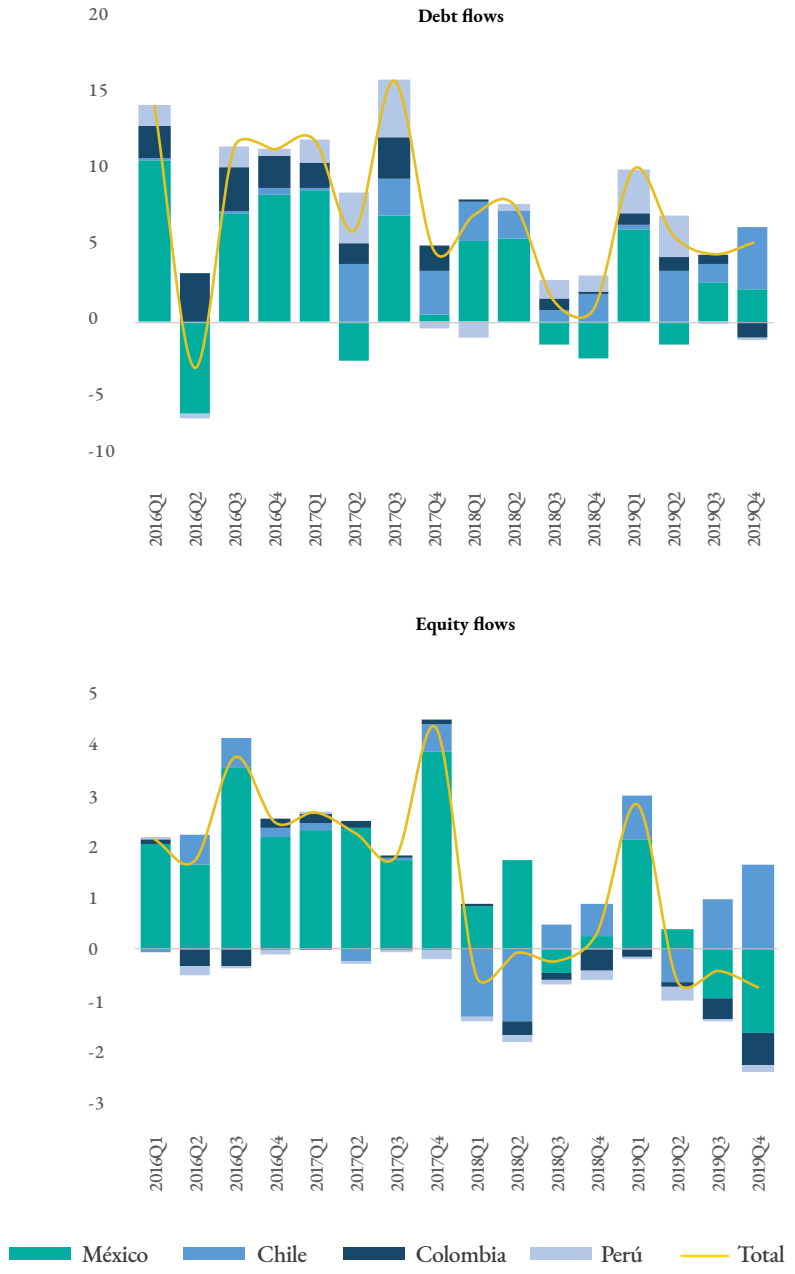
Capital flows to Pacific Alliance member countries (gross, in million USD)



Our focus is in debt portfolio inflows[5] due to their impact on financial systems' lending dynamics. The first and most direct channel is by materializing in the liabilities side of balance sheets as these flows not only increase the quantity, but also reduce the price of foreign funding. The second channel, which is indirect, is related to a substitution effect between the loan and the investment portfolio at a bank-level dimension. When debt portfolio inflows are driven by foreign appetite on Government securities, banks might sell off their investment positions to these investors, releasing liquidity that enables them to increase their loan portfolio. Because of the reasons mentioned above, we use debt portfolio inflows (portrayed in the first graphic of Figure 3) as our main independent variable. Moreover, as shown here, debt flows are also more evenly distributed in the sample countries than equity flows, aiding the identification strategy and yielding more interesting and valuable results.

Figure 3.

Portfolio capital flows (debt and equity) to Pacific Alliance member countries (gross, in million USD)



[5] Portfolio capital inflows, debt capital inflows, and debt portfolio inflows are used interchangeably from here onwards all referring to debt portfolio inflows.

In relation to bank-level data, in the final sample the Mexican financial system accounts for the largest share of total observation units with 36 banks (36%), followed by Chile and Colombia, both with 26 banks (26%), and lastly, Peru with 13 banks (13%). As of December 2019, the banks included in the sample represented 100% of total assets of the Colombian and Chilean financial system, respectively, 98% of the Mexican and 80% of the Peruvian.

Table 2 summarizes key statistics from the main items in the balance sheet for an average bank in the sample. The net loan portfolio accounts for the majority of assets (54.1% on average), although it might vary widely, taking into account its 25.6% standard deviation. The investment portfolio has a share of around 14.3% in total assets, while cash and interbanks represent on average 10% and 2.4%, respectively, of total assets. In terms of liabilities, the average bank across the sample is deposit-reliant with deposits having a 61.5% share in total liabilities, with a 24.8% standard deviation though. Credits and bonds follow deposits with a share of 1.2% and 5.1%, respectively, meanwhile interbanks represents less than 1% of liabilities. Lastly, capital as a percentage of total assets stands, on average, around 15.9% with a standard deviation of 15.6%.

Table 2.

Summary Statistics – Bank-Level Characteristics

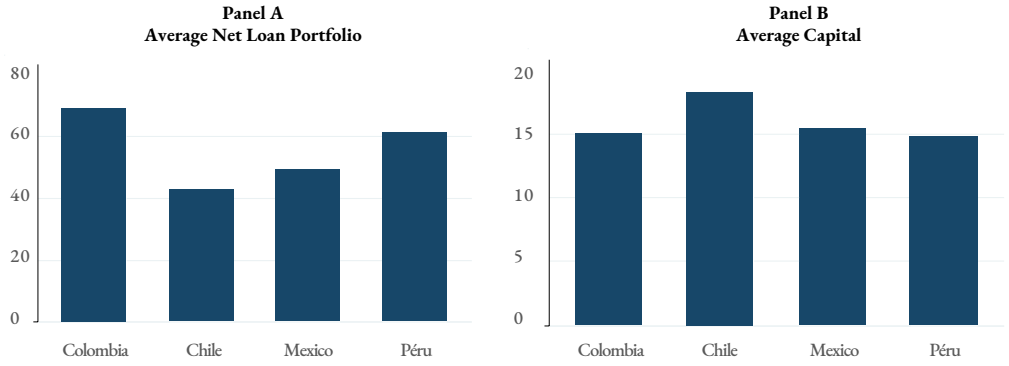
		Obs.	Mean	Std. Dev.	Percentile 25	Median	Percentile 75
Cash	(% of total assets)	1,616	10.0%	10.0%	4.5%	6.9%	11.8%
Interbanks		1,616	0.8%	2.4%	0.0%	0.0%	0.4%
Investments		1,616	14.3%	14.8%	3.3%	10.1%	21.1%
Net Loan Book		1,616	54.1%	25.6%	36.6%	64.0%	73.7%
Deposits	(% of total liabilities)	1,616	61.5%	24.8%	55.7%	66.8%	77.2%
Interbanks		1,616	0.5%	1.9%	0.0%	0.0%	0.3%
Bonds		1,616	5.1%	7.9%	0.0%	0.0%	7.6%
Credits		1,616	7.4%	11.2%	0.1%	3.6%	9.2%
Capital	(% of total assets)	1,616	15.9%	15.6%	7.9%	11.7%	16.9%

To characterize the banking sector in each country, we focus first on differences across the average bank, in terms of assets and capital, before digging deeper on liabilities which is our focus of study. According to Figure 4 panel A, the average Colombian bank is more concentrated in traditional lending activities, than the Chilean, Mexican, and Peruvian. For the entire period, the net loan portfolio as a percentage of total assets stood around 69% in Colombia, followed by Peru with an average 61%, while in Chile and Mexico it averaged between 40%-50%. This means banks in these two countries had a larger share of other assets such as investments. Concerning capitalization ratios, measured as the total equity divided

by total assets, Figure 4 panel B depicts Chile as the country with the highest indicator on average (18%), with a considerable distance from the Colombian, Mexican, and Peruvian, all averaging approximately 15%. This implies banks in Chile have higher buffers to absorb unexpected losses for which the cushions of capital are built.

Figure 4.

Average Net Loan Portfolio and Capital (as % of total assets)

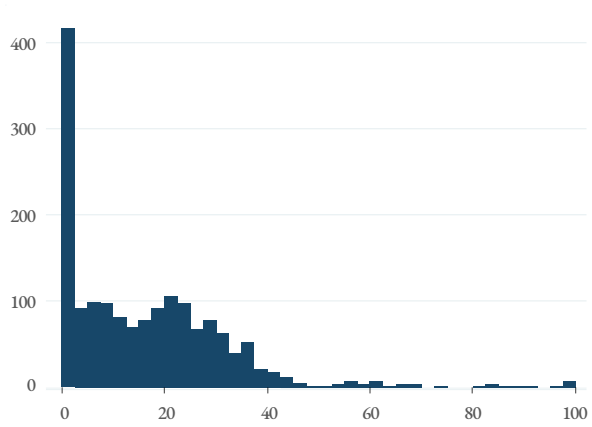


Turning to the liabilities-side of balance sheets, our main variable of interest is the non-core ratio. In line with Harutyunyan et al. (2015), we define it as the total sum of non-deposits as a share of total funding sources.[6] Figure 5 portrays the frequency distribution of this ratio across banks in the sample. The majority of them have between 20% and 30% of their funding sources represented in non-deposits, although there is also a significant concentration of observations between 0% and 20% and some observations piling up around 60% and 99%.

[6] Funding sources is considered as the sum of deposits, credits, bonds and interbank liabilities. Non - core ratio is defined in this way due to the availability of balance sheet data and taking into account the differences in information disclosure by each of the regulators in the Pacific Alliance member countries.

Figure 5.

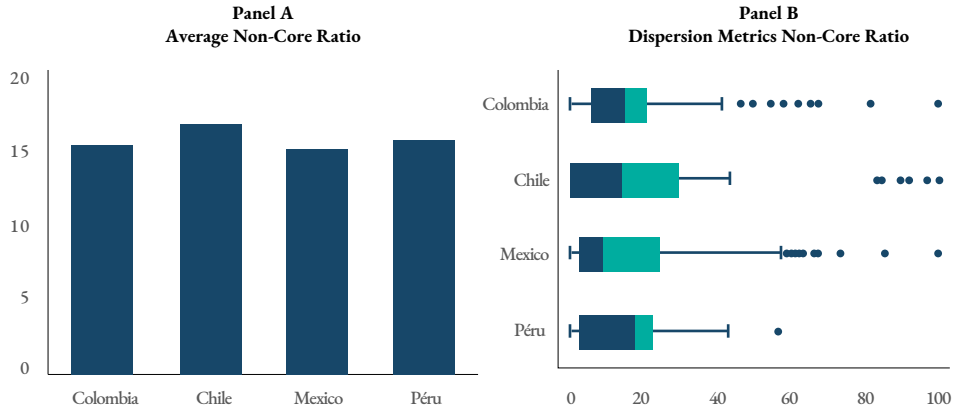
Frequency distribution of Non-Core Ratio (non-deposits as % of total funding sources)



Discriminating by country (see Figure 6 panel A), the average non-core ratio is higher in Chile (16.9%) than Peru (15.8%), Colombia (15.6%), and Mexico (15.3%). This result is in line with the average share of the net loan portfolio in terms of assets, where Chile recorded the lowest share. It is reasonable that banks tilt their funding sources towards deposits as their assets are more concentrated in the loan (rather than investment) portfolio to keep an ample liquidity profile. Moreover, from Figure 6 panel B it is clear that Chile has not only the highest mean of non-core ratio, but also the broadest dispersion across the sample.

Figure 6.

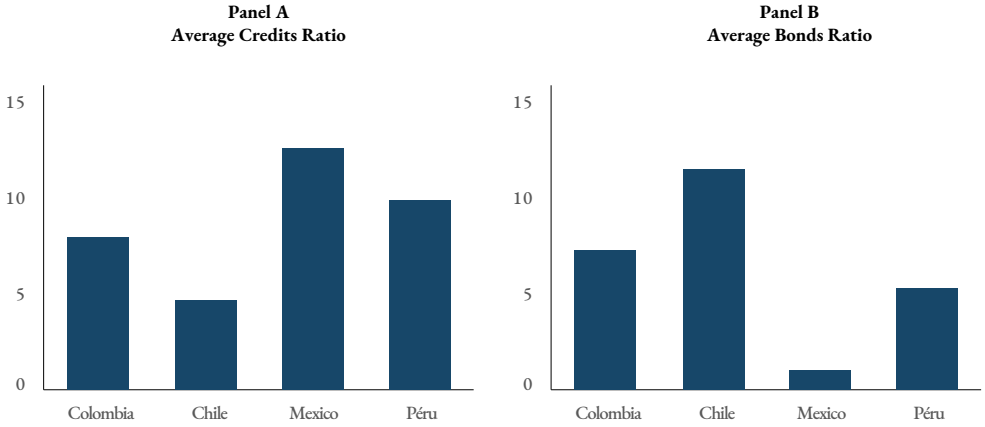
Non-Core Ratio by country
(non-deposits as % of total funding sources)



Since the beginning, we have classified funding sources for the banking sector between deposits and non-deposits. Figure 7 shows for the set of countries in our sample the two primary sources of non-core funding (credits and bonds). As shown in panel A, the Mexican and Peruvian financial systems are the most reliant on credits, representing almost 12.6% and 10.0%, respectively, of funding sources. They are followed by Colombia (slightly above 7.8%), and Chile (around 4.6%). On the other hand, according to Panel B, Chile leads in terms of bonds with a mean share of total funding sources above 11.5%, followed by the average Colombian bank with 7.4%, the Peruvian (5.4%), and the Mexican (almost 1%).

Figure 7.

Bonds and credits with other entities (as % of total funding sources)



Finally, we highlight the innate characteristics from each of these banking systems –and their relation to capital flows– in order to have a better grasp of the underlying dynamics. According to Batini et. al. (2020) article titled *IMF Advice on Capital Flows to Latin America*, Chile is the deepest financial system of the sample countries. However, banks are not the largest intermediaries of the system, but rather pension funds and insurance companies. In the same vein, Chile has "managed to differentiate from other financially-integrated Latin American economies, showing higher resilience to external financial shocks (...) both in normal times and in periods of market stress, as evidenced by the episodes of heightened market volatility" (IMF Chile Article IV, 2018) mainly because of its strong and credible institutional and macroeconomic framework, and a free floating exchange regime.

In the case of Colombia, the IMF Article IV from 2019 highlights the implementation of several regulatory measures as positive since they enhance financial system stability. In this country, where the largest domestic banks have a significant international exposure with their Central American subsidiaries, regulation such as Basel III contributes to mitigate risks associated with their overseas operations. On the other hand, the capital account is less open than the average in emerging markets which might mitigate the domestic impact of large external financial shocks.

As of year-end 2019, Peru's financial system was still largely dollarized, although the Government had been pursuing several de-dollarization policies. Financial dollarization, along concentration and off-balance sheet exposures, has been identified as one of the main vulnerabilities of the financial system (IMF Peru Article IV, 2020). Moreover, its capital account has been categorized as being 100% open and banks are considered the main intermediaries in the country. Therefore, these two effects might amplify any type of external shock.

Lastly, Mexico stands out for having a numerous financial system (many banks operating within the country) but a high concentration of market share, mainly stemming from the subsidiaries of large international banks. The authorities have been carrying an agenda that includes mainly efforts to boost financial sector competition and to improve inclusion with a multi-pronged strategic approach (IMF Mexico Article IV, 2019).

3. Empirical strategy

As mentioned earlier, this paper identifies the impact of international capital flows using as outcome variables two banking sector dynamics: credit supply and risk. Credit supply is captured by the natural logarithm of the net loan portfolio and risk is measured with the Z-score. Since we are interested in testing whether the impact is conditional on

the funding strategy, we include in a second model specification an interaction term between debt portfolio inflows and the non-core ratio. Therefore, we would like to estimate the following two equations for each bank *i* at period *t* in country *j*:

$$1)\ \underline{Y}_{i.i.t} = \beta_0 + \beta_1 DebtInflows_{i.i.t-1} + Bank'_{i,j,t}\ \gamma + Country'_{i,t}\ \delta + \mu_i + \mu_t + \mu_i + \varepsilon_{i.t}$$

$$2)\ \underline{Y}_{i,j,t} = \beta_0 + \beta_1 DebtInflows_{i,j,t-1} + \beta_2 (DebtInflows * Non - Core\ Ratio)_{i,j,t-1} + Bank'_{i,j,t}\ \gamma + Country'_{i,t}\ \delta + \mu_i + \mu_t + \mu_i + \varepsilon_{i.t}$$

Our coefficient of interests are β_1 and β_2, which address the impact of capital flows in banks' credit supply and risk, and the differential effect according to the balance sheet's structure. We add the following covariates to our model to control for bank-specific characteristics that might affect lending and risk. First, we include the total assets' logarithm to control for the banks' size because larger and more prominent banks are able and more prone to increase their loan books. Second, we control for other liabilities in the banks' balance sheets that are not funding sources to make equivalent comparisons across banks with different proportions of total funding sources to total liabilities. In the third place, we include asset quality measured as 90-days past due loans over the total loan portfolio. Moreover, we consider the macroeconomic and institutional environment, in line with the existing literature body, by including the real GDP annual growth rate as bank lending is positively associated with higher economic dynamism. In the same fashion, we include inflation and monetary policy stance, taking into account that both factors determine real interest rate returns. Lastly, we include time, individual, and country fixed effects.[7]

Our baseline model specification violates the assumption of strict exogeneity because of two reasons: causality and omitted variable bias. First of all, causality may run in both directions - from capital flows to loan growth and risk, and vice versa – making the independent variables correlated with the error term-. It is reasonable that a domestic shock, either from the demand or supply side, causes credit to rise steadily, which at the same time -or consequently- generates an asset price boost, fueling the credit boom and attracting even more external flows. On the other hand, capital flows might influence credit growth via the broadening of funding options for the banking sector, which modifies lending standards and determines the level of loans granted. In this sense, since credit supply and risk might also be determined by demand (and not only the amount of capital inflows), it is essential to disentangle the supply-driven factors, which are our matter of interest, from the demand-side ones.

Besides reverse causality, strict exogeneity is violated due to omitted variable bias generated by the unobserved time-invariant heterogeneity stemming from bank-level and country-specific characteristics. For example, banks with a more prudent risk appetite framework and conservative risk management strategy might, even without the presence of capital inflows, issue less credit to the public. Other factors, such as the business strategy or the regulatory moral hazard underlying the banks' operations, are also unobservable and can influence outcome variables even when capital inflows are not considered. Moreover, it is clear from a macro point of view that other elements, such as the regulator's macro prudential framework and the Governments' commitment to safeguard financial stability in case of stress, might also affect investment decisions guiding the amount of capital inflows to a country.

[7] Descriptive statistics of all the variables employed in our baseline model specification are included in Table 28 (Section 8 Annexes).

To deal with endogeneity concerns, we start by including bank fixed effects that enable us to control for all those time-invariant individual characteristics from banks that might affect our variables of interest, thus eliminating a potential bias source. Additionally, we incorporate time fixed effects – on a yearly and quarterly basis- and country fixed effects to further strengthen our identification strategy. These coefficients absorb the effect of all possible heterogeneities in capital flows behavior from a specific period or macroeconomic event in a particular country. Lastly, we leverage the panel structure of the dataset and include the regressors of interest lagged one period intending to avoid mere contemporaneous correlations.

Nonetheless, the foremost strategy we use to address endogeneity relies on the external variable employed to instrument our endogenous variable: capital flows. To find this external instrument, we use the traditional analytical framework of push versus pull factors in which capital flows determinants are embedded. Push factors "relate to the behavior of external variables such as interest rates, economic growth, risk premiums, and the monetary and fiscal policies of the advanced economies" (Arias et al., 2013), while pull factors refer to the domestic "macroeconomic conditions, policies and institutional frameworks of those countries receiving the resources" (Arias et al., 2013), that may attract or repel capital.

Since we want to extract the exogenous portion of portfolio inflows from the macro-financial domestic conditions of the recipient country, we distance ourselves from pull factors and focus on the push ones. The global financial cycle stands out as the main determinant of risk perception, which is in turn the principal driving force of capital inflows (Kalemli-Özcan, 2020). The former notion is reflected in the VIX index, a measure of the U.S. equity market volatility derived from the price inputs of the S&P 500 options, as aforementioned. In essence, our empirical strategy revolves around an instrumental variables model that uses as a source of exogenous variation the changes in debt portfolio inflows brought about by the changing levels of the VIX index. We are confident that this instrument satisfies the assumptions required in a two-stage least squares framework: relevance and exclusion as we will discuss it hereby.

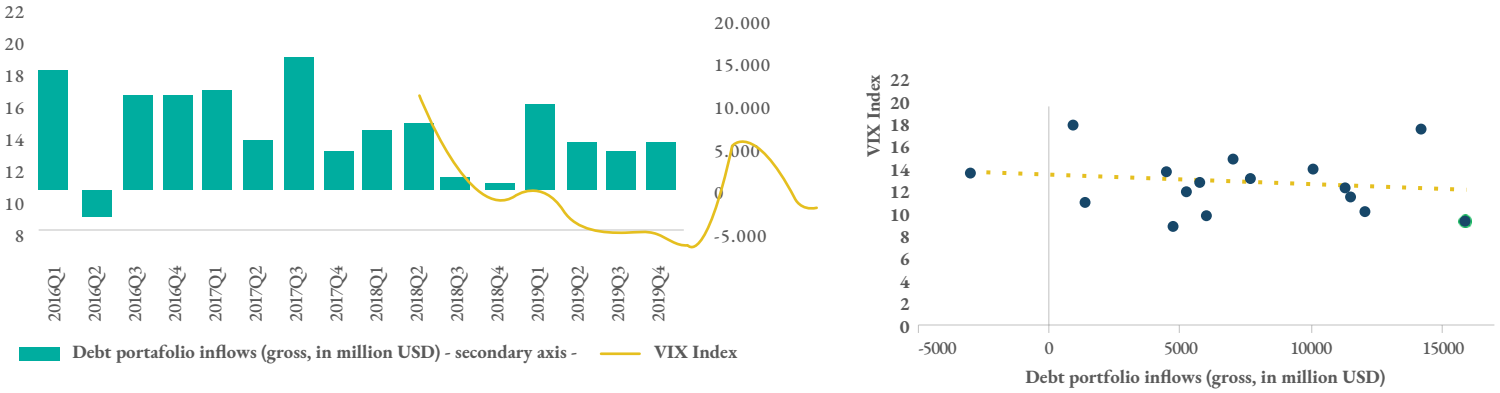
Relevance

First, the VIX index is used by investors, research analysts, and portfolio managers to gauge risks, fear, or stress in the markets, compare returns, and take investment decisions that are later translated into capital inflows or outflows. When the VIX is high, risk perception is heightened, and capital inflows to emerging markets decrease, as investors prefer safer assets such as the US Dollar or gold. This behavior has been widely documented as the *flight to quality*. The

relevance assumption is satisfied not only conceptually but also empirically. Figure 8 displays the VIX index along with debt portfolio inflows to the Pacific Alliance member countries and showcases a scatter plot of these two variables yielding a negative and significant coefficient. These two preliminary approaches give some confidence on the instrument since it is significantly and negatively correlated, as we had hypothesized.

Figure 8.

Debt portfolio inflows and VIX Index



To better grasp on the instrument's relevance we must evaluate the results of the first stage regression displayed in Table 3. Based on Staiger and Stock's (1997) rule of thumb, this assumption's assessment is straightforward running the first stage. They state that the F-statistic of joint significance on this stage should exceed 10 in order to be trustworthy. However, it is desirable to have not only relevant but also strong instruments, avoiding the weak instruments problem.

To do this, we rely on a more rigorous and formal test procedure devised by Stock and Yogo (2005) using the Cragg-Donald Statistic. The null hypothesis is that instruments are weak and the estimator's critical values are calculated taking into account the number of instruments, endogenous regressors, and the size of estimator bias. Both exercises are presented in Table 3. We can conclude

with enough evidence that our instrument is strong, since the F statistics satisfy the 10 rule of thumb and the Cragg–Donald statistics are above the 25% critical values of maximum bias accepted.[8]

Table 3.

Relevance of the instrument – First Stage Statistics

Model Specification	Endogenous Variables	Instruments	Bank Controls	Macro Controls	F-Statistic	Cragg-Donald Wald F Stat
With interaction	$DebtInflows_{i,jt-1}$	$VIXIndex_{i,jt-1}$	Yes	Yes	2549	50,53
Without interaction	$DebtInflows_{i,jt-1}$ $(DebtInflows * NonCore\ Ratio)_{,jt-1}$	$VIXIndex_{i,jt-1}$ $(VIXIndex * NonCore\ Ratio)_{i,jt-1}$	Yes	Yes	1941 40	25,21

Exclusion

Having discussed our instrument's relevance, we turn to the exclusion assumption. Hereby, we defend the idea that the only mechanism through which the instrument (VIX index) affects the dependent variables (net loan book and Z score) is via portfolio capital inflows.[9] We argue that the VIX index is not likely to affect the banks' lending supply nor the risk level by any other channel distinct to higher/lower capital inflows in a **short time** window.

We begin by recalling the existent theory surrounding capital flows, as this gives us a sense of confidence in our results. As mentioned earlier, capital inflows have been attributed to push (external) factors and pull (domestic) factors. If push factors are more relevant in explaining portfolio inflows than pull factors (domestic), we can establish the existence of causality and in a certain way the exclusion restriction is satisfied. This is true because otherwise, if domestic factors were more important, that would mean that portfolio inflows are just the consequence of an already dynamic economy which attracts external funds to continue fueling up credit growth and domestic economic activity.

[8] To view Cragg-Donald critical values please refer to the Annexes.
[9] Since we have an exactly identified system, we can not rely on the Sargan– Hansen J-test to test exclusion.

Nonetheless, the literature has proven our thesis in several documents. Calvo G. et. al (1993) and Calvo and Reinhart (2000) revealed that capital inflows to Latin America were determined by outside conditions of the region. Moreover, Fratzscher (2012) stressed that push factors acted as the main drivers of capital flows during the GFC, especially for emerging countries. In the same vein, Passari and Rey (2015) and Ghosh et al. (2014) provided evidence that global factors such as VIX explained to a large extent the capital flows cycle. All these studies gives us some sense of reassurance on causality as the VIX index (being a key push factor) would extract the portion of capital inflows which is exogenous to the current state of economies in the countries involved.

However, our main argument to prove the exclusion restriction relies on the frequency of our dataset. Taking into account that we employ quarterly figures, the exclusion restriction is satisfied because this instrument affects the availability and magnitude of capital inflows **before** any other variable that might influence our outcomes of interest. We acknowledge it is feasible that the VIX index influences in the medium and long term other variables that might also impact our outcomes. Nevertheless, all of these variables follow the initial response of capital flows to the VIX index, as these higher amount of external inflows unleashes a set of events that may affect the demand side generating the confounding factors we want to omit. In the following lines we will focus on three variables which could be thought as challenging the exclusion restriction, but actually end up complying with it.

First, we would like to highlight the FED monetary policy, as it could lead to changes in interest rates in Latin America which affect the level of credit supplied by banks in this region. We recognize that US policy makers may react to a nervous and uncertain stock market. Rigobon and Sack (2003) showed that the Federal Reserve responds systematically to stock prices (captured by the VIX index) by loosening monetary policy. This change in US monetary policy could influence central banks in Pacific Alliance member countries to modify their domestic monetary policies too because of the *fear of floating channel,* distorting bank-level decisions in which we are interested. However, Bekaert et. al. (2010) provided evidence that "high VIX readings are correlated with expansionary monetary policy in the medium run-future", more specifically after 13 months. That helps us to discard immediate changes in monetary policies by the central banks in Mexico, Colombia, Chile, and Peru, which could affect the loan supply and risk through a channel distinct to the one we are defending.

In the second place, we explore how emerging markets country risk is affected by the VIX index using EMBI (Emerging Markets Bond Index) to gauge it. The EMBI is technically the spread between the sovereign yields of any particular country and the US sovereign bonds yield. According to Jubinski and Lipton (2012), investors overreact to changes in stock volatility (captured in the VIX index) increasing yields on the riskiest bonds (emerging markets) and decreasing them in the safest ones (advanced economies). In this way, in the event of higher VIX levels recorded, EMBI of emerging countries such as the ones considered in this document increase significantly. As a result of higher country risk perception, interest rates for emerging markets rise, causing international funding costs for domestic banks to increase and lending ends up being constrained. However, this only happens due to the fact that lower levels of capital inflows recorded in the balance of payments make the foreign currency a scarce fund increasing the price of it domestically. Again, the only channel that ties the VIX index with banks' supply and risk is the influence stock volatility might exert on foreign capital flows attracting or repelling them.

Lastly, one could think of changes in the foreign exchange rate (FX) motivated by the VIX index which could affect bank-level credit supply and risk. Cairns, et.al. (2007) defended the idea that at times of heightened global stock volatility (reflected in high levels of the VIX index) high-yielding currencies depreciate, while low-yielding (typically from advanced economies) appreciate. Moreover, as revealed by Coudert, et. al. (2011), "investors in need for liquidity to meet their margin calls may sell risky assets across the board, including emerging currencies" setting the stage further for a depreciation. In our case, all Pacific Alliance member countries belong to the emerging markets category, suggesting they will depreciate in the light of increased VIX levels. Likewise to the EMBI case, external funding would become more expensive insofar as the decreasing levels of capital inflows make the foreign currency supply scarcer, which translates into a depreciation of the local currency. In this way, lending would be tightened due to FX depreciation but via capital inflows.

Empirically, we estimate the first stage of debt portfolio inflows as a function of the VIX index and a set of exogenous covariates. This enables us to extract the exogenous variation of capital inflows. Next, we regress our outcome variables over the fitted values of debt portfolio inflows and the same set of covariates and fixed effects in the second stage. This is our instrumental variables approach to deal with endogeneity. Results will be discussed in the next chapter.

4. Results

We have highlighted above that our baseline model specification has two potential endogeneity concerns yielding inconsistent and biased estimators when estimated through OLS: omitted variable bias and reverse causality. Using within fixed effects as the estimation method solves the endogeneity arising from time-invariant unobserved heterogeneity by demeaning all variables, but do not addresses reverse causality. Results from these two estimation methods are shown in Table 16 and Table 17 for the net loan book and the Z score as the outcome variable, respectively, in the Annexes (Section 8). Differences between the model specifications shown is that two of them include the interaction term between debt portfolio inflows and non-core ratio while the other two do not. At a first glance, when the interaction term is included, results suggest that a higher share of non-core liabilities mitigates rather than amplifies the effects on credit supply and risk. It is also relevant to stress out that although FE Within coefficients (and standard errors) are smaller in magnitude to OLS, these preliminary results must be read with precaution since both methods yield biased and inconsistent estimators.

Compared to OLS and FE Within, the two-stage least squares framework yields more reliable and trustworthy estimators (see Table 4). It does so by getting rid of the initial endogeneity concerns instrumenting portfolio flows with the VIX index, as it is a valid instrument in both dimensions (relevance and exclusion). We will analyze first the results having the net loan book as outcome variable to proceed thereafter with the Z-score results.

According to the first model in Table 4, a 10% increase in portfolio inflows leads to a 0.7% expansion of the net loan book. Testing the differential impact due to the funding strategy confirms that non-core liabilities in the balance sheet drive the net loan book's procyclicality with respect to debt capital inflows. This can be concluded as the interaction term of flows and non-core funding sources in the second model of Table 4 is positive and relevant, amplifying –rather than mitigating- the initial results. In terms of magnitudes, for each 10% increase in debt portfolio inflows, a bank with an average share of non-core liabilities to total funding sources increases its lending by 0.9%; while a fully-funded bank with non-core sources expands its net loan book by 2.1% *ceteris paribus* (an additional 120 basis points).[10] These effects are persistent over time: with a 95% level of confidence we can say that lagged capital flows up to three quarters (t-3) impact the net loan book in t positively.[11]

Regarding Z-score results (last two models in Table 4), capital inflows' coefficients are in both specifications negative, suggesting that debt portfolio flows actually increase the exposure to insolvency risk. To size this impact, this ratio decreases 1.0% as inflows rise 10%. Besides this, the funding dependence on non-core liabilities also amplifies the increased exposure to insolvency. This result makes sense with the previous one of credit supply, as an increased net loan portfolio might add another layer of risk to the ongoing operation. More specifically, a 10% increase in debt portfolio inflows, decreases the Z score for an average bank (in terms of non-core liabilities share in total funding sources) by 1.0% and by 1.1% (1 basis point more) in the case of a fully-funded bank by non-core sources *ceteris paribus*. However, there is not enough statistical evidence to confirm this indicator is negatively impacted by portfolio flows, neither it is possible to conclude if it is differential upon the funding strategy as the interaction term also lacks significance (see Table 4). However, assessing persistence of the effects, it is worth highlighting that with a 95% level of confidence debt capital inflows from two quarters earlier impacts the Z score in *t* negatively as displayed in Table 19).

[10] At the end of this chapter, Table 11 summarizes results for the different outcome variables according to certain positions in the non-core ratio distribution.
[11] Estimation results for the two-, three-, and fourth-period lagged debt capital inflows are included in the Annexes.

Table 4.

2SLS –Net loan book and Z score

VARIABLES	2SLS Ln Net Loan Book	2SLS Ln Net Loan Book	2SLS Ln Net Loan Book	2SLS Ln Net Loan Book
Ln Debt Flows$_{t-1}$	0.0678	0.0670	-0.102	-0.102
	(0.142)	(0.149)	(0.0912)	(0.0914)
Interaction term$_{t-1}$		0.138**		-0.00323
		(0.0540)		(0.0254)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,515	1,515	1,515	1,515
Number of id_banco	101	101	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Focusing on the model's results using risk as the outcome variable, we decompose the Z-Score into its two additive components to gain further insights, in line with previous literature about banks' risk-taking (Köhler, 2014). The two elements that make up the Z-score are the portfolio and leverage risk, respectively, which are mathematically expressed as follows:

$$Portfolio\ Risk = \frac{ROA}{SD(ROA)} \qquad Leverage\ Risk = \frac{CAR}{SD(ROA)}$$

Portfolio and leverage risk are measures of returns and capital to total assets, respectively, adjusted by risk levels (the ROA standard deviation). They are also interpreted as more unstable as they decrease in value. Running the two-stage least squares regression with the same instrument for debt portfolio inflows (the VIX index), while changing the outcome variables to these two risks, we arrive to results shown in Table 5. At first sight, it is noticeable that leverage risk is driving the Z score results as the coefficients are negative, while portfolio risk yielded positive coefficients meaning that banks are distanced from the insolvency threshold in this dimension. Taking into account the funding structure (in the second and

fourth model of Table 5) non-core liabilities amplify the initial results in both cases: decreasing portfolio risk exposure and increasing leverage risk exposure.

Table 5.

2SLS –Z Score Decomposition: Leverage and Portfolio Risk

VARIABLES	2SLS Ln Net Loan Book	2SLS Ln Net Loan Book	2SLS Ln Net Loan Book	2SLS Ln Net Loan Book
Ln Debt Flows$_{t-1}$	0.0251	0.0249	-0.0480	-0.0480
	(0.141)	(0.141)	(0.0681)	(0.0683)
Interaction term$_{t-1}$		0.0224		-0.00771
		(0.0454)		(0.0213)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,515	1,515	1,515	1,515
Number of id_banco	101	101	101	10

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In Table 6 we dig deeper into what is driving the increase in leverage risk exposure and the reduced exposure to portfolio risk. We find out in model (1) that capital is decreasing in response to an increase in inflows. Moreover, that non-core funded banks decrease it more than deposit-reliant institutions (see model (2)). These two facts explain why leverage risk exposure rises.

On the other hand, regarding portfolio risk, we encounter that banks depending on non-core sources of funding decrease their assets less than core funded entities in model (4). In particular, they decrease the low-yielding assets (mainly cash and cash equivalents) which are the most liquid, as shown in model (6). Meanwhile, they increase the most high-yielding assets: loans as proved earlier and investment assets (model (8)). Therefore, quarterly profits record a steeper fall in the case of deposit reliant banks as shown in model (10).

The Z-score decomposition sheds light on the drivers of higher risk and how the funding structure does affect the original results. While there is not enough statistical evidence to draw conclusions, we might argue that the main threat stemming from capital inflows is related to leverage risk and, in the second place, to liquidity risk as both capital and liquid assets decrease more in non-core funded banks than deposit-reliant institutions. In terms of portfolio risk, there is no additional risk on this dimension for banks tilting their funding structure towards alternative funds to deposits.

Table 6.

2SLS –Z Score Decomposition: Capital, Liquid Assets, Investment Assets, and Profits

	2SLS (1)	2SLS (2)	2SLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	2SLS (7)	2SLS (8)	2SLS (9)	2SLS (10)
VARIABLES	Ln Capital	Ln Capital	Ln Assets	Ln Assets	Ln Liquid Assets	Ln Liquid Assets	Ln Investment Assets	Ln Investment Assets	Ln Quarterly Profit	Ln Quarterly Profit
Ln Debt Flows$_{t-1}$	-0.0418	-0.0416	-0.334	-0.329	-0.143	-0.143	0.305	0.305	-0.248	-0.248
	(0.0594)	(0.0606)	(0.317)	(0.307)	(0.122)	(0.126)	(0.289)	(0.289)	(0.526)	(0.526)
Interaction term$_{t-1}$		-0.0295		0.242		-0.0508		0.000768		0.00594
		(0.0241)		(0.195)		(0.0566)		(0.0836)		(0.218)
Bank Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,515	1,515	1,515	1,515	1,515	1,515	1,515	1,515	1,515	1,515
Number of id_banco	101	101	101	101	101	101	101	101	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The following Table 7 summarizes results for our baseline model. Figures represent percentage change in outcome variables as capital inflows increase 10%. For the sake of interpretation and comparison, effects are shown unconditional and conditional on the funding strategy for a set of banks (fully funded by deposits, in the percentile 25th of the non-core ratio distribution, average, in the percentile 75th, and fully funded by non-core funds).

Table 7.

Summarized Results Baseline Model

Baseline Model										
(% change in outcome variables in response to a 10% increase in debt inflows)										
		Ln Net Loan Book	Ln Z Score	Ln Portfolio Risk	Ln Leverage Risk	Ln Capital	Ln Assets	Ln Liquid Assets	Ln Investment Assets	Ln Quarterly Profit
Unconditional on funding strategy		0,7	-1,0	0,3	-0,5	-0,4	-3,3	-1,4	3,1	-2,5
Conditional on funding strategy	Fully funded by deposits	0,7	-1,0	0,2	-0,5	-0,4	-3,3	-1,4	3,1	-2,5
	Percentile 25	0,7	-1,0	0,3	-0,5	-0,4	-3,2	-1,4	3,1	-2,5
	Mean	0,9	-1,0	0,3	-0,5	-0,5	-2,9	-1,5	3,1	-2,5
	Percentile 75	1,0	-1,0	0,3	-0,5	-0,5	-2,7	-1,6	3,1	-2,5
	Fully funded by non core	2,1	-1,1	0,5	-0,6	-0,7	-0,9	-1,9	3,1	-2,4

Table 8 displays summarized results by country of the main outcome variables: net loan book and Z score. The individual estimations are shown in Section 8 Annexes (Table 22 to Table 25). From these results, stands out the fact that, in response to capital inflows, the net loan book increases more in Peru than everywhere else, while decreases the most in Chile. Colombia and Mexico show average results, the former keeping the net loan book stable and the latter growing slightly (0.1%). Regarding the solvency risk, results are mixed with no evident or conclusive pattern.

Results in Table 8 are easily explained with the characterization of the Pacific Alliance financial systems done in Section 2. Peru is the only dollarized economy of the sample with the most open capital account and a financial system where banks are the main intermediaries. Conversely, Chile has a relatively open capital account but a financial system which is already quite deep and where pension funds, along insurance companies, are the main intermediaries rather than banks. Moreover, as mentioned earlier, this country has managed to shield itself from external shocks – no matter if they are positive or negative- which might also explain the negative coefficient on the loan book for this country.

Table 8.

Summarized Results Baseline Model by Country

Baseline Model										
(% change in outcome variables in response to a 10% increase in debt inflows)										
Country		Colombia		Chile		Mexico		Peru		Total	
		Ln Net Loan Book	Ln Z Score	Ln Net Loan Book	Ln Z Score	Ln Net Loan Book	Ln Z Score	Ln Net Loan Book	Ln Z Score	Ln Net Loan Book	Ln Z Score
Unconditional on funding strategy		0,0	-0,1	-0,4	-0,1	0,1	0,1	0,2	0,2	0,7	-1,0
Conditional on funding strategy	Fully funded by deposits	0,1	-0,2	-0,9	-0,1	0,0	0,1	0,4	0,3	0,7	-1,0
	Percentile 25	0,0	-0,2	-0,9	-0,1	0,0	0,1	0,3	0,2	0,7	-1,0
	Mean	0,0	-0,1	-0,5	-0,1	0,1	0,1	0,2	0,1	0,9	-1,0
	Percentile 75	0,0	0,0	-0,3	-0,1	0,2	0,1	0,2	0,1	1,0	-1,0
	Fully funded by non core	-0,5	0,8	1,1	-0,2	0,8	0,1	-0,5	-0,7	2,1	-1,1

a. Extensions to our baseline model

As mentioned in the introduction and literature review, we differentiate effects across non-core sources of funding since not all of them have the same characteristics neither lead banks to take the same decisions regarding loan supply and risk. To do this, we run the baseline model adjusting the interaction term between capital inflows and non-core funding sources. Instead of using the non- core ratio in this interaction, we rely on the ratio of bonds and credits, respectively, to total funding. Table 9 summarizes results for the net loan book and the Z score as outcome variables where we test the conditional effects of certain types of funding.

Table 9.

2SLS –Net loan book and Z score results according to type of funding

VARIABLES	2SLS Bonds Ln Net Loan Book	2SLS Credits Ln Net Loan Book	2SLS Bonds Ln Z Score	2SLS Credits Ln Z Score
Ln Debt Flows $_{t-1}$	0.0111	0.0729	-0.0895	-0.101
	(0.151)	(0.144)	(0.0867)	(0.0903)
Interaction term $_{t-1}$	0.652**	0.0556	-0.149	0.0185
	(0.291)	(0.0690)	(0.107)	(0.0335)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,515	1,515	1,515	1,515
Number of id_banco	101	101	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

First of all, it stands out that the procyclicality of the net loan book to capital inflows is driven mainly by banking institutions reliant on public debt (bonds). Although the share of private debt (credits) also leads to an increased net loan portfolio, we lack enough statistical evidence to affirm this mechanism operates. Second, evaluating the effect on the Z score, the banks that depend more on bonds are more exposed to insolvency risk than banks that use credits as sources of funding. To sum up, what is interesting from this exercise is that effects on both outcome variables are larger in the case of the interaction with the share of bonds in funding sources than with credits. This is especially clear-cut mainly on banks that are fully funded with bonds and credits, respectively. In terms of magnitudes, a 10% increase in portfolio inflows, increases the loan portfolio by 1.3% in the case of fully-funded banks by credits, against 1.5% in the case of fully-funded banks by bonds (a 20 basis points of difference). In terms of risk, the Z score decreases by 0.8% for credit-reliant institutions and by 1.2% for bond-reliant banks (a distance of 40 basis points).

Results obtained so far are aligned to what the literature has laid out. Public debt is considered the "unmonitored type of lending" since the risk-sharing instrument makes the multiple individual investors suffer from the free-riding syndrome, ending up executing poor monitoring. In comparison, using credits as funding sources implies a higher level of monitoring and a reduced leeway for borrower banks as lenders have access to private information and more incentives to execute adequate supervision. Our empirical evidence proves that bond concentration in the funding structure might lead to excessive lending and, therefore, increased exposure to insolvency risk.

The second extension to our baseline model has also been explored in the extant literature. In this exercise, we test whether capital flows' impact is differential upon banks' ownership, i.e., if they are foreign-owned -belonging to an international group- or domestic but foreign-funded. We do so by dividing the sample among foreign banks (42 institutions) and domestic entities (59 banks). Table 10 displays results for foreign banks, while Table 11 does the same for domestic banks.

Table 10.

2SLS – Net loan book and Z score results for foreign banks

	2SLS	2SLS	2SLS	2SLS
VARIABLES	**Bonds Ln Net Loan Book**	**Credits Ln Net Loan Book**	**Bonds Ln Z Score**	**Credits Ln Z Score**
Ln Debt Flows$_{t-1}$	0.881	1.418	-1.009	-1.306
	(4.692)	(9.583)	(4.846)	(8.196)
Interaction term$_{t-1}$		0.463		-0.256
		(1.720)		(1.480)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
Observations	630	630	630	630
Number of id_banco	42	42	42	42

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11.

2SLS – Net loan book and Z score results for domestic banks

	2SLS	2SLS	2SLS	2SLS
VARIABLES	Bonds Ln Net Loan Book	Credits Ln Net Loan Book	Bonds Ln Z Score	Credits Ln Z Score
Ln Debt Flows$_{t-1}$	0.00785	0.00984	-0.0335	-0.0379
	(0.0135)	(0.0136)	(0.0468)	(0.0466)
Interaction term$_{t-1}$		-0.0218		0.0486
		(0.0407)		(0.0348)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
Observations	885	885	885	885
Number of id_banco	59	59	59	59

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Differences across the two sub-samples dominate. From Table 10, it is evident that foreign banks drive loan procyclicality with respect to capital inflows yielding a positive -although insignificant- coefficient for the interaction term. Conversely, in domestic banks shown in Table 11, the interaction term is actually negative -while insignificant too-. To dimension these results, a fully funded bank with non-core sources increases the net loan portfolio by 18.8% if it is foreign and reduces it by 0.1% if it is domestic, when capital inflows increase 10%.[12] Along these lines, the Z score decreases in the light of capital inflows for foreign banks while increases for domestic institutions. However, only in the sub-sample of foreign banks, the fact of having non-core sources of funding amplifies the already increased risk exposure to insolvency.

The second extension results can also be rationalized with the present-day literature. On the one hand, loans' procyclicality to capital inflows in foreign banks is explained by the fact that these banks benefit disproportionately from current account deficits. In other words, they have a more direct credit channel –*the internal capital markets* within the holding company- where funds flow more smoothly impacting bank-level decisions significantly, such as loan supply. Meanwhile, domestic banks depend exclusively on external *capital markets*, which means the international credit channel is less direct nor straightforward. As a result, the latter type of institutions can only raise capital from outsiders (third parties) and by means of accessing capital markets, while the former can do it via insiders and without reaching capital markets.

[12] Our net loan portfolio results are in line with previous studies. Kamil and Rai (2010) report that in Latin America lending supply tightening from a global liquidity squeeze was mitigated thanks to foreign banks.

In line with net loan portfolio results, once again exposure to risk increases more for the banks that extend more credit in the light of capital inflows (foreign banks). In contrast, balance sheet's tight control in domestic banks allows them to even move farther away from the insolvency threshold. In the case of foreign banks, although they are risking their solvency position, it is possible that this effect is not as relevant, since they usually have support from the matrix company so they can play with the risk-returns trade-offs inherent to the financial intermediation activity. Evidence for this hypothesis is out of this work's scope, although it might be valuable to develop in further strands of this literature field.

The following Table 12 summarizes results for our model extensions. Figures represent percentage change in outcome variables as capital inflows increase 10%.

Table 12.

Summarized Results for Model Extensions

(% change in outcome variables in response to a 10% increase in debt inflows)

Type of bank according to Non-core Ratio	Type of non-core funding				Type of ownership			
	Public Debt	Private Debt	Public Debt	Private Debt	Foreign Banks	Domestic Banks	Foreign Banks	Domestic Banks
	Ln Net Loan Book		Ln Z Score		Ln Net Loan Book		Ln Z Score	
Fully funded by deposits	0,1	0,7	-0,9	-1,0	14,2	0,1	-13,1	-0,4
Percentile 25	0,1	0,7	-0,9	-1,0	14,2	0,1	-13,1	-0,4
Mean	0,5	0,8	-1,0	-1,0	14,9	0,1	-13,4	-0,3
Percentile 75	0,7	0,8	-1,0	-1,0	15,3	0,0	-13,7	-0,3
Fully funded by non core	1,5	1,3	-1,2	-0,8	18,8	-0,1	-15,6	0,1

The third and last extension to our baseline model revolves around the concern of financial instability generated by non-core foreign currency funding. Recalling the previous discussion of non-core funding, we mentioned that these type of funding is deemed as more volatile and unstable than traditional funding concentrated in public deposits. Yet, there is an additional layer of risk added when this type of non-core funding is denominated in foreign currency since it exposes financial institutions to foreign exchange risk mismatch. Nonetheless, it is also reasonable that banks funded by foreign currency denominated liabilities are more responsive to capital inflows as they would try to use these funds to reduce the foreign exchange mismatch by lending also in foreign currency. In

order to test these effects over the net loan book and risk conditional upon the currency denomination, we carry out the same estimation only for the Peruvian banks since they operate in the most dollarized economy, changing only the specification for the interaction term. Instead of using the non-core ratio (% of funding sources alternative to deposits) we replace it by the non-core foreign currency funding ratio (% of funding sources alternative to deposits which are denominated in foreign currency). Results of this estimation are shown in Table 26 (Section 8 Annexes). However, Table 13 depicts the percentage change in outcome variables when the currency denomination of non-core funds is taken into account. In line with the previous hypothesis, as banks fund themselves more with alternative sources to deposits which are denominated in foreign currency, the net loan book is more procyclical to capital inflows. It increases 0.5% when debt portfolio inflows rise 10%. However, contrary to what we initially suspected, the exposure to insolvency risk actually decreases as capital inflows increase, very likely because of the reduction in the exchange risk mismatch as more loans might be granted in foreign currency too.

Table 13.

2SLS – Net loan book and Z score results for Peru's Non-Core Funding in Foreign Currency

Baseline Model
(% change in outcome variables in response to a 10% increase in debt inflows)

Type of bank		Ln Net Loan Book	Ln Z Score
According to % of Non Core Ratio in Foreign Currency	Percentile 10	0,0	-0,2
	Percentile 25	0,0	-0,2
	Mean	0,3	0,2
	Percentile 75	0,5	0,5
	Percentile 90	0,5	0,6

5. Robustness tests

Our objective in this section is to prove the validity and strength of our results. We do this by altering some aspects of the sample or empirical strategy and checking if the same conclusions hold. With this in mind, we propose two robustness tests we will describe in the following lines.

First, we reinforce net loans' procyclicality to capital inflows by testing the impact on a variable that depends strictly on the expansion of credit: loan loss reserves. This refers to the stock of accumulated provisions banks hold to cope with credit risk associated with clients' probability of default. No matter the accounting regime, by regulation, all banks are compelled to provision according to the loan growth they record. Hence, we would expect that loan loss provisions increase in so far as the net loan portfolio expands too.

Table 14 confirms our initial thoughts: a bank increases its loan loss reserves by 1.16% with respect to a 10% increase in capital inflows. This effect is almost 50 basis points higher than the effect of capital inflows on the net loan portfolio which we must recall was of 0.7%. Since the impact on the stock of provisions is more prominent than the expansion of loans, we could interpret this as a prudent control of banks' risk from a management perspective. Lastly, once again, we can conclude with a 95% level of confidence that results are amplified as the share of non-core liabilities in the funding sources increases.

Table 14.

Loan Loss Reserves

VARIABLES	2SLS Ln Loan Loss Reserves	2SLS Ln Loan Loss Reserves
Ln Debt Flows$_{t-1}$	0.116	0.114
	(0.140)	(0.164)
Interaction term$_{t-1}$		0.249**
		(0.0993)
Bank Controls	Yes	Yes
Macro Controls	Yes	Yes
Year Fixed Effects	Yes	Yes
Quarter Fixed Effects	Yes	Yes
Observations	1,515	1,515
Number of id_banco	101	101

The second test we perform involves the measure we are using to assess funding strategies: the non-core ratio. Instead of using this indicator, we replace it by the loans-to-deposits (LTD) ratio which is also used to evaluate funding and more precisely liquidity risk by determining how large the share of loans covered by stable funding is. It is constructed as the quotient of dividing total loans in the total amount of deposits, therefore it should be positively associated with the non-core ratio, since a higher ratio implies that the bank relies more in alternative funding sources to deposits.

We perform this estimation running the two-stage least squares with the baseline model specification changing only the interaction term and expecting to get similar results to what we obtained using the non-core ratio (see Section 8 Annexes Table 27). Results shown in Table 15 prove our initial thoughts: debt portfolio inflows generate a more procyclical trend in credit insofar as the loans-to-deposits ratio increases. Regarding risk, debt portfolio inflows impact is negative, moving banks closer to the insolvency threshold, but the effect is not amplified when funding is taken into account since it has the same magnitude for all type of banks.

Table 15.

2SLS – Net loan book and Z score results using Loans-to-Deposits Ratio

Baseline Model
(% change in outcome variables in response to a 10% increase in debt inflows)

Type of bank according to Loans to Deposits Ratio	Ln Net Loan Book	Ln Z Score
Percentile 10	-0,3	-1,0
Percentile 25	0,2	-1,0
Mean	0,4	-1,0
Percentile 75	0,6	-1,0
Percentile 90	0,9	-1,0

a. Limitations

As with any other study, we are not exempt from limitations. First, a notable limitation when performing cross-country analyses revolves around data. In our case, although the balance of payments structure is homogenous across the countries considered and published by the same entity (IMF), bank-level information is heterogeneous. Each of the regulators publishes balance sheet data with a different structure and degree of disclosure. Although some balance sheet items are clearly equivalent such as total assets, liabilities, equity, etc., trying to disentangle effects in more particular accounts becomes difficult as Mexico, for example, had the lowest degree of disclosure.

This implication leads us to our second point. The non-core ratio, one of our main variables of interest, was defined in the best way we could, considering the information available and the differences in disclosure. Nonetheless, we acknowledge that there are several ways to construct it and that there is no consensus on the literature of how it should be assessed. Since our primary concern is to assess the funding strategy roughly (mainly categorizing funding sources between deposits and non-deposits), we are confident we have taken an adequate approach throughout this investigation.

Another limitation is that we can only talk about causality at the local level as our identification strategy relies on an instrumental variables framework. That means we are capturing only those debt capital inflows which were actually influenced by the VIX index. Lastly, we have to recall that these results hold only for the countries (Pacific Alliance members) and for the time window selected (2016-2019).

6. Conclusions

In synthesis, as debt portfolio inflows to the Pacific Alliance member countries increase by 10%, the net loan portfolio expands on average by 0.7% keeping everything else constant. Testing whether these effects are amplified or mitigated by banks' funding structure choice, we find out that a higher reliance on non-core liabilities amplifies -rather than mitigates- the former results. To size the differential impact of funding, in the same event of a 10% increase in debt capital inflows, a fully funded bank by non-core sources expands its net loan book by 2.1% *ceteris paribus*, meanwhile a bank with an average share of non-core liabilities increases it by 0.9%. This is a 120 basis points of difference suggesting that banks' decision to issue credit in the light of capital inflows is to a large extent determined by the current funding strategy they use, as it makes it more (or less) easy to tap the enlarged pool of resources made available thanks to foreign creditors.

Unlike net loan portfolio results, debt capital inflows seem to have no significant impact on banks' solvency risk in the countries considered. However, while there is not enough evidence to conclude about this effect, we interpret it as a 1.0% decrease in the Z-score –an increased exposure to insolvency risk- in response to a 10% increase in debt portfolio inflows. Decomposing the solvency risk between portfolio and leverage risk, our findings showcase that if anything the potential negative impact of capital inflows on solvency risk comes through leverage risk as banks decrease their capital levels. Once again, a more concentrated funding structure towards non-deposits amplifies the initial results by actually increasing the banks' overall exposure to insolvency risk. While a fully-funded bank with non-core liabilities decreases its Z-score by 1.0% *ceteris paribus* in the event of a 10% increase in debt capital inflows, an average bank in terms of funding decreases it by 1.1% (a 1 basis point of difference). This is in line with previous results setting out that banks with a higher share of non-core sources of funding expanded more their loan portfolios and this could possibly result in an added layer of risk to their ongoing operation.

As extensions to our baseline model, we test differential impacts across banks' type of non-core funding (public debt versus private debt) and type of ownership (foreign-owned versus domestic). Our findings point out that among non-core sources of funding, public debt is more procyclical to capital flows as it is considered the "unmonitored lending" where investors free ride others' monitoring activities. Thus, this might lead to excess lending and increased levels of insolvency risk. On the other hand, we encounter that foreign-owned banks play an important role in the international credit channel, also driving the procyclicality of credit to capital flows as these banks have a more direct channel through which funds flow: *the internal capital markets* within the holding company.

To sum up, throughout this investigation we examined in tandem two sources of macroeconomic imbalances (capital flows and financial stability), and provided evidence for the link between this two: the international credit channel. We disclosed that in open economies, such as the ones we analyzed, debt capital inflows permeate financial systems through the liabilities side of banks' balance sheets, as they are used as alternative funding sources to deposits. As a consequence of this, we concluded banks were able to increase the size of their high-yielding assets, mainly the loan and the investment portfolio, while decreasing their most liquid assets (cash and cash equivalents). However, we argue that they did this without seriously risking the solvency profile - as we had not enough evidence to affirm it was otherwise-. If anything, we discovered that leverage risk exposure increases due to the decrease in capital levels, certainly in some banks more than others, and that there could be some potential stress in the liquidity levels at a systemic dimension.

Another of our key objectives in this investigation was to assess funding strategies along a complete "standard" cycle of capital inflows. This exercise yielded that the non-core sources of funding were the main drivers behind net loans' procyclicality to capital inflows, as lending increased more when this factor was taken into account. However, compared to the literature and the empirical evidence regarding this type of funding, the estimations were far less negative. Not only credit grew, but also provisions increased at a higher rate than the loan growth meaning a prudent risk approach from the management perspective.

Therefore, the final balance is in our view mixed since non-core funding enhances lending dynamics without causing excessive loan growth and aiding financial system deepening which is key in emerging markets. Taking into account the risks analyzed, there is no conclusive evidence over the risk of insolvency, however there could be some evidence of potential stress in liquidity levels.

Regarding the extensions to our baseline model, we were able to exploit successfully the dataset's granularity to disentangle the micro effects of capital flows. In this sense, we exposed that the impact of debt portfolio inflows was not homogenous across the type of banking institution (foreign-owned vs. domestic) and of non-core liabilities (bonds vs. credits), being the resulting effects -in terms of loan growth and solvency risk exposure- larger in the bonds case vs. credits and in the foreign-owned vs. domestic banks. Both conclusions are not concerning for us at a systemic level. First, because the use of bonds as funding sources is still not too deep in Latin America and second, because the share of foreign-owned and domestic banks in these countries is approximately "balanced" so contrasting results might offset each other.

Lastly, our main takeaway from this document is that debt capital inflows altered bank-level credit supply decisions, most notably in banks with a higher share of non-deposit funding, which were more responsive to the entry of foreign funds by increasing more their lending. Interestingly, this increased size of their loan portfolio not necessarily came along with a stress of their solvency nor liquidity position, although these two risks must be important to watch cautiously at a systemic level in order not to risk financial stability in periods of large capital inflows.

References

Altunbas, Y., et al. (2011). Bank Risk During the Financial Crisis—Do Business Models Matter? ECB Working paper series No. 1394; Frankfurt: European Central Bank.

Arias, F., et al. (2013). Do the different types of capital flows respond to the same fundamentals and in the same degree? Recent evidence for emerging markets. Bank of International Settlements, Basel.

Baskaya, Y., et al. (2017). Capital flows and the international credit channel. Journal of International Economics 108 (2017) S15–S22.

Batini, N., et al. (2020). IMF Advice on Capital Flows to Latin America. Background Paper of the Independent Evaluation Office, BP/20-02/06, International Monetary Fund.

Bekaert, G., et al. (2013). Risk, Uncertainty and Monetary Policy. Working Paper Series, No 1565, European Central Bank.

Cairns, J., et al. (2007). Exchange rates and global volatility: implications for Asia-Pacific currencies. Bank of International Settlements (BIS) Quarterly Review.

Calomiris C. and Kahn C. (1991). The Role of Demandable Debt in Structuring Optimal Banking Arrangements. American Economic Review, 81(3), 497-513.

Calvo, G., et. al. (1993). Capital Inflows and Real Exchange Rate Appreciation in Latin America: The Role of External Factors. IMF Staff Papers, 40, (1), 108-151.

Calvo, G. (1998). Capital Flows and Capital Market Crises: The Simple Economics of Sudden Stops. Journal of Applied Economics (1), November, 35-54.

Calvo, G. and Reinhart, C. (2000). When Capital Inflows Come to a Sudden Stop: Consequences and Policy Options. MPRA Paper, University Library of Munich, Germany.

Cetorelli, Nicola and Linda S. Goldberg (2009) "Banking Globalization and Monetary Trans- mission" Forthcoming in Journal of Finance.

Cetorelli, N. and Goldberg, L. (2011) Global Banks and International Shock Trans mission: Evidence from the Crisis. IMF Economic Review, 59 (1), 41-76.

Cetorelli, N. and Goldberg, L. (2012a) Follow the money: Quantifying domestic effects of foreign bank shocks in the great recession. American Economic Review, 102 (3), 213-218.

Cetorelli, N. and Goldberg, L. (2012b) Follow the money: Quantifying domestic effects of foreign bank shocks in the great recession. Journal of International Economics, 88 (2), 299-311.

Damar, H., Meh, C. and Terajima, Y. (2012). Leverage, Balance Sheet and Wholesale Funding. 16th International Panel Data Conference and 44th Annual Conference of the Canadian Economics Association.

Coudert, V., et al. (2011). Exchange rate volatility across financial crises. Journal of Banking & Finance, 35, 3010–3018.

Demirgüc-Kunt, A. and Huizinga, H. (2010). Bank Activity and Funding Strategies: The Impact on Risk and Returns. Journal of Financial Economics, 98, 626-650.

Diamond, D. and Dybvig, P. (1983). Bank Runs, Deposit Insurance, and Liquidity. Journal of Political Economy, 91(3), 401-19.

Diamond, D. (1984). Financial intermediation and delegated monitoring. The Review of Economic Studies, 51 (3), 393-414.

Dinger, V. and Te Kaat, D. (2017). Cross-Border Capital Flows and Bank Risk-Taking. Journal of Banking and Finance. Volume 117, August 2020, 105842.

Fratzscher, M. (2012). Capital flows, push versus pull factors and the global financial crisis. Journal of International Economics, 88 (2), 341-356.

Ghosh, A, et al. (2014). Regulating Capital Flows at Both Ends: Does it Work? Working Paper 14/188, International Monetary Fund (IMF).

Hahm, J., et al. (2011a). Macroprudential Policies in Open Market Economies. Prepared for the 2011 Reserve Bank of San Francisco Asia Economic Policy Conference.

Hahm, J., et al. (2011b). Non-Core Bank Liabilities and Financial Vulnerability. Preliminary draft of paper for the Federal Reserve Board and JMCB Conference on "Regulation of Systemic Risk"", Washington D.C.

Hamann, F., et al. (2014). Leverage Pro-cyclicality and Bank Balance Sheet in Colombia. Ensayos sobre Política Económica 32 (73), 50-76.

International Monetary Fund, (2018). Chile Article IV- Staff Report. IMF Country Report No. 18/311.

International Monetary Fund, (2019). Colombia Article IV- Staff Report. IMF Country Report No. 19/106.

International Monetary Fund, (2019). Mexico Article IV- Staff Report. IMF Country Report No. 19/336.

International Monetary Fund, (2020). Peru Article IV- Staff Report. IMF Country Report No. 20/3.

Ivashina, V. and Scharfstein, D. (2010). Bank Lending during the Financial Crisis of 2008. Journal of Financial Economics, 97, 319-38.

Jubinski, D. & Lipton, A. (2012). Equity volatility, bond yields, and yield spreads. The Journal of Futures Markets, 32 (5), 480 – 503.

Kalemli-Özcan, S. (2020). US monetary policy and international risk spillovers. Proceedings of the Jackson Hole Symposium, Forthcoming. 1, 6, 8.

Kamil, H. & Rai, K. (2010). The Global Credit Crunch and Foreign Banks' Lending to Emerging Markets: Why Did Latin America Fare Better? Working Paper 10/02, International Monetary Fund (IMF).

Kim, H., et al. (2012). Monetary Aggregates and the Central Bank's Financial Stability Mandate. Paper prepared for the Federal Reserve Conference Central Banking Before, During and After the Crisis.

Köhler, M. (2014). Which banks are more risky? The impact of loan growth and business model on bank risk-taking. Discussion Paper No 33/2012, Deutsche Bundesbank.

Lane, P. and McQuade, P. (2013). Domestic Credit Growth and International Capital Flows. ECB Working Paper No. 1566.

Mendoza, E. and Terrones, M. (2008). An Anatomy of Credit Booms: Evidence from Macro Aggregates and Micro Data. National Bureau of Economic Research, Working Paper 14049.

Ostry, J., et al. (2011). Managing Capital Inflows: What Tools to Use? IMF Staff Discussion Note SDN/11/06.

Rigobon, R. & Sack, B. (2003). Measuring the Reaction of Monetary Policy to the Stock Market. The Quarterly Journal of Economics, 118 (2), 639 – 669.

Sharpe, S.A. (1990). Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships. Journal of Finance, 45(4), 1069-87.

Steiger, D. & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. Econometrica, 65(3), 557-586. Retrieved from https://www.jstor.org/stable/2171753.

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. National Bureau of Economic Research. Retrieved from https://www.nber.org/papers/t0284.pdf.

Rajan, Raghuram G. (1992). Insiders and outsiders: the choice between informed and arm's- length debt. Journal of Finance, 47, 1367-1400.

Passari, E. & Rey, H. (2015). Financial Flows and the International Monetary System. The Economic Journal, 125 (584), 675-698.

8. Annexes

Table 16.

OLS and FE Within - Net Loan Book

VARIABLES	OLS Ln Net Loan Book	FE WITHIN Ln Net Loan Book	OLS Ln Net Loan Book	FE WITHIN Ln Net Loan Book
Ln Debt Flows$_{t-1}$	0.0171	0.00283	0.00965	-0.00775
	(0.0122)	(0.00558)	(0.0133)	(0.00806)
Interaction term$_{t-1}$			0.0492	0.0700***
			(0.0353)	(0.0258)
Bank Controls	No	Yes	No	Yes
Macro Controls	No	Yes	No	Yes
Year Fixed Effects	No	Yes	No	Yes
Quarter Fixed Effects	No	Yes	No	Yes
Country Fixed Effects	No	Yes	No	Yes
Observations	No	Yes	No	Yes
R-squared	1,515	1,515	1,515	1,515
Number of id_banco	0.918	0.866	0.918	0.867

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 17.

OLS and FE Within – Z Score

VARIABLES	OLS Ln Z Score	FE WITHIN Ln Z Score	OLS Ln Z Score	FE WITHIN Ln Z Score
Ln Debt Flows$_{t-1}$	-0.0142**	-0.00263	-0.0115	-0.00381*
	(0.00667)	(0.00159)	(0.00728)	(0.00212)
Interaction term$_{t-1}$			-0.0183	0.00779
			(0.0194)	(0.00888)
Bank Controls	No	Yes	No	Yes
Macro Controls	No	Yes	No	Yes
Year Fixed Effects	No	Yes	No	Yes
Quarter Fixed Effects	No	Yes	No	Yes
Country Fixed Effects	No	Yes	No	Yes
Observations	No	Yes	No	Yes
R-squared	1,515	1,515	1,515	1,515
Number of id_banco	0.608	0.782	0.608	0.782

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18.

First Stage Regression: Critical Values Cragg- Donald Wald F Statistic

Model without interaction term		Model with interaction term	
Weak identification test (Cragg-Donald Wald F statistic):	50.53	Weak identification test (Cragg-Donald Wald F statistic):	25.210
Stock-Yogo weak ID test critical values: 10% maximal IV size	16.38	Stock-Yogo weak ID test critical values: 10% maximal IV size	7.03
15% maximal IVsize	8.96	15% maximal IVsize	4.58
20% maximal IV size	8.66	20% maximal IV size	3.95
25% maximal IV size	5.53	25% maximal IV size	3.63
Source: Stock-Yogo (2005). Reproduced by permission.		Source: Stock-Yogo (2005). Reproduced by permission.	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 19.

2SLS –Net loan book and Z score with lagged portfolio inflows (t-2)

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-2}$	0.0113	-0.00148	-0.0282*	-0.0305**
	(0.0210)	(0.0226)	(0.0149)	(0.0143)
Interaction term$_{t-2}$		0.114***		0.0208
		(0.0437)		(0.0218)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	No	Yes	No	Yes
Observations	1,414	1,414	1,414	1,414
Number of id_banco	101	101	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 20.

2SLS –Net loan book and Z score with lagged portfolio inflows (t-3)

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-3}$	0.0201	0.00582	-0.00153	-0.00441
	(0.0166)	(0.0141)	(0.00573)	(0.00549)
Interaction term$_{t-3}$		0.0980*		0.0198
		(0.0532)		(0.0206)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	No	Yes	No	Yes
Observations	1,313	1,313	1,313	1,313
Number of id_banco	101	101	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21.

2SLS –Net loan book and Z score with lagged portfolio inflows (t-4)

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-4}$	-0.0179	-0.0207	-0.00614	-0.00941
	(0.0211)	(0.0210)	(0.00775)	(0.00742)
Interaction term$_{t-4}$		0.0190		0.0219
		(0.0184)		(0.0163)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	No	Yes	No	Yes
Observations	1,212	1,212	1,212	1,212
Number of id_banco	101	101	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 22.

2SLS –Net loan book and Z score results for Colombia

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-1}$	-0.000975	0.00752	-0.00890	-0.0244**
	(0.00262)	(0.00461)	(0.0107)	(0.0120)
Interaction term$_{t-1}$		-0.0573		0.104**
		(0.0415)		(0.0433)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Observations	390	390	390	390
Number of id_banco	26	26	26	26

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 23.

2SLS –Net loan book and Z score results for Chile

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-1}$	-0.0428	-0.0875	-0.0137	-0.0112
		(0.0654)	(0.0208)	(0.0239)
Interaction term$_{t-1}$	(0.0533)	0.201*		-0.0110
		(0.117)		(0.0307)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Observations	390	390	390	390
Number of id_banco	26	26	26	26

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24.

2SLS –Net loan book and Z score results for Mexico

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-1}$	0.00981	0.00252	0.0105	0.0105
	(0.00811)	(0.00732)	(0.00708)	(0.00686)
Interaction term$_{t-1}$		0.0753***		-0.000384
		(0.0182)		(0.0216)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Observations	540	540	540	540
Number of id_banco	36	36	36	36

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25.

2SLS –Net loan book and Z score results for Peru

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-1}$	0.0240	0.0356**	0.0171	0.0295
	(0.0147)	(0.0167)	(0.0311)	(0.0367)
Interaction term$_{t-1}$		-0.0888**		-0.0955
		(0.0447)		(0.0604)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Observations	195	195	195	195
Number of id_banco	13	13	13	13

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 26.

2SLS –Net loan book and Z score results for Peru Non-Core Funding in Foreign Currency

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-1}$	0.0240	-0.00167	0.0171	-0.0190
	(0.0147)	(0.0280)	(0.0311)	(0.0117)
Interaction term$_{t-1}$		0.0618		0.0868**
		(0.0503)		(0.0398)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Observations	195	195	195	195
Number of id_banco	13	13	13	13

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 27.

2SLS –Net loan book and Z score results using Loans-to-Deposits Ratio

VARIABLES	2SLS Ln Net loan book	2SLS Ln Net loan book	2SLS Ln Z Score	2SLS Ln Z Score
Ln Debt Flows$_{t-1}$	0.0678	-0.0351	-0.102	-0.0964
	(0.142)	(0.141)	(0.0912)	(0.0910)
Interaction term$_{t-1}$		0.0775***		-0.00456***
		(4.87e-05)		(1.48e-05)
Bank Controls	Yes	Yes	Yes	Yes
Macro Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Quarter Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,515	1,515	1,515	1,515
Number of id_banco	101	101	101	101

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 28.

Descriptive Statistics – Variables in Baseline Model Specification

Depend Variable	Obs.	Mean	Std. Dev.	Percentile 25	Median	Percentile 75
Ln Net Loan book	1.616	12,01	4,64	9,53	13,25	15,38
Ln Z score	1.616	3,31	1,16	3,02	3,54	4,03
Control Variables						
Bank-Level Regressors						
Non-Core Liabilities Ratio	1.616	42,39	30,05	22,43	35,25	53,28
Ln Assets	1.616	12,80	4,37	10,77	13,78	15,76
Ln Capital	1.616	10,90	3,92	8,60	11,85	13,65
Ln Past Due Loans Ratio	1.616	10,56	5,70	9,34	11,63	14,58
Macroeconomic Regressor						
Gross portfolio debt inflows	1.616	2025	3133	195	1478	3243
GDP growth rate (Y-o-Y)	1.616	6,25	2,24	4,82	6,69	7,52
Monetary policy rate	1.616	4,95	2,04	3,00	4,25	7,00
Consumer price index growth rate	1.616	3,72	1,62	2,60	3,23	4,65

ASOBANCARIA