A Theory of Payments-Chain Crises*

Saki Bigio† **Esteban Méndez**‡ **Diana Van Patten**§

First Released Draft: December 2021

Abstract

This paper introduces an endogenous network of payments chains into a business cycle model. Motivated by evidence of linked payments in Costa Rica, we present a model where agents order production in bilateral relations. Some payments are executed immediately. Other payments, chained payments, are delayed until other payments are executed. Because production starts only after orders are paid, chained payments induce production delays. In equilibrium, agents choose the amount of chained payments given interest rates and access to internal funds or credit lines. This choice determines the payments-chain network and aggregate total-factor productivity (TFP). The paper characterizes equilibrium dynamics and their innate inefficiencies. Agents internalize the direct costs of their payment delays, but do not internalize the costs induced onto others. This externality produces novel policy insights and rationalizes permanent reductions in TFP under excessive debt.

Keywords: Payments, Networks, Business Cycles

JEL Classification: E32, E42, G01

*We thank Luigi Bocola, Ricardo Lagos, Jennifer La'O, Ezra Oberfeldt, Michael Peters, Guillaume Rocheteau, Gary Richardson, Pierre-Olivier Weill, and Yuliy Sannikov as well as seminar participants at UC Irvine, UCLA, CEMLA, CESifo, for early discussions. Sebastian Merkel provided a thoughtful discussion. We are in debt with Ken Miyahara and Luis Yepez for their outstanding assistance, multiple discussions, and their creativity in tackling proofs. The views expressed herein are those of the authors and do not necessarily represent the views of the Central Bank of Costa Rica (BCCR). Results have been reviewed by the BCCR to ensure no confidential information is disclosed.

†UCLA and NBER. Email: sbigio@econ.ucla.edu. Address: 315 Portola Plaza Los Angeles, CA 90095.

‡Central Bank of Costa Rica. Email: mendezce@bccr.fi.cr. Address: Avenida Central y Primera, Calles 2 y 4, San José 10101.

§ Yale University and NBER. Email: diana.vanpatten@yale.edu. Address: 165 Whitney Avenue, New Haven, CT 06511.

Says A: *I could use some of B's goods; but I have no cash to pay for them until someone with cash walks in here!*
Says B: *I could buy some of C's goods, but I've no cash to do it 'till someone with cash walks in here.*

From the book Stamp Scrip, Irving Fisher, 1933

1. Introduction

During financial crises, there are visible declines in credit variables. Though harder to measure, there is also a widespread perception that the normal, seamless flow of transactions and goods slows down. In ways still not fully understood, productive resources appear to remain idle—not because of technical constraints or lack of demand, but because agents must wait for others to pay, who in turn are waiting on others.

A simple example illustrates the idea. Consider a general contractor hired to build a commercial property. The contractor relies on a network of subcontractors—electricians, plumbers, and carpenters—who agree to begin or continue work only after receiving partial payment. The contractor also places orders with material suppliers, who may customize some of the materials and are unwilling to ship them until payment is completed or, at minimum, a proof of funds is provided. However, the contractor is waiting for a payment from the project's developer, who is waiting for a loan installment or the proceeds from the sale of another property. During crises, banks may delay loan payments and property sales may slowdown, disrupting the developer's funding sources, and therefore, the entire chain: the contractor cannot pay the electrician or the supplier; the electrician postpones work; the supplier holds off on shipping; and the project stalls. The slowdown is not due to a lack of labor, materials, or technical capacity, but rather to each party waiting to be paid by someone else. These cascading delays can leave productive resources idle, and the national accounts will record the outcome as a decline in productivity.

Despite how natural they seem, such transactional delays are rarely featured in macroeconomic models. Production and trade are typically centralized, with goods exchanged on spot markets and payment frictions largely abstracted away. While supplier financing and credit constraints are sometimes modeled, the usual assumption is binary: agents are either constrained or unconstrained. What is missing is a tractable framework in which economic activity can stall simply because agents must wait for

other transactions to clear before making their own payments. In such settings, even solvent and willing producers may remain idle—not because they face traditional financial constraints, but because payment is delayed somewhere upstream.

This paper proposes a model of payments chains and studies their macroeconomic implications, particularly during financial crises. Developing a theory of payments-chain disruptions is important for understanding observed dynamics in productivity. Since the onset of modern business cycle analysis, economists have argued that fluctuations in total factor productivity (TFP) often reflect credit-market conditions (e.g., Summers, 1986). This view becomes especially salient during crises in developing economies, where contractions are frequently driven by large declines in measured TFP. Financial crisis models often explain these declines through input misallocation. However, replicating large observed TFP drops remains challenging, particularly given the relatively smooth labor flows during such episodes.[1]

Payment-chain disruptions offer an alternative explanation for this co-movement. Delays in the timing of payments can propagate through production networks, leading to widespread idle time and reduced output, even when factor inputs are available. Through this mechanism, it is not the quantity or allocation of inputs that constrains production, but the timing: the slowdown of payments delays the utilization of resources.

These forces are salient empirically. Section 2 leverages detailed microdata on the universe of interbank transfers in Costa Rica, along with transaction-level details for each sale in the economy, to present the first evidence that many firms depend on payments made to them to execute further payments. In particular, about one-half of firms consistently execute payments on the exact day they receive a payment. We further provide on the impact of chained payments given a financial disruption by exploiting a natural experiment in which bank accounts were suddenly frozen for a subset of firms, and analyze how the the consequences of this payment disruptions largely depend on whether impacted firms were part of a payments-chain.

[1]The puzzle of large drops in TFP and capital utilization alongside small labor movements was first noted by Meza and Quintin (2005). It has since appeared in several contexts: Meza and Quintin (2007), Mendoza (2010), Oberfield (2013), and Karabarbounis et al. (2021) document similar patterns in the Chilean banking crisis of the early 1980s, the Mexican and East Asian crises of the 1990s, and the recent Greek crisis. These studies cite numerous additional examples.

Guided by this evidence, and to model payment-chain disruptions and their impact on production, we introduce a payment-chain production network in Section 3. This framework explicitly links the timing of economic transactions to the timing of production. In the network, production takes place through random bilateral relationships, where customers place orders with suppliers. Some of these are spot orders, which are paid upfront and trigger immediate production. Others are chained orders, which are only paid once a preceding order has been paid. Since funds are transferred with a delay, chained orders introduce production lags that can cascade through the network. Chained transactions correspond to situations like the construction example above: a subcontractor, such as a carpenter, is scheduled to work but waits to begin work until the general contractor receives payment from the developer. Each chained payment depends on the clearance of prior payments along the chain.

A payments-chain network consists of many chains, each representing a linked transaction sequence. The first payment in each chain is a spot payment, financed by resources external to the network—such as savings, working capital credit, or accounts payable—and initiates the sequence. All subsequent payments are chained and rely on the completion of earlier payments in the chain. Chained orders, therefore, represent contractual obligations whose fulfillment is contingent on the settlement of prior transactions.

A key implication of the model is that the further downstream an order lies in a payment chain, the longer the production delay, since each chained transaction introduces a lag in transaction times. The fraction of chained orders in the economy determines the overall structure and length of these chains: as the share of chained orders rises, the distribution of chain lengths shifts to longer chains, and average waiting times increase. These delays, in turn, reduce aggregate productivity. Specifically, measured TFP in this environment is given by: $\mathcal{Y}(\mu;\delta) = (1-\mu) + \mu\frac{\delta-\delta\mu}{\mu-\delta\mu}\ln\left(\frac{1-\delta\mu}{1-\mu}\right) < 1$, where μ is the fraction of chained orders relative to total orders and δ is a parameter that captures production/transactions delays.

The payments-chain network developed in Section 3 is a stand-alone, mechanical production block where transactions are taken as given. However, it is portable to a range of contexts where agents decide to make spot or chained transaction agreements.

In the second part of the paper, Section 4, we embed this network into a tractable, deterministic business-cycle model.

In that setting, a new payments-chain network is formed in each period. There is a natural borrower and a natural saver which, for simplicity, we model as households. By construction, the saver always has funds to place spot orders. By contrast, the borrower carries outstanding debt. The borrower can obtain funds to place spot orders by borrowing from exogenous credit lines. However, her credit lines are limited by her outstanding debt. When outstanding debt impairs her access to short-term funds, she can still place chained orders. However, placing chained orders is privately costly, as goods bought through chained orders are more expensive. Critically, whereas borrowers internalize that chained orders are more expensive, they do not internalize that by placing chained orders, they delay other transactions.

The formula for TFP $\mathcal{Y}$ showcases how credit conditions translate into TFP declines. When short-term funds are limited, borrowers place chained orders to maintain a consumption level. This tightening of short-term funds increases the fraction of chained orders, μ. Measured TFP falls as labor inputs remain idle.

In the environment, the evolution of debt impacts TFP by impairing credit lines. In turn, TFP influences the desire to accumulate debt. There are three critical regions of the state space. For debt levels below an efficient-debt threshold, the economy is in a steady state without delays. For moderate debt levels, the economy features a *temporary* payments-chain crisis. Production delays last until the economy transitions toward an efficient steady state. For such moderate debt levels, crises are only temporary because borrowers have incentives to repay debts—to access short-term funding in the future. When debt is above a threshold, the economy features hysteresis, permanent payments-chain crises. Hysteresis occurs when borrowers decide to maintain high debt levels because the benefits of deleveraging happen too late into the future and require an excessive sacrifice of current consumption. The economy remains with permanent production delays in this debt overhang region of the state space.

The environment leads to novel policy implications presented in Section 5. We study a Ramsey planner that has instruments to influence the path of debt, but cannot tax spot and chained expenditures differentially. This Ramsey planner respects the pay-

ments technology but, as opposed to households, internalizes the effects of chained orders on TFP. The exercise illustrates that transitions are constrained inefficient for two reasons: savers spend too little in spot orders, and borrowers spend excessively in chained orders. Because the inefficiency is two-sided, debt may be excessively high or excessively low relative to the social optimum during a payments-chain crisis.

We also revisit fiscal multipliers. We allow the government to make spot expenditures or expenditures chained to tax receipts. Although all forms of government spending are waste, government spending can produce positive multipliers during a payment-chain crisis. A novel insight is that multipliers are positive only if the government makes spot expenditures. If government expenditures are chained to tax income, they are detrimental. The reason for positive multipliers is not the stimulus of aggregate demand, a conventional view regarding the benefits of fiscal policy. Rather, government expenditures stimulate output by speeding up payments, a monetarist reinterpretation.

Literature Review. The literature on financial crises is vast. This paper connects with theories that underscore the sharp declines in aggregate TFP. The link between financial crisis and TFP is not at all obvious because financial crises can manifest themselves through distorting labor-market outcomes, not productivity. One branch of the literature explains declines in aggregate TFP through increased misallocation—see Pratap and Urrutia (2012) or Oberfield (2013). However, a common finding is that heterogeneity can only partially explain TFP declines once models are disciplined with data on input use and heterogeneity. Here, the channel is utilization. Other studies also explain declines in TFP as aggravated financial conditions increase the cost of utilizing capital—see Meza and Quintin (2007). The mechanism here is different: financial conditions impact TFP through a slowdown of payments.

Beyond the focus on financial crises, the paper falls at the crossroads of several areas. Namely, the monetary-payments literature, the economic-networks literature, and the literature on aggregate-demand externalities. The issue of how payments instruments affect production is a classic theme: Lucas and Stokey (1987) analyzes a stochastic cash-in-advance economy; Kiyotaki and Wright (1989) studies trade with indivisible tokens; Lagos and Wright (2005) a model with divisible money and explicit trading ar-

rangements.[2] Recent work focuses on how the distribution, and not the instruments per se, affect production—see Lippi et al. (2015), Rocheteau et al. (2016), and Brunnermeier and Sannikov (2017). In common with this literature, the distribution of funding affects allocations. The main distinction is that we focus on delays in sequential payments.

Sequential payments appear in many other studies. The payments-chain network is inspired by the model in Kiyotaki and Moore (1997). Kiyotaki and Moore (1997) focus on finding micro-foundations of production disruptions in a single chain. The contribution here is to present a network of transactions whose structure depends on payment decisions. Furthermore, we embed the payments-chain network into a standard business-cycle model. Other models of sequential payments include Townsend (1980), which studies sequences of payments with spatial separation, Freeman (1996) and Green (1999), which study sequential transactions in overlapping generation environments, La'O (2015), which studies a circular flow of transactions, or Guerrieri and Lorenzoni (2009), which studies sequential transactions in a Lagos and Wright-type environment. Recent work by Hardy et al. (2022) and Bocola (2022) contrasts payments funded externally against trade credit.[3] Relative to these papers, there are two distinctions: here, transactions are formed in a network, and the network is endogenous to expenditure-savings decisions. The importance of this body of theoretical work is substantiated by a body of empirical evidence found in a number of recent papers: Boissay and Gropp (2007), Jacobsen (2015), Barrot (2016), and Costello (2020) among many others.

With respect to the economic networks literature, the paper connects with models with endogenous network formation. The contribution relative to that literature is modest, as network formation is not strategic. By contrast, in Oberfield (2018), a network is formed through strategic partnerships. In Kopytov et al. (2022) and Elliott et al. (2022) firms form strategic links, being aware of possible supply-chain breakdowns. Here, the network is randomly formed, but the distribution of chains is endogenous to financial decisions. Like in Elliott et al. (2014), Alvarez and Barlevy (2021)

[2] See Shi (1997); Lagos et al. (2011); Lagos and Rocheteau (2009); Li et al. (2012); Nosal and Rocheteau (2011); Rocheteau (2011) for many other directions in that area.

[3] See also Biais and Gollier (1997).

and Taschereau-Dumouchel (2022), there are externalities here too. In those models, externalities occur when individual defaults provoke subsequent defaults. Here, externalities occur through payment delays. Bigio and La'O (2013) considers the propagation of financial shocks that induce misallocation in a production network. Here, there is no misallocation in the production network, but there are production delays.

Finally, the paper connects with models of aggregate demand externalities. An early model of these externalities is Diamond (1982) where, via search, consumption decisions affect output. In most of the literature, demand externalities result from nominal rigidities. There has been a recent interest in coupling nominal rigidities with financial constraints—for example, Eggertsson and Krugman (2012) and Guerrieri and Lorenzoni (2017). Recent papers have further introduced sequential transactions into environments with nominal rigidities—for instance, Woodford (2022) and Guerrieri et al. (2022). In those models, demand externalities occur when agents cut back on any form of expenditure. The nature of demand externalities here is different. In particular, the type of expenditures by the private or public sector matters: spot orders may stimulate output, but chained orders depress it. The demand externalities provoked by the slowdown of payments is part of a classic narrative. Almost a century apart, the opening quotation taken from Fisher (1933) and the renowned ``babysitting co-op" analogy of Krugman (1998) belong to that same tradition. The rest of the paper attempts to provide a convenient analytic formulation.

2. Empirical Motivation

This section documents empirical evidence on payment chains, their widespread prevalence, and their substantial effects on firm-level productivity.

Data. We leverage two exceptionally rich data sources. The first is SINPE—Costa Rica's National Electronic Payments System—which processes and settles every interbank transfer in the country.[4] In 2024 alone, SINPE recorded over 60 million transfers,

[4]SINPE is an acronym for Costa Rica's National Electronic Payments System (*Sistema Nacional de Pagos Electrónicos*, and is operated by the Central Bank of Costa Rica to process and settle all electronic interbank transactions.

capturing, for each transaction, the precise timestamp, unique identifiers of sender and receiver, the banks involved, and the exact amount transferred. This unparalleled coverage yields a complete universe of interbank activity for 281 ,324 distinct firms, with an average transaction size of USD 3,158.

The second dataset derives from Costa Rica's VAT (Value-Added Tax) reporting system, which mandates an electronic invoice for every formal sale. For each invoice, the registry records the sale price, product code, and quantity sold, and flags whether payment was settled on the spot or was deferred. Deferred payments are recorded, in particular, because as long as the seller has not received full payment, classifying the payment as deferred allows her to defer the VAT payment on the unpaid amount until the payment is due, for up to 90 days. For deferred payments, the system also records the contractual repayment window. This feature provides direct, time-stamped insight into short-term financing arrangements and payment timing at a level of granularity rarely available.

Documented Regularities. We exploit the universe of interbank transfers among firms to test for systematic payment–chain behavior. Absent such chains, there should be no systematic relationship between injections of liquidity and the timing of payments a firm. Thus, to empirically detect chained payments, we start by examining whether firm j is more likely to issue a payment within 24 hours of receiving one.[5] Specifically, we estimate:

$$sent_{jt} = received_{j,t-1} + \lambda_t + \lambda_{j\times d} + \varepsilon_{jt}, \tag{1}$$

where $sent_{jt} = 1$, if firm j made a payment on business day t, and $received_{j,t-1} = 1$ if firm j received a payment in the preceding 24 hours.[6] To prevent this relationship from being purely mechanical, wwe include both time fixed effects and firm × day-of-week fixed effects.[7]

[5] We begin with this narrow window to strengthen the inference of a direct link between payments received and payments sent; later, we extend our analysis to wider intervals.

[6] We restrict attention to business days; payments received on weekends or holidays are assigned to the next business day.

[7] Time fixed effects capture calendar-driven patterns—e.g., firms that always pay or receive on month-end or the 15th. Firm × day-of-week fixed effects absorb idiosyncratic weekly rhythms—e.g., some firms systematically pay on Thursdays, others on Fridays—thereby isolating only the abnormal deviations linked to incoming payments. The latter fixed effect would also discipline variation, for instance, from a

Table 1 reports the regression results. Column (1) shows that, on average, firms are significantly more likely to send a payment after receiving one within the prior 24 hours. Columns (2)–(4) reveal heterogeneity: the payment–chain effect is strongest among small firms (5 workers or less) and attenuated for firms with higher asset-to-sales ratios or greater loan access relative to sales.

Table 1: Triggered Payments Across Firms

Dependent variable: Prob. firm j makes a payment at time t

	(1)	(2)	(3)	(4)	(5)	(6)
$received_{j,t-1}$	0.032	0.011	0.042	0.044		0.032
	(.0005)***	(.0009)***	(.0008)***	(.0008)***		(.0005)***
$Small_j \times received_{j,t-1}$		0.022				
		(0.001)***				
$\frac{Assets_j}{Sales_j} \times received_{j,t-1}$			-0.019			
			(0.001)***			
$\frac{Liabilities_j}{Sales_j} \times received_{j,t-1}$				-0.022		
				(0.001)***		
$received_{j,t-5}$					0.011	0.011
					(.0004)***	(.0005)***
Observations	10,628,428	6,936,593	7,793,996	7,793,996	10,468,698	10,468,698
Adjusted R^2	0.494	0.483	0.483	0.483	0.494	0.494
$t, j \times$ d FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The independent variable in column (1) is an indicator that equals one if firm j received a payment from another firm within the last 24 hours. This variable is interacted with firm j's size, assets-to-sales and liabilities-to-sales in columns (2), (3), and (4) respectively. Column (5)'s independent variable is an indicator that equals one if firm j received a payment within 5 days of time t. Column (6) includes both the latter variable and a corresponding indicator for payments within 24 hours of t. Robust standard errors, adjusted for clustering by firm, are in parentheses. We include time and firm×weekday fixed-effects. Data is daily and spans 2024.

Columns (1)–(4) of Table 1 focus on a one-day window to approximate chained-payment responses. To trace the temporal profile of this effect for all firms, we implement a local projection exercise (Jordà, 2005; Jordà and Taylor, 2025). Figure 1 illustrates that, while many firms issue triggered payments within one business day, there is also a second peak—nearly as large as the initial response—around five business days (one week) after receipt of funds.

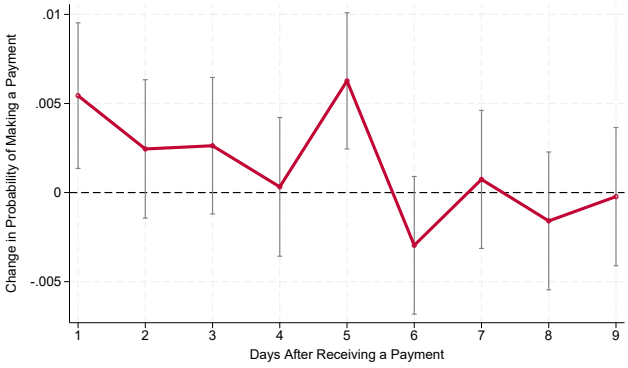
This two-wave structure reflects both the speed of Costa Rica's modern payment

firm that tends to make payments every day, and instead would force the variation to come from out-of-the-ordinary responses correlated with receiving a payment.

infrastructure and residual internal frictions. In 2024, when SINPE processes most transfers electronically, same-day responses are technologically feasible. By contrast, in 2005 most payments settled by check—requiring several days to clear—so that the fastest possible response was around four business days (see Figure 11, using SINPE data back to 2005).[8]

Even with a rapid payments network, frictions within firms—such as internal approvals, inspections of the products to be paid, or compliance checks—can induce delays, explaining the pronounced second wave at business day five. Columns (5) and (6) of Table 1 extend the specification in equation (1) by including an indicator for receipt of any payment in the preceding five business days; consistent with the local projection, this variable significantly predicts payment activity. Moreover, comparing coefficients in columns (1), (5), and (6) reveals remarkable stability, suggesting that payments unfold in two distinct and orthogonal waves—an immediate response and a delayed one around business day five.[9]

Figure 1: Triggered Payments: Local Projection



Notes: The figure shows a local projection (Jordà, 2005; Jordà and Taylor, 2025), where the dependent and independent variables have 13 lags, as determined by a VAR-based optimal lag selection criteria. Data is daily and spans 2024.

Placebo Exercises These relations are not mechanical. In fact, we can conduct placebo tests, where we change the timing so that the independent variable represents payments received h days *after* making a payment, with $h \in \{1, 2, 3\}$. As shown in Table

[8] Although firm-level transfer records are unavailable for the 1990s, one would expect even longer lags then, as checks could take up to 15 days to clear (Cerdas and Melegatti, 2014).

[9] We also examine the intensive margin in Table 3, which yields results consistent with the extensive-margin analysis: firms are more likely to make payments once they receive transfers.

4, results are consistently negative or insignificant, speaking against spurious temporal correlations.

Constructing Payments-Chains Leveraging the comprehensive ACH and wire transfers data, we aim to identify firms where the receipt of a payment triggers the issuance of subsequent payments. Our goal is to classify firms into three distinct groups: (i) those that tend to make chained payments without delay; (ii) those that tend to make chained payments with delay; and (iii) those that tend to make spot payments.

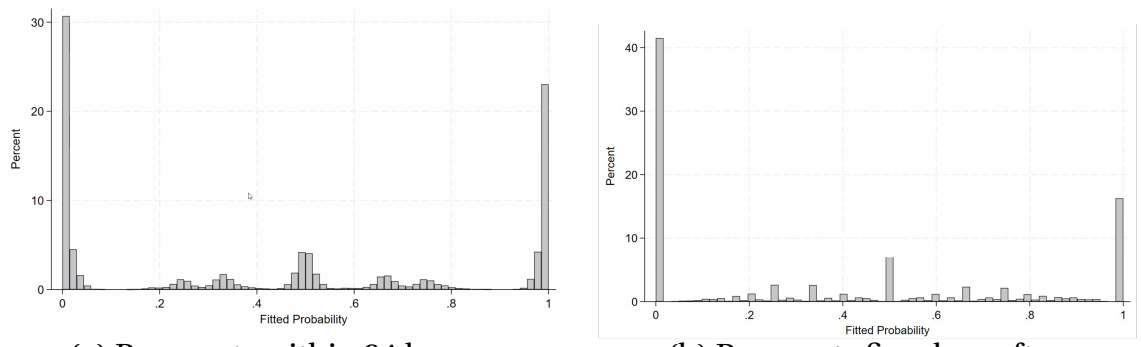
We proceed in two stages. First, to isolate firms chained without delay (type (i)), we compute for each firm the monthly frequency with which receiving any payment leads to issuing a payment *within one business day*. As shown in Panel (a) of Figure 2, this distribution has large masses of firms concentrated in a value of one (firms with perfect same-day response rates) and zero (firms with no same-day responses). Given this bimodal distribution, we then use k-means clustering to split the sample into two groups: firms type (i) that make chained payments without delay and other firms.

Second, among those firms that were *not* classified as type (i), and informed by the local projection exercise in Figure 1, we consider the distribution of monthly occurrences by firm in which receiving a payment led to making a payment five days later. This distribution, shown in Panel (b) of Figure 2 is, once again, largely bimodal, and allows us to split this subsample in two groups: firms concentrated in one, which correspond with type (ii)—chained with delay—and firms concentrated in zero, which correspond with type (iii)—spot, for whom payments sent are uncorrelated with payments received. This two-stage clustering yields a firm-month classification into three types; although firm types are highly persistent, our approach allows for type-switching from month to month.[10]

Stylized Facts on Payments-Chains Once firms are classified, we construct payments-chain networks. Several remarks are in order. First, on a given month, 43% tend to make spot payments, while the remaining 57% of firms engage mainly in chained payments.

[10]As an alternative, one can estimate each firm's fitted probability of triggered payments from the specification in column (6) of Table 1 and then apply k-means clustering with $k = 3$. The resulting three-group classification highly correlates with our baseline method at 0.94.

Figure 2: Histogram of Fitted Probabilities of Triggered Payments by Firm



(a) Payments within 24 hours (b) Payments five days after

Notes: The figures show the distribution of monthly average triggered payments by firm based on an extensive margin estimation. Firms with a value of zero have a payment behavior that is orthogonal to receiving money transfers, while firms with a value of one have a perfect correlation between receiving a transfer and making a payment either the day after (panel a) or five days later (panel b). Data is monthly and spans 2024.

Among the chained firms, 79% make payments without delay and 21% make payments with delay.[11] Second, interactions across firm types are common. As shown in panel (a) of Figure 3, 32% of all links in the payments network are "mixed," involving both a spot and a chained firm.[12] Third, if we focus on transactions between same-type firms, network structures differ significantly. On the one hand, Panel (b) of Figure 3 shows the network for transactions between firms classified as spot; 1% of links.[13] The network structure lacks a tight core, with a more radial or diffuse structure. The dense and uniform pattern suggests a decentralized transaction ecosystem with many firms directly transacting with others. Thus, the network appears relatively resilient, as disruptions in one firm's payments do not propagate as easily. On the other hand, panel (c) consists of transactions between firms classified as chained, blue (green) dots representing firms chained without (with) delay; 67% of all links. These links are particularly relevant as

[11] The k-means clustering partitions firms at a fitted probability threshold of 0.41 when considering chained without delay vs. other firms, and of 0.44 when considering chained with delay vs. spot firms.

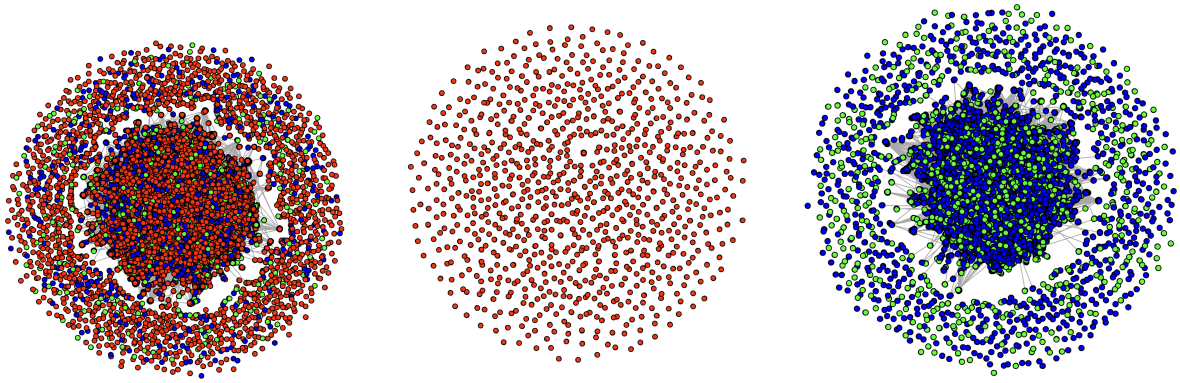
[12] This classification is based on December 2024. As shown below, the shares of links by types are stable over time. We also show a breakdown of mixed transactions into spot (S)→chained (X) and X→S.

Month	S→S	X→X	X→S	S→X	Total	Month	S→S	X→X	X→S	S→X	Total
1	1%	66%	2%	32%	187,026	7	1%	67%	1%	31%	194,151
2	1%	66%	2%	32%	190,099	8	1%	67%	2%	31%	183,187
3	1%	66%	2%	32%	178,270	9	1%	66%	2%	31%	184,636
4	1%	67%	2%	31%	195,494	10	1%	66%	2%	32%	191,507
5	1%	66%	2%	32%	197,412	11	1%	66%	2%	32%	186,960
6	1%	66%	2%	32%	178,841	12	1%	67%	1%	31%	190,737

[13] These red dots look less numerous than in Panel (a), as spot firms trade mostly with chained firms.

consecutive chained links ultimately form the payment chains. The network structure shown is a dense core with isolated nodes. The core cluster represents a critical mass of firms in tight transactional relationships and a circular dependency loop, while the outer nodes are less connected to same-type firms. Firms in this network's core are interdependent, which would relate to high systemic risk: if one core firm delays or defaults, it could ripple through the network.

Figure 3: Network Structure by Firm Type



(a) All (32% mixed links) (b) Spot-to-spot (1% of links) (c) Chained-to-chained (67% of links)

Notes: The figures show the network structure by link type. Panel (a) shows all links between firms. Panel (b) focuses on links between firms classified as spot. Panel (c) shows links between firms classified as chained. Data is collapsed for December 2024 to construct the network. Breakdowns for other months are similar and reported in Footnote 12.

With firm classifications in hand, we construct payment chains as network paths that begin and end at firms making spot payments, with chained firms potentially occupying the intermediate nodes. For each pair of spot firms connected through chained firms, we define the chain's length as the minimum number of links (firms) separating them.[14] In total, we identify approximately 2,000 distinct chains in the economy. Figure 12 displays the distribution of chain sizes for December 2024; results are nearly identical in other months, exhibiting only minimal month-to-month variation.

[14]To avoid double counting of overlapping fragments and to maintain computational feasibility at a monthly frequency, we measure chain length by the shortest path between the two spot firms in the payments network.

We further exploit the VAT registry of electronic invoices—linked at the firm level to SINPE transfers—to observe prices and timing of payment—immediate or deferred, and if deferred, for how long—for firms classified as either chained or spot. Because sellers must issue an electronic invoice upon receipt of any partial payment, deferred transactions in the VAT system naturally capture chained-payment behavior. Indeed, firms classified as chained are 83% more likely to engage in deferred payments than spot-payment firms, with an average deferral of 34 days. Leveraging this rare level of detail, we find that, for a given product, invoice prices are 12% higher when payment is deferred, and this premium is even larger when the purchasing firm is one that regularly makes chained payments. Appendix D.5 presents details of these results.

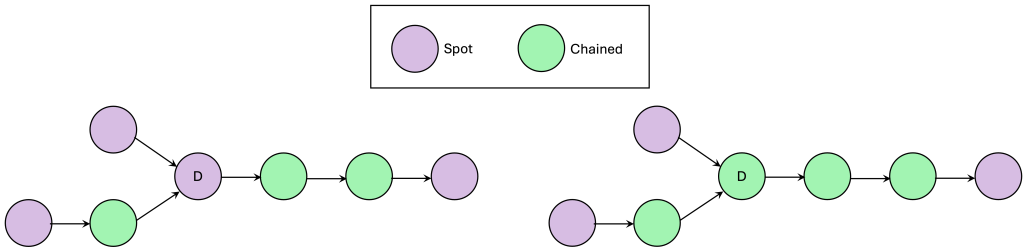
A Payments-Chains Crisis A payments-chains crisis may arise when shocks to short-term liquidity propagate through chained transactions and impair downstream payments. We exploit a natural experiment to assess the empirical relevance of such crises.

In August 2024, Costa Rica's *Financiera Desyfin*—a financial institution catering mainly to enterprises and business clients—was abruptly and unexpectedly intervened by the financial regulator, who immediately froze all customer accounts due to insolvency concerns. The regulator then declared the institution unviable and placed it into a resolution process under the Deposit Guarantee Fund law, leaving those funds inaccessible for several months.

Using SINPE data, we can identify the directly shocked firms and trace how disruptions ripple through their surrounding payment networks. To build intuition, we will provide an example. Consider the pre-freeze network in panel (a) of Figure 4, in which arrows represent flows of payments. Firm D is classified as a spot payer, issuing transfers to chained firms downstream (to its right in the figure) with no systematic link between payments it received and sent. After the Desyfin freeze, however, firm D's sent payments started happening only after receiving a payment, so that it "switched" to being classified as chained, as depicted in panel (b).

With this example in mind, let $Desyfin_i = 1$ if firm i maintained an account at Financiera Desyfin during 2024, and let $Shock_t = 1$ for $t \geq$ August 2024, when the freeze occurred. We also define $Switch2chained_{it} = 1$ if firm i changes from being classified as

Figure 4: Example: Firm D with a Frozen Account "Switches" from Spot to Chained



(a) Chain prior to accounts' freeze (b) Chain after accounts' freeze

Notes: The figures show an example of how the accounts' sudden freeze might have impacted a chain in which one, formerly spot, firm "switched" to being classified as chained. Arrows represent payment flows, so circles positioned farther to the right correspond to firms that are further downstream.

spot to being classified as chained in month t.[15]

We first explore if firms which had an account at Desyfin become more likely to "switch" to being classified as chained after the freeze. Column (1) of Table 2 reports that this occurs; the probability of switching from spot to chained increased in 0.008 pp—equivalent to a 17% rise relative to the low 0.05 baseline. These estimates include province×industry and time fixed effects to absorb regional or sectoral shocks.

Next, we examine how a "switch" to chained status might impact a directly shocked firm (D, in our example above). To do so, we now consider the indicator $Switch2chained_{it}$ as an independent variable, and study if firm i that had an account at Desyfin *and* switched to being classified as chained made fewer purchases after the shock, which would align with a hindered ability to make payments. As shown in column (2) of Table 2, this is indeed the case, with a 42% fall in purchases made. Remarkably, the same column shows that $Switch2chained_{it}$, on its own, has the opposite effect on purchases; and that it is the interaction with the liquidity freeze that produces the negative effect. The latter is particularly relevant, as one would expect $Switch2chained_{it}$ to have a positive coefficient, as firms might start making chained payments as a way to expand their production. The term $Desyfin_i \times Shock_t$ also has, on its own, a positive coefficient, aligned with positive pre-shock selection into Desyfin, which catered to corporate clients. In line with fewer purchases, column (3) also shows sales fell in a similar proportion.

[15]For this construction, recall that the classification of firms by type is month-specific and allows for type-switching across time. While the classification is flexible, firm types tend to be highly persistent.

We finally ask how chained suppliers fare when selling to a directly shocked firm. Define $Selling2Desyfin_{jt}$ as an indicator equal to one if chained firm j was *selling to* a directly shocked "Desyfin" firm i. Panel (B) of Table 2 shows a dramatic 60% reduction in both purchases and sales after the freeze, *but only when they were selling to a firm that switched to chained payments and held a Desyfin account* (columns 4 and 5). These results underscore how payment chains can undermine economic resilience. Just as a project developer's lack of funding can trigger cascading delays for the general contractor, subcontractors, and so on, a shock—such as a bank account freeze—that disrupts an initial spot payment can cause delays and reduced productivity for all firms down the payments-chain.

Table 2: A Payments-Chains Crisis: Frozen Bank Accounts

A. Directly Shocked Firm i			
	$Switch2chained_{it}$	ln $Purchases_{it}$	ln $Sales_{it}$
	(1)	(2)	(3)
$Switch2chained_{it} \times Desyfin_i \times Shock_t$		-0.421	-0.431
		(0.065)***	(0.058)***
$Switch2chained_{it}$		0.311	0.205
		(0.008)***	(0.010)***
$Desyfin_i \times Shock_t$	0.008	0.052	0.042
	(0.003)***	(0.004)***	(0.006)***
%Δ w.r.t. mean [mean dep. var]	17% [0.05]		
Observations	1,202,560	1,193,436	916,594
Province×industry, t FE	Yes	Yes	Yes
B. Firm j Selling to a Directly Shocked Firm i			
		ln $Purchases_{jt}$	ln $Sales_{jt}$
		(4)	(5)
$Selling2Desyfin_{jt} \times Switch2chained_{it} \times Desyfin_i \times Shock_t$		-0.600	-0.569
		(0.029)***	(0.058)***
$Selling2Desyfin_{jt} \times Switch2chained_{it}$		0.311	1.175
		(0.041)***	(0.045)***
$Desyfin_i \times Shock_t$		0.086	0.082
		(0.002)***	(0.002)***
Observations		967,341	1,336,981
Province×industry and t FE		Yes	Yes

Notes: The dependent variable in column (1) is an indicator equal to 1 if firm i switched from being classified as a spot firm to a chained firm in period t. Columns (2) and (3) investigate the impact of the freeze on firms who held accounts at Desyfin, before and after the freeze, differentiating between impacts if the firm switched to being classified as chained or not. Columns (4) and (5) instead focus on firms that were selling to a firm directly connected to Desyfin, before and after the shock. Robust standard errors, clustered by firm, are in parentheses. We include province×industry (ISIC-4) and time fixed-effects. Data is monthly and spans 2024.

3. Payments-Chains and Productivity

Motivated by the empirical evidence of a payments-chain network, this section presents an approach to map a set of income and expenditures in the economy into a payments-chain network. When transaction delays entail an economic loss, the payment-chain network translates into TFP losses. We then then embed this module into a dynamic business-cycle model.

Bilateral Relations. To fix ideas, assume that production is organized through bilateral agreements in which a customer orders a product from a supplier owning one production input. The agreement is exclusive: only the agent placing the order values the output once produced. In turn, production units are exclusively dedicated to producing for a specific client.[16] As examples of such relations, we can think of a home renovation project, a medical service, manufacturing an engineered product with a specific blueprint, or commissioning a piece of art.

There are two types of orders: *spot* and *chained* orders. Spot orders are paid immediately. Chained orders are contingent on previous payments; they are paid after the client receives a payment from another transaction. There are N production units. In turn, there are N^s spot orders and N^x chained orders. There are equal numbers of production units and orders, $N = N^s + N^x$. In the following section, we work with a limit for $N \to \infty$ and recast this condition as a market-clearing condition. Each production unit is assigned an identifier, $i \in \mathcal{N} = \{1, 2, \ldots, N\}$. Likewise, each order is assigned a unique identifier, $i \in \mathcal{N}$. We partition $\mathcal{N}$ into two sets, $\mathcal{N}^s$ and $\mathcal{N}^x$, to denote the set of identifiers of spot and chained orders, respectively.[17] Also, we work under the assumption that payments are identical across orders, a condition we explain in the following section. We also provide an interpretation of spot and chained once we introduce financial decisions in that section.

We define two relations that together define the payments-chain network. First, $\mathcal{P} : \mathcal{N} \to \mathcal{N}$, is a one-to-one assignment from an order to a production unit. By as-

[16] The agents' identity in these agreements is not explicit at this stage. Identities are explicit, and these matter for the economy's dynamics in the following section.

[17] $\mathcal{N}^s$ and $\mathcal{N}^x$ are a partition of $\mathcal{N}$: $\mathcal{N}^s \cap \mathcal{N}^x = \emptyset$ and $\mathcal{N}^s \cup \mathcal{N}^x = \mathcal{N}$. The number of elements is $N^s = \mathcal{N}^s$, $N^x = \mathcal{N}^x$.

sumption $i \neq \mathcal{P}(i)$. The interpretation is that $\mathcal{P}(i)$ is the unit that produces order i. The assignment $\mathcal{P}$ is entirely random: any unit can be assigned to any order with equal probability. Second, a chained income-expenditure relation associates a chained order with a production unit. This relation is the identity function defined on $\mathcal{N}^x$. That is, order i uses the payments to production unit i if $i \in \mathcal{N}^x$, as a source of funds to execute order i. The idea is that although order i is not externally funded, the customer who places order i owns production unit i. Hence, order i can be funded after order $j = \mathcal{P}^-(i)$ executes a payment. The assumption that the income-expenditure relation is the identity is innocuous.[18]

To anticipate how the two relations induce a payments-chain network, consider production unit j assigned to order i, $\mathcal{P}(i) = j$. The client placing order i must pay for j's production. This creates a payment link from i to j. In turn, if order j is spot, $j \in \mathcal{N}^s$, the funds used to pay for order i are unused in further payments. However, if order j is chained, the funds are used again to pay unit $k = \mathcal{P}(j)$. In other words, when $j \in \mathcal{N}^x$, there is a flow of payments from i to j, and from j to k. If order k is also chained, the same funds are used to pay unit $\mathcal{P}(k)$, and so on. The payments chain continues until a final order in a chain is placed on some production unit i not associated with a chained order. Since every order is paired with a production unit, the economy features an entire network of transactions forming a collection of payment chains.

The payments network is inconsequential if there are no delays in transferring funds. Indeed, if every payment was executed immediately, the transfer of funds in the entire network would be instantaneous. Moreover, even if transfers are not immediate, the network would be inconsequential if no resources were lost as transactions are delayed. Of course, transaction delays can occur for many reasons: technology restrictions, inattention, or financial frictions. The economic consequences can be various: inventory depreciation, time-discounting, or utilization.

Here, we provide an example of how the network is consequential. We introduce two financial frictions that delay payments and, consequently, production. The first

[18] Formally, we can define the chained income-expenditure relation $\mathcal{X}$ as the identity function on $\mathcal{N}^x$, that is $\mathcal{X} : \mathcal{N}^x \to \mathcal{N}^x$ such that $\mathcal{X}(i) = i$. The idea is that $i \in \mathcal{N}^x$ obtains funds from production unit, $\mathcal{X}(i) = i$. Indeed, the identity function $\mathcal{X}(i)$ can be replaced by any injective function $\mathcal{X} : \mathcal{N}^x \to \mathcal{N}$ so that production units and associated chained orders do not have the same identifiers. Changing identifiers does not affect the results.

friction is that production can only start once the buyer shows proof of funds. If an order is spot, by definition, proof of funds is immediate since the order is funded externally. By contrast, chained orders can provide evidence of funds only after the buyer receives a payment from another transaction and the money enters their account. Production requires proof of funds to deter fraudulent behavior; otherwise, buyers could promise payments they know will never occur. Since production is customized, this would lead to a disadvantageous ex-post renegotiation.

We need a reason to delay transactions: here, we invoke limited commitment. The idea is that after the customer shows proof of funds, the funds are only released after the fraction $1-\delta$ of the output is inspected. Without inspection, The producer could produce a good for himself, taking away the funds. Since there is no way to verify a customization, inspection is necessary.

The delay in transactions provokes production losses. Namely, the payments-chain network determines production: production occurs within a unit time interval. Each order starts at some endogenous time $\tau \in [0,1]$ when the payment is executed. Once production starts, it cannot stop. Production is linear in time: if production starts at τ, production is $1-\tau$. The delay in the time to execute payments then translates into idle resources which map into TFP losses.

Placing these frictions together, proof of funds and limited commitment, we obtain a production structure where the greater the number of chained orders, the more payment delays and the longer the production delays. To produce predictions about aggregate TFP, we first formalize the definition of the payments chain network and derive statistical properties that help us map it into production losses.

Payments-Chain Network. Once we enumerate transactions, we can construct a network of transactions. A sequence of payments in the network that starts from a single original source of funds defines a **payments chain**. Naturally, every chain must start with a spot order, followed by a sequence of chained orders in which those initial funds are transferred from one transaction to another. The number of chained orders defines the **chain length**. A **payments-chain network** is the collection of all payment chains, encompassing the universe of transactions during the production time interval. We employ a formal definition.

Definition 1. *A **payments-chain network** $\mathcal{K}$ is an acyclical directed network with degree 1 with nodes $\mathcal{N} = \{1, 2, ..., N\}$ and links $\mathcal{V} = \{(i, j) | \mathcal{P}(i) = j, j \in \mathcal{N}^x\}$, $\mathcal{K} = (\mathcal{N}, \mathcal{V})$. A **payments-chain** of length n is a finite sequence of nodes $\{i_k\}_{k=0}^{n}$ such that the sequence starts with some $i_0 \in \mathcal{N}^s$ and $\forall k \in \{1, \ldots, n\}$, $\mathcal{P}(i_{k-1}) = i_k \in \mathcal{N}^x$. By convention, if $i \in \mathcal{N}^s$ and $j = \mathcal{P}(i) \notin \mathcal{N}^x$, i and j define a chain of length zero.*

Nodes corresponding to the payment chain network represent both orders and production units. The directed links represent the direction of the flows of funds. Because production is bilateral, one or no links stem from each node. A link from i to j indicates that i orders from production unit j and that j is a chained order, $\mathcal{P}(i) = j \in \mathcal{N}^x$. The source of funds of order j is the funds paid in order i. In turn, if a node does not receive a link, it represents a spot order. Furthermore, if no links are directed toward order $j = \mathcal{P}(i)$, then order j is also spot—in which case i and j form a zero-length chain. Notice that in this construction, each order has a source of funds. Any (longest) path of links defines a payment chain. The collection of payment chains is the payment-chain network.

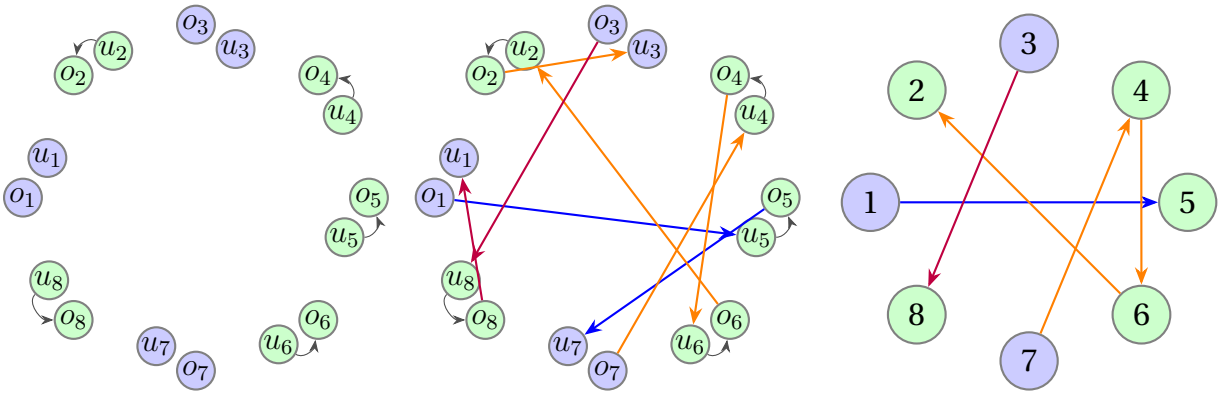
Examples. Let us provide an example. Set $\mathcal{N} = \{1, 2, \ldots, 8\}$ and let the subset of spot orders be $\mathcal{N}^s = \{1, 3, 7\}$. Also, define the production relation as follows: let $\{i_n\}_{n \in \mathcal{N}} = \{1, 5, 7, 4, 6, 2, 3, 8\}$ such that $i_{n+1} = \mathcal{P}(i_n)$ and $i_1 = \mathcal{P}(i_N)$. Thus, the links in this payments-chain network are $\mathcal{V} = \{(1, 5), (7, 4), (4, 6), (6, 2), (3, 8)\}$.

Several graphs are associated with this example. The left panel of Figure 5 depicts the chained income-expenditure relation. In that panel, we split each node into counterparts: the production units $\{u_n\}$ and production orders $\{o_n\}$ for nodes, $n \in \{1, 2, ..., 8\}$. The links represent the flow of funds from production units to their corresponding chained orders, defined by the chained income-expenditure relation.

The middle panel adds the links to the flow of payments for production, corresponding to $\mathcal{P}$. That is, the links in that panel add the payments production units. Adding the links of the chained income-expenditure and production relations allows us to trace funds. Notice that the links from orders to production units with the same color share a common source of funds.

The right panel depicts the payment chain network. In this example, there are three

chains. The first chain is of length one—from 1 to 5. The second chain starts at 7, and links nodes 7, 4, 6, and 2. Since 4, 6, and 2 are chained orders, the length of the second chain is three. The third chain links node 3 with node 8 and is also of length one.



(a) Income relation (b) Network in detail (c) Payments-Chain Network

Figure 5: Components of the Payments-Chain Network

Chain-Length Distribution. We now derive the distribution of payment-chain length as the number of transactions increases, $N \to \infty$. As the number of transactions increases, we keep the ratio $\mu \equiv N^x/N$ constant. This fraction is a parameter of the probability distribution $G(n;\mu)$ chain length in the network.

Proposition 1. Let $n \in \{0,1,2,\ldots\}$ be the length of a payment chain in the payments-chain network. Then, n is a random variable with probability mass function (p.m.f.) $G(n;\mu)$ where $G(n;\mu)$ is the *geometric p.m.f.* with parameter μ, that is, $G(n;\mu) = (1-\mu)\,\mu^n$.

Proof. Recall that the production relation is random. As $N \to \infty$, a node has a link directed toward it if it corresponds to a chained order. Thus, a node is linked with probability μ. A node does not receive a link if it is spot, with probability $1-\mu$. Furthermore, recall that each chained order is funded and thus belongs to specific payment chain which, in turn, starts with a specific spot order. Thus, there is a one-to-one relation between each payment chain and a spot order. Hence, there are N^s chains in total,

which we can also index by $i \in \mathcal{N}^s$. If chain i forms a chain of length zero, it must be that $\mathcal{P}(i) \in \mathcal{N}^s$. This, happens with probability $1-\mu$—because each production link indeed occurs with equal probability. Likewise, chain i is of length 1 if $\mathcal{P}(i) \in \mathcal{N}^x$ but $\mathcal{P}(\mathcal{P}(i)) \notin \mathcal{N}^x$. This happens with probability $\mu \times (1-\mu)$. In a chain of length two, there are two consecutive chained orders followed by a specific spot order. This occurs with probability $\mu^2(1-\mu)$. Proceeding by induction, we arrive at the geometric distribution. □

The length distribution is convenient for obtaining a measure of economic losses. For that, we turn to our specific micro-foundation.

TFP. We use this distribution to obtain aggregate production: we first derive the production in a chain of arbitrary length. In a chain of length zero, there is always a spot order whose production begins immediately. In a chain of length one, the production of the first spot order begins immediately, but there is a delay in the second order. For the second order, the funds are received after $1-\delta$ of the production of the first order is finished. This happens at time $\tau = 1-\delta$. Hence, this leaves only δ time to produce the second order in the chain. For the second order, production is δ of which the $1-\delta$ fraction must be inspected. If there is a third order in the chain, the transfer of funds occurs $(1-\delta)\delta$ time after the first transfer at $1-\delta$. Adding these consecutive delays, the production of the third order can only start by $(1-\delta)+\delta(1-\delta) = 1-\delta^2$. This leaves δ^2 time to produce that third order.

We can deduce a pattern by forward induction.[19] In a chain of length n, the production vector of $n+1$ consecutive orders in a chain is $\{1, \delta, \delta^2, \ldots, \delta^n\}$. Using the distribution of chain lengths to compute expectations, we obtain aggregate production.

Proposition 2. (Output per worker): *Given μ and δ, the average output corresponding to chained orders converges, as $N \to \infty$, to:*

$$\mathcal{A}(\mu;\delta) = \frac{1-\mu}{\mu}\frac{\delta}{1-\delta}\ln\left(\frac{1-\delta\mu}{1-\mu}\right) < 1. \tag{2}$$

[19] If the k-th node initiates production at time $1-\delta^{k-1}$, the delay from its inspection is $(1-\delta)\delta^{k-1}$, which added to previous delays leads to a transfer of funds only by time $1-\delta^{k-1}+(1-\delta)\delta^{k-1} =: 1-\delta^k$. This, leaves δ^k time for production to the subsequent unit. Since we computed the delay for $k = 1, 2, \ldots$. the productions in a chain of length n are $\{1, \delta, \delta^2, \ldots, \delta^n\}$.

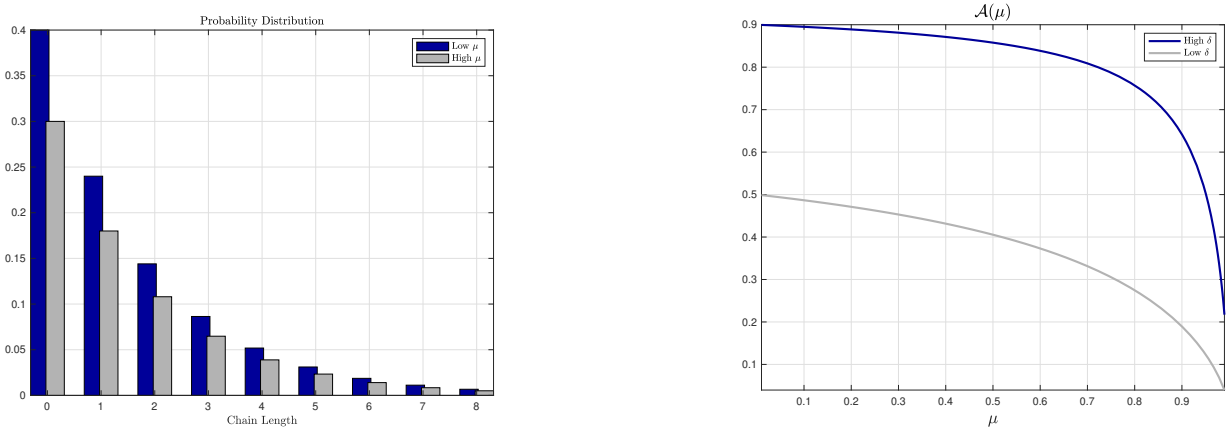
$\mathcal{A}$ is strictly concave, decreasing, and satisfies $\lim_{\mu\to 0}\mathcal{A}(\mu;\delta) = \delta$ *and* $\lim_{\mu\to 1}\mathcal{A}(\mu;\delta) = 0$ *and* $\lim_{\delta\to 0}\mathcal{A}(\mu;\delta) = 0$ *and* $\lim_{\delta\to 1}\mathcal{A}(\mu;\delta) = 1$. *Average output is* $\mathcal{Y}(\mu) = (1-\mu) + \mu\mathcal{A}(\mu) \leq 1$.

This theorem is key as it yields a formula output: To obtain output, we first compute the average production among chained orders, $\mathcal{A}(\mu;\delta)$.[20] $\mathcal{A}$ is strictly concave and decreasing in μ and has well-behaved limits.[21] Total output $\mathcal{Y}(\mu;\delta)$ is constructed by noting there are $(1-\mu)$ spot orders for which production is 1 and μ for chained orders whose average production is $\mathcal{A}$. Since labor is in fixed supply, $\mathcal{Y}$ is both output and TFP. However, since efficiency losses are fully driven by utilization among chained orders, we refer to $\mathcal{A}$ as TFP directly. To illustrate the formula for $\mathcal{A}(\mu;\delta)$ and its construction, the left panel of Figure 6 plots the distributions of chain lengths—corresponding to two values of μ. For the higher μ, the distribution shifts mass to chains of greater length. The right panel graphs $\mathcal{A}$ as a function of μ, for two values of δ. Productivity falls as the fraction μ increases. In turn, the lower δ, the greater the delays and the lower TFP.

The main insight is that TFP losses result from payment delays. It is worth distinguishing the sources of these losses from other environments. Unlike search models, there are no congestion externalities—the production assignment s one-to-one. Unlike sticky-price models, the fact that payments are the same across orders does not have implications for production (input uses do not depend on prices). The source of TFP losses are the production delays caused by payment delays. The economy is at full capacity when if all transactions are spot. Importantly, TFP losses are magnified by the random assignment of orders. Holding fixed the number of spot orders, if payment chains could be reorganized to be all of the same length, the economy would still

[20] To obtain output, we first compute the average production of chained orders in an n-length chain, $\left(n^{-1}\sum_{i=1}^{n}\delta^{i}\right)$. With the average production of each chained orders, we can compute the expected value of production in chains with at least one chained order. Integrating across all possible lengths, we obtain $\mathcal{A}(\mu;\delta)$, the average production for chained orders. For that, we use the discrete probability $(1-\mu)\mu^{n}/\mu$. This is the distribution of chain length conditioned on $n>0$, obtained from $G(n;\mu)$. Integrating across all possible lengths, we obtain $\mathcal{A}(\mu;\delta)$, the average production for chained orders.

[21] With respect to the limits of $\mathcal{A}$, as $\mu\to 1$ the chains become larger but their average production decreases to zero since the additional production of chained order decreases exponentially. On the other hand, when $\mu\to 0$ the chain length tends to 1 and thus per-worker productivity tends to δsince average delay times converge to one period. In turn, when $\delta\to 0$ transfers take the full production period leaving not time for production. Conversely if $\delta\to 1$, chained orders feature no delays. Interestingly, $\mathcal{A}(\mu;\delta)$ resembles an entropy function. We are unaware of a connections between the geometric distribution, a discounted sequence, and the entropy function.



(a) Chain length distribution (b) TFP

Figure 6: Chain-Length Distributions and TFP

Notes: Left panel graphs $G(n)$ for $\mu = \{0.6, 0.7\}$. Right panel $\mathcal{A}(\mu)$ for $\delta = \{0.5, 0.9\}$.

feature TFP losses, but less so than under random assignment (Jensen's inequality).

Discussion: Reduction of Economic Complexity. The payments-chain network here is simple. In practice, economies involve more complexity than exhibited here: Transactions differ by size and are coupled in more intricate production networks. Moreover, production ends by the end of the period and orders are not withdrawn.[22] Studying these richer dimensions would make the problem more realistic, but complicate the analysis.[23] Despite the crude assumptions, the payments-chain network illustrates the effects of delays. A virtue is that we can obtain a mapping from a financial quantity, μ, to measurable TFP. Because μ depends on agent decisions in the following section, we can study policy exercises immune to the Lucas critique.

[22] To allow withdrawals, we would have to find an endogenous maximal chain-length provided that we obtain a relative price for funds and production, as we do in the following section. In that case, the production in a chain would drop to zero for orders above a given position. An analog, but more complicated, formula to TFP, (2), can be derived for the case.

[23] Statistical physics handles combinatorial problems that lead to this complexity. Potentially, such methods can be used in a similar environment to the one here. In particular, statistical physics offers tools to calculate statistical properties of the system.

3.1 Extensions

The formulas for the chain distribution and TFP can be generalized by relaxing several assumptions that bring the model closer to the data.

Correlation in order types. The payment-chain network introduced above we assumed no correlation among order types: a buyer faces the same chances of meeting a spot or chained seller. We now allow for correlated order types, matching the clustering we observe empirically (see Figure 5).

To introduce correlation, recall that the payments-chain network induces a binary sequence of order types—spot or chained—along each chain. Fix a chain of length n; the sequence of order types can be written as $S^N = \{s, x, s, x, s, ..., x\}$. To introduce correlation, it is convenient to model S^n as the realization of a Markov chain, not in time, but across positions in the payments network. That is, we interpret the realization of the sequence of payments as evolving by position along a production chain, with transitions governed by a Markov matrix.

Each realization of the overall network S^N, for $N \to \infty$, can then be understood as a sample path from this position-based Markov process. The construction of the network thus reduces to finding a stochastic matrix Q (a 2×2) matrix with non-negative entries and column sums equal to one) that governs the transition probabilities across payment types. Importantly, this matrix must preserve the unconditional frequencies of spot and chained orders. That is, it must satisfy:

$$\begin{bmatrix} 1-\mu \\ \\ \mu \end{bmatrix} = Q \begin{bmatrix} 1-\mu \\ \\ \mu \end{bmatrix},$$

where μ is the fraction of chained orders that preserves the populations of order types. A special case is the i.i.d. network introduced earlier, where links are independently

assigned as spot or chained (with probabilities $(1-\mu)$ and μ), with transition matrix:

$$Q^{\text{iid}} = \begin{bmatrix} 1-\mu & 1-\mu \\ \mu & \mu \end{bmatrix}.$$

To allow for *correlation* between adjacent positions in the chain (i.e., conditional dependence), we generalize this to a one-parameter family of matrices $Q(\rho)$:

$$Q(\rho) \equiv Q^{\text{iid}} + \begin{bmatrix} \frac{\mu}{1-\mu}\rho & -\rho \\ -\frac{\mu}{1-\mu}\rho & \rho \end{bmatrix},$$

where the correlation parameter $\rho \in [-\mu(1-\mu), \min\{\mu, 1-\mu\}]$, satisfies the stationarity condition. This form ensures that the overall fraction of chained orders remains, μ, while introducing correlation: a positive value of ρ implies that chained orders are more likely to be followed by chained orders, and similarly for spot orders. We will describe how TFP will change in this case in brief, and provide more details on this representation in Appendix B.1.

Default When default is possible, the payment-chain distribution is still distributed geometric (in terms of contracted payments). However, the realization of payments and production is truncated after some chain position η. This would be the case, for instance, if there was a freeze in bank accounts that did not allow payments to flow for certain firms. Thus, aggregate production among chained orders in a chain of length n is given by:

$$y^x(n) = \begin{cases} \delta\left(\frac{1}{1-\delta} - \frac{\delta^n}{1-\delta}\right), & \text{if } n \leq \eta \\ \\ \delta\left(\frac{1}{1-\delta} - \frac{\delta^\eta}{1-\delta}\right), & \text{if } n > \eta. \end{cases}$$

More details behind these results are presented in Appendix B.2.

Heterogeneity in Delays Empirically, we observed that an important share of firms made chained orders, but upon receiving funds, they transferred them immediately to their suppliers. To better align the theory with this finding, we can allow for some payments associated with chained orders to be transferred immediately when a payment is received. In particular, we assume that the receiver of funds experiences a delay with probability q. This probability is i.i.d., but clearly affects the production of chained orders only. In this case, given a chain of length n, whose size is still distributed geometrically, the probability of k delays is given by a multinomial:

$$M(k,n) = \left(\begin{array}{c} n \\ \\ k \end{array} \right) q^k (1-q)^{n-k}.$$

Thus, output is given by the joint distribution of chain length and the distribution of delays in the chain, which is now a multinomial. Details are provided in Appendix B.3.

Modified Formulas. We present the formulas for TFP for the variations of the model described above, with correlation, default, and heterogeneity. When we allow for correlation, the chain-length distribution is again distributed Geometric with parameter $p = \mu + \rho$ instead of μ. Moreover, formula for TFP is given is the same as in the baseline model, but with argument p instead of μ. That is, the TFP is $\mathcal{A}(p)$.

- If furthermore, chained-orders in position η or above default TFP is given by:

$$\mathcal{A}^d(p;\eta) = \frac{1-p}{p}\frac{\delta}{1-\delta}\left(\log\left(\frac{1-\delta p}{1-p}\right) - \left(\frac{\delta^\eta}{\delta p}\Phi(p,1,\eta) - \frac{1}{\delta p}\Phi(p\delta,1,\eta)\right)\right).$$

  where $\Phi(z,1,\eta)$ is the Lerch transcendent:

$$\Phi(z,1,\eta) = \int_0^\infty \frac{\exp(-\eta t)}{1 - z\exp(-t)} dt.$$

- If we allow for immediate payment transfers:

$$\mathcal{A}^i(\mu;q) = (1-q) + \frac{1-p}{p}\frac{\delta}{1-\delta}\left(\log\left(\frac{1-p(1-q+\delta q)}{1-p}\right)\right).$$

Notice that $\rho = 0$, we have no correlation and the formulas boil down to the baseline for $\eta \to \infty$ or $q = 1$, respectively.

Resampling. This extension aims at reproducing the empirical distribution of chain lengths in Figure 12. We assume that after observing their position in a chain, chained types can attempt to resample. Potentially there is a cost to do so, but for starters, let's assume there is no cost, and only assume that types above $n > \eta$, instead of defaulting, can resample. When they do so, they resample, they contact another supplier at random, which can indicate the chained supplier at the end of their chain. If they do so, these orders can be be placed at the end of the chain of the supplier they contact. Thus, all chains of $n > \eta$ now shrink to η. The process repeats itself indefinitely until there are no further resampling. In this further extension, we show that the final distribution is no longer Geometric, but governed by the evolution of Fokker-Planck equation that depends on several details of the chain formation process. The salient feature is right truncation and the movement of mode of the distribution of chain lengths towards numbers above one, as described in more detail in Appendix B.4.

Alternative use of time. In the baseline version, we assume that labor remains entirely idle as sellers are waiting for payment. As a final extension, we make the, perhaps more realistic assumption, that labor has an alternative use, namely, to produce a generic product that can be sold in a common market. The generic product is worth $\alpha < 1$ units of the customized product. Then, TFP losses should be scaled up by:

$$\mathcal{Z}(\mu) = \alpha + (1 - \alpha)\,\mathcal{A}^i(\mu; q)\,,$$

so that when output is entirely wasted $\alpha \to 0$ and we recover the original formula whereas when the alternative use of time is a perfect substitute $\alpha \to 1$ and there is no loss.

4. Payments-Chains in a Business-Cycle Model

We now incorporate the payments-chain network into a business cycle setting. We use the results of the previous section as inputs.

4.1 Environment

Timing. The horizon is infinite. Expenditure and savings decisions are programmed at integer dates, $t = \{0, 1, 2, ...\}$. Production happens within the time interval between integer dates, according to the payments-chain network of the previous section. It is no longer necessary to refer to the time interval of production; it is understood to happen within dates. There is perfect foresight. Labor units are the numeraire.

Demographics. The economy is populated by two households. One is the worker household that inelastically supplies labor and starts with negative financial wealth. The other is the saver household that has positive financial wealth but does not produce. Each household consumes goods by placing production orders in the payments-chain network. We use superscripts s and w to distinguish saver from worker variables. The advantage of this setup is that it allows for closed forms, akin to endowment economies with log-utility, while allowing for labor income.

Income, Expenditures, Transactions, and Prices. Production and consumption depend on the expenditure mix between spot and chained orders by both households. The worker household is endowed with N labor units. As in the previous section, a labor unit is assigned to a single order. We normalize payment per labor unit to $1/N$ so that total labor income is one.

We work with the limit as $N \to \infty$. As we increase the number of orders, we scale the production of each labor unit by $1/N$, to keep maximal feasible production equal to one. Implicitly, households place a large number of orders. Aggregating across all those orders, the risk in the position of a chain order is diversified. This assumption is akin to the classic 'big-family" assumption of Lucas and Stokey (1987). In the previous section, we assumed that all production units are paid the same, regardless of their

production. A motivation for this assumption is pairwise stability.[24]

At any t, both households choose amounts of spot and chained orders. To place spot orders, households must possess funds. Savers do not produce, so they are restricted to only making spot payments. For the rest of the paper, it is understood that the saver's period consumption C^s is purchased via spot orders. By contrast, the worker household has to choose between spot and chained orders. The total of goods purchased by workers is C^w

$$C^w = S^w + X^w, \tag{3}$$

where S^w and X^w are her goods purchased via spot and chained orders (henceforth, spot and chained goods respectively). As in the previous section, each chained order must be backed by a labor income unit.

Aggregating across households, there is a total of S spot goods and X chained goods produced. In turn, adding the expenditures of both households, expenditures in spot and chained orders are E^s and E^x, respectively. In equilibrium, the total income (of households) must equal total expenditures, so $E^x + E^s = 1$. The ratio of chained expenditures to total expenditures is $\mu = E^x$ and, thus, $E^s = 1 - \mu$.

In the previous section, we worked with integer amounts of production units and orders. We further imposed the countability condition, $N = N^s + N^x$. Here, expenditures are not restricted to be natural numbers when households decide on them. Moreover, we impose a labor market-clearing condition. Both properties are internally consistent. To see this, we can count the number of spot and chained orders with a floor and ceiling function, $N^s = \lfloor E^s_t \times N \rfloor$, $N^x = \lceil E^x_t \times N \rceil$, given expenditure choices. The countability condition is satisfied as long as $E^x + E^s = 1$; any inconsistencies—due to rounding errors—between the expenditure choices and the number of goods bought vanish as $N \to \infty$.

Recall that total output depends on μ. Since for each spot order, there is one unit of output, $E^s = S = 1 - \mu$. Thus, using Theorem 2 we have that $\mu \mathcal{A}(\mu) = X$. If we

[24] See Bloch and Jackson (2006) for a definition. Applied here, pairwise stability requires agents to accept the links in a network at the moment of placing orders—without knowledge of the location in the network. To accept an order, labor units must be paid the same. Otherwise, they would sever links and a new network would dissolve. In tun, to place an order, households must be willing to do so, anticipating the average amount of goods obtained by placing spot and chained orders, respectively. This is the case in this model.

substitute $\mu = E^x$ into this relation, we find that $E^x = \mathcal{A}^{-1}(\mu) X$. Because the number of orders tends to infinity, we can treat $q(\mu) \equiv \mathcal{A}^{-1}(\mu)$ as a price of chained goods per unit of chained expenditure. We use this auxiliary price, to define a worker expenditure bundle:

$$S^w + q(\mu) X^w = E^w. \tag{4}$$

where E^w are the worker expenditures.

Savers. The saver's period utility is $\log(\cdot)$. He maximizes discounted lifetime utility over the sequence $\{C_t^s\}$. Savers begin each t with real deposits, D_t, their only source of wealth. Deposits earn an equilibrium return R_t. Given D_t, savers choose future savings, D_{t+1}, and current expenditures.

Problem 1. (Saver Problem): *Given* D_0 *and* $\{R_{t+1}\}_{t\geq 0}$,

$$\max_{\{C_t^s\}_{t\geq 0}} \sum_{t\geq 0} \beta^t \log(C_t^s),$$

subject to the budget constraint, $R_{t+1}^{-1} D_{t+1} + C_t^s = D_t$, $\forall t \geq 0$.

Workers. The worker's preferences are the same. Different from savers though, workers begin each t with debt, B_t, and choose between current expenditures and future debt B_{t+1}. The choice B_{t+1} is limited by a natural debt limit, $\bar{B} = 1/(1-\beta)$. Because they are in debt, the worker must borrow intra-period to make spot expenditures, . Namely, she comes with B_t, but increases her debt to $B_t + S_t^w$ at the start of the period. This intra-period debt carries no interest. Its repayment is always feasible since total labor income always exceeds worker expenditures—because saver and worker expenditures add up to labor income in equilibrium.By the end of the period, intra-period debt is either paid or added to the balance of B_{t+1}, depending on the worker's overall expenditures. Critically, intra-period debt is limited by a time-varying spot-borrowing line (SBL), $\tilde{B}_t$:

$$S_t^w \leq \bar{S}_t \equiv \max\left\{\tilde{B}_t - B_t, 0\right\}. \tag{5}$$

An interpretation of $\tilde{B}_t$ is that it is a credit line that caps the amount of intra-period borrowing by $\bar{S}_t$.

The worker can consume without placing spot orders by placing chained orders. However, chained orders are costlier because $q \geq 1$.

Problem 2. (Worker Problem): *Given* B_0 *and* $\left\{R_{t+1}, \tilde{B}_t\right\}_{t\geq 0}$,

$$\max_{\{S_t^w, X_t^w\}_{t\geq 0}} \sum_{t\geq 0} \beta^t \log\left(C_t^w\right),$$

subject to the budget constraint, B_t+E_t^w=$R_{t+1}^{-1}B_{t+1}$+1, $\forall t \geq$0, *to the expenditure mix (4) and total consumption (3), to the intra- and inter-period constraints (5), and to the natural debt limit* $B_t \leq \bar{B}$.

Clearly, the worker will always prefer to make spot payments. However, if the worker has little intra-period borrowing capacity she has to make costly chained expenditures to reach a desired level of consumption.

Market Clearing. Clearing in the asset market requires:

$$D_t = B_t. \tag{6}$$

Recall that since savers do not work, they must always maintain positive savings, $D_t > 0$. Thus, without loss of generality, we work under the assumption that the worker is always in debt. Hence, from now on, we no longer refer to D_t, and use B_t to represent both saver deposit and worker debt. Given the consumption choices of both households, $\{X_t^w, S_t^w, C_t^s\}$, the goods-market clearing condition is:

$$C_t^s + S_t^w + X_t^w = \mathcal{Y}\left(\mu_t\right). \tag{7}$$

Adding both household's budget constraints yields an income-expenditure identity:

$$C_t^s + S_t^w + q_t X_t^w = 1. \tag{8}$$

Definition 2. *Given a sequence of* $\left\{\tilde{B}_t\right\}_{t\geq 0}$, *a sequence* $\{B_t, C_t^s, S_t^w, X_t^w, R_t, q_t\}_{t\geq 0}$ *is a* ***symmetric competitive equilibrium*** *if:*

1. *Given* $\{R_t, q_t\}_{t\geq 0}$, $\{B_t, S_t^w, X_t^w\}_{t\geq 0}$ *solves the worker's problem and* $\{B_t, C_t^s\}_{t\geq 0}$ *solves the saver's problem.*

2. *Markets clear; (6) and (7) are satisfied.*

In equilibrium, the ratio of chained expenditures satisfies the following payments identity:

$$\mu_t = q\left(\mu_t\right) \cdot X_t^w. \tag{9}$$

4.2 Characterization

Household problem solutions. The solution to the saver's problem is typical of log utility and goes without proof.

Proposition 3. *Given* B_0*, the solution to the saver's problem is:*

$$C_t^s = (1-\beta)\, B_t \text{ and } B_{t+1} = R_{t+1}\beta B_t \ \forall t \geq 0. \tag{10}$$

The worker's problem is more involved. Given total expenditures E_t^w, since $q_t \geq 1$, cost minimization requires the worker to spend spot as much as her SBL allows. Thus, given E_t^w, spot and chained expenditures are respectively:

$$S_t^w = \min\left\{\bar{S}_t, E_t^w\right\} \tag{11}$$

and

$$X_t^w = \left(E_t^w - \min\left\{\bar{S}_t, E_t^w\right\}\right)/q_t. \tag{12}$$

Given this optimal split, we invoke the principle of optimality to cast the worker's problem into a Bellman equation:

Problem 3. (Worker (Recursive) Problem): *Given* B_0 *and* $\left\{\tilde{B}_t, R_{t+1}, q_t\right\}_{t\geq 0}$*, workers choose a sequence of debt holdings* $\{B_{t+1}\}_{t\geq 0}$ *which follow from the solution to:*

$$V_t^w\left(B\right) = \max_{B'\leq \bar{B}} \log\left(S^w + X^w\right) + \beta V_{t+1}^w\left(B'\right) \tag{13}$$

where S^w and X^w are given by (11) and (12), respectively, and total expenditures by $E^w = B' R_{t+1}^{-1} + 1 - B$.

The time index in the value function reflects the time dependence on R_{t+1}, q_t, and $\tilde{B}_t$. The Bellman equation reformulates the worker's problem as an expenditure-savings problem. The solution yields total expenditures. Using total expenditures as inputs to the optimal expenditure rules, (11) and (12), we obtain optimal spot and chained consumption. The Bellman equation is key to understand the business-cycle implications of payments-chain networks. The following lemma identifies two threshold points that characterize the worker's expenditure mix in terms of their debt.

Lemma 1. (Expenditure Threshold Points): *Define the **efficiency threshold**, $B_{t+1}^{\star} \equiv R_{t+1}\left(\tilde{B}_t - 1\right)$. Then, $S_{t+1}^w = 0$ if and only if $B_{t+1} > \tilde{B}_{t+1}$. In addition, $X_t^w > 0$ if and only if $B_{t+1} > B_{t+1}^{\star}$.*

Proof. The proof is immediate. If $B_{t+1} > \tilde{B}_{t+1}$, the worker cannot spend spot. In turn, $B_{t+1} > R_{t+1}\left(\tilde{B}_t - 1\right)$ happens if and only if $E_t^w = B_{t+1}/R_{t+1} - B_t + 1 > \tilde{B}_t - B_t$. In that case, the worker spends sufficiently high that his chosen expenditure E_t^w requires him to spend on some chained orders.

□

It is convenient to define some relevant objects derived from these threshold points.

Definition 3. (Marginal Prices and Marginal inflation):

I. *The **average price** per unit of worker expenditure is $Q_t \equiv E_t^w / C_t^w$.*

II. *Given B', the price of the good at t bought with a marginal increase in B' is:*

$$\tilde{q}_t^E\left(B'\right) \equiv 1 + \left(q_t - 1\right) \mathbb{I}_{\left[B' \geq B_{t+1}^{\star}\right]},$$

the price of the good purchased at $t+1$ after a marginal decrease in B' is

$$\tilde{q}_{t+1}^B\left(B'\right) \equiv 1 + \left(q_{t+1} - 1\right) \mathbb{I}_{\left[B' > \tilde{B}_{t+1}\right]}.$$

*These terms define **marginal prices.***

III. ***Marginal inflation*** *is* $\Pi_{t+1}(B') \equiv \tilde{q}^B_{t+1}(B')/\tilde{q}^E_t(B')$*, a continuous function of* B' *except at discontinuity points* $\left\{B^\star_{t+1}, \tilde{B}_{t+1}\right\}$.

The average goods price is the ratio of expenditures to the quantity of goods the worker buys. Marginal prices have the following interpretation: If at t the worker spends on chained goods, the reduction in her debt is financed with a reduction in chained expenditures. Otherwise, if she does not spend on chained goods, a reduction in her debt is financed with a reduction in spot goods. Since the price of chained goods is q_t and the price of spot goods is 1, the worker sacrifices $1/\tilde{q}^E_t$ units of consumption per unit of debt reduction. Likewise, if she spends spot at $t+1$, any past savings can translate into spot expenditure at $t+1$. Otherwise, if she only spends in chained goods at $t+1$, any past savings translate into chained expenditures. Hence, the worker can buy $1/\tilde{q}^B_{t+1}$ additional goods by reducing her debt at t on the margin. Marginal inflation is the ratio of marginal prices, a definition that enters the following generalized Euler equation.

Proposition 4. (Workers's First-Order Condition): *Fix a sequence* $\left\{\tilde{B}_t, R_{t+1}, q_t\right\}_{t\geq 0}$ *such that* $\tilde{B}_t$ *is an increasing and* $\beta R_{t+1} \leq 1$*. Then, any solution* $\{B_{t+1}\}_{t\geq 0}$ *to the worker's problem satisfies the following generalized Euler equation:*

$$\frac{E^w_{t+1}}{E^w_t}\frac{Q_t}{Q_{t+1}} = \beta\frac{R_{t+1}}{\Pi_{t+1}(B_{t+1})} \quad \textit{if} \quad B_{t+1} \neq B^\star_{t+1} \tag{14}$$

and

$$\beta q_t R_{t+1} \geq \frac{E^w_{t+1}}{E^w_t} \geq \beta R_{t+1} \quad \textit{if} \quad B_{t+1} = B^\star_{t+1}.$$

This Euler equation is unconventional, but the interpretation is standard. The left-hand side is the ratio of marginal rates of substitution between t and $t+1$ consumption—expressed in expenditures over average prices.[25] The right-hand side captures the relation between discounting and rate of return, βR_{t+1}; the novelty is that it is deflated by marginal inflation. Marginal inflation enters the expression because this is the relevant ratio of prices that deliver marginal utilities at t and $t+1$, respectively $\tilde{q}^E_t$ and $\tilde{q}^B_{t+1}$. This generalized Euler equation holds exactly except at the discontinuity point $B^\star_{t+1}$. The inequalities that are satisfied when $B_{t+1} = B^\star_{t+1}$ correspond to a sub-differential optimality condition: the condition says that increasing debt is optimal to the left of $B^\star_{t+1}$,

[25] The ratio of expenditures to average prices, is the ratio of marginal utilities under log preferences.

but not to its right.[26]

Importantly, the Euler equation is a necessary condition for optimality, but not sufficient condition. The reason is that the objective function in the worker's problem, (13), is not concave in B'. This introduces further challenges. In particular, multiple (finite) sequences of B_{t+1} indexed by an initial choice of B_1, may satisfy the Euler equation. This property leads to the possibility of hysteresis in general equilibrium. Before explaining this, we present a characterization of the worker's problem in partial equilibrium.

Proposition 5. (Stationary Worker Problem): *Fix $\tilde{B}_t = B_{ss}$ and $R_t = \beta^{-1}$ and let $\tilde{B}_{ss} > 1$. Let $B^\star_{ss} \equiv B^\star\left(1/\beta, \tilde{B}_{ss}\right)$. Then, a solution to the worker's problem satisfies:*

I. *If $B_0 \leq B^\star_{ss}$, then $B_t = B_0 \; \forall t$.*

II. *There exists a threshold $B^h > B_{ss}$ such that:*

$\diamond$ *If $B_0 < B^h$, then $B_t \to B^\star_{ss}$ in finite time*

$\diamond$ *If $B_0 > B^h$, then $B_t = B_0 \; \forall t$.*

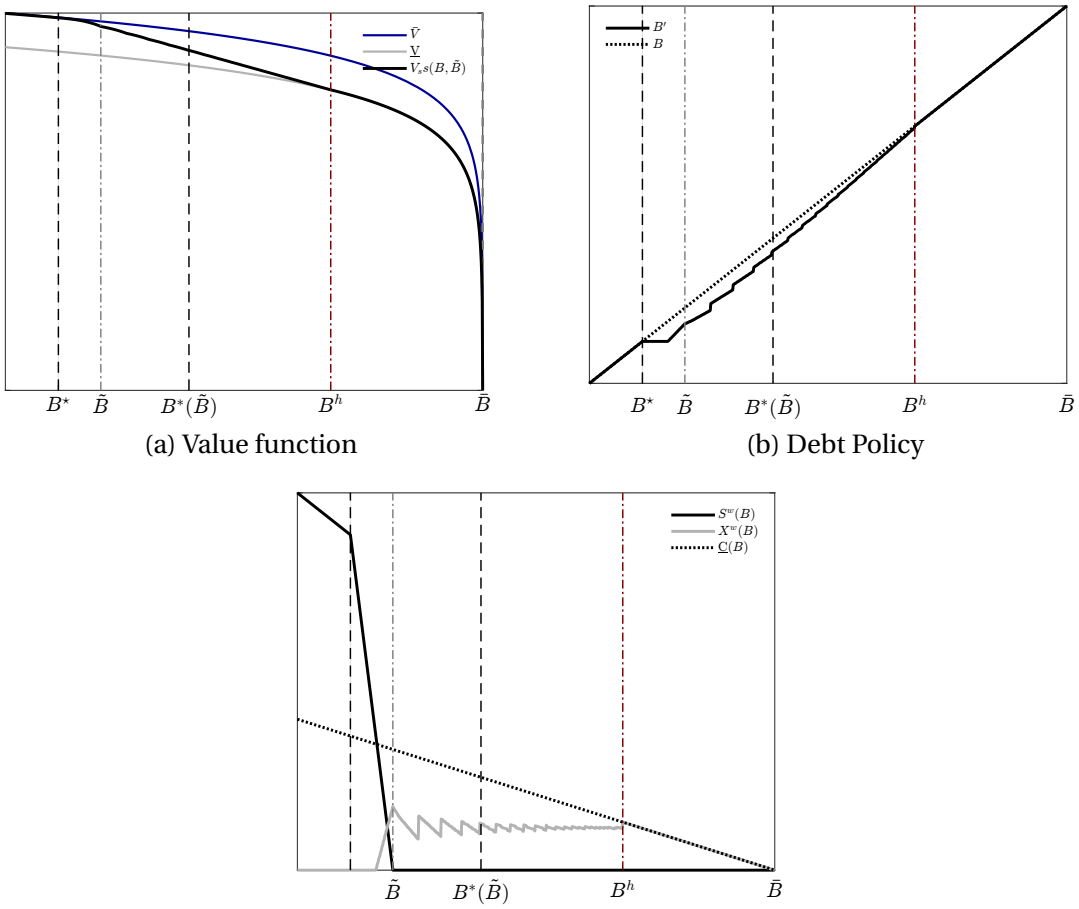
The threshold B^h and the convergent sequence is given in the proof.

The proposition characterizes the dynamics of B_t: a sharper characterization of convergence in terms of convergence times is provided in its proof. The proposition showcases the domains of attraction toward steady-state solutions. To illustrate the proposition, Figure 7 plots a numerical solution to the worker's Bellman equation and its policy functions when $R_{t+1} = \beta^{-1}$ and $\tilde{B}_t = \tilde{B}_{ss} > 0$. We also overlay the value functions corresponding when $\tilde{B}_t = 0$ and when $\tilde{B}_t = \infty$. We denote these value functions by $\underline{V}$and $\bar{V}$ respectively.

For debt levels below $B^\star$, V lies on top of $\bar{V}$. When $B < B^\star$, the worker can consume the annuity of his human capital minus the interest on his debt, entirely on spot consumption. This possibility yields the same value, $\bar{V}$, as if the SBL is never binding, $\tilde{B}_t = \infty$. When debt is above a threshold B^h, the value function lies on top of $\underline{V}$. When

[26]There is also a discontinuity in the Euler equation at $B_{t+1} = \tilde{B}_{t+1}$, but the corresponding subdifferential does not yield an optimality condition.

$B > B^h$, the worker can consume the annuity of his human capital minus the interest on his debt, but in this case, entirely on chained consumption, paying an effective price $q > 1$. When $B > B^h$, even though the worker could potentially save to enjoy better prices in the future, she has no incentives to save. She has no incentives to save because the benefits of deleveraging come too far in the future and require an excessive sacrifice of current consumption. In that case, her value function is $\underline{V}$, as if she could never access short-term funding. For moderate debt levels, when $B \in (B^\star_{ss}, B^h]$, the worker will delever until her debt falls below $B^\star$. All in all, saving away from financial constraints happens if the SBL constraint is not too tight. This property leads to the possibility of hysteresis.



(a) Value function

(b) Debt Policy

(c) Consumption Policy

Figure 7: Bellman Equation and Policy Functions: Worker Problem

Note: Figures are calculated using value function iteration: $\beta = 0.8$, $q = 1.75$ and $\tilde{B} = 0.4 \cdot \bar{B}$.

Equilibrium Dynamics: From Sequential to a Functional Representation. We now analyze the equilibrium dynamics of B_t. The equilibrium is recursive with state variable $B_t \times \tilde{B}_t \times \tilde{B}_{t+1} \in [0, \bar{B}]^3$. A recursive representation requires that we have a relationship between a sequence and a recursive formulation: a variable m_t, can be expressed as an equilibrium function, $m : [0, \bar{B}]^3 \to \mathbb{R}_+$ such that $m_t = m\left(B_t, \tilde{B}_t, \tilde{B}_{t+1}\right)$. Hence, from now we use m to represent the function that maps the state into its equilibrium value m_t. We also adopt the convention of using m' to refer to m_{t+1}.

Using the recursive formulation, from (10), (8), and (11) we have:

$$C^s(B) = (1-\beta)B, \quad E^w(B) = 1-(1-\beta)B, \tag{15}$$

and

$$S^w\left(B, \tilde{B}\right) = \min\left\{\bar{S}\left(B, \tilde{B}\right), E^w(B)\right\}.$$

Several other equilibrium functions are deduced from these functions directly: $\mu = E^w - S^w$, $q \equiv \mathcal{A}^{-1}(\mu)$, $X^w \equiv \mu/q$, $Q \equiv E^w/C^w$, and, finally, $C^w \equiv S^w + X^w$.

The only endogenous argument of the state is B_t. Thus, we need a map $\mathcal{B}$ from the current state to its future value, $B' = \mathcal{B}\left(B, \tilde{B}, \tilde{B}'\right)$. If we obtain that map, the equilibrium rate will satisfy $R_{t+1} = \mathcal{R}\left(B_t, \tilde{B}_t, \tilde{B}_{t+1}\right)$ where $\mathcal{R}\left(B, \tilde{B}, \tilde{B}'\right) \equiv \beta^{-1}\mathcal{B}\left(B, \tilde{B}, \tilde{B}'\right)/B$, following the saver's optimal rule. With that equilibrium function, we can define the threshold function $B^\star\left(B, \tilde{B}, \tilde{B}'\right) \equiv \mathcal{R}\left(B, \tilde{B}, \tilde{B}'\right) \cdot \left(\tilde{B} - 1\right)$. We solve for $\mathcal{B}$ below, after defining a functional representation for the equilibrium prices:

$$\Pi\left(B'; B, \tilde{B}, \tilde{B}'\right) \equiv \tilde{q}^B\left(B'; \tilde{B}'\right)/\tilde{q}^E\left(B'; B, \tilde{B}, \tilde{B}'\right),$$

where

$$\tilde{q}^E\left(B'; B, \tilde{B}, \tilde{B}'\right) \equiv 1 + \left(q\left(B, \tilde{B}\right) - 1\right) \cdot \mathbb{I}_{\left[B' \geq B^\star\left(B, \tilde{B}, \tilde{B}'\right)\right]}$$

and

$$\tilde{q}^B\left(B'; \tilde{B}'\right) \equiv 1 + \left(q\left(B', \tilde{B}'\right) - 1\right) \cdot \mathbb{I}_{\left[B' > \tilde{B}'\right]}.$$

To find $\mathcal{B}$, we combine the worker and saver Euler equations, (14) and (10). We then substitute out R_{t+1} and use the equilibrium worker expenditures (15), to obtain

an equilibrium condition B':

$$\underbrace{\frac{B}{1-(1-\beta)\,B}\cdot Q\left(B,\tilde{B}\right)}_{\equiv\mathcal{E}\left(B;\tilde{B}\right)}=\underbrace{\frac{B'}{1-(1-\beta)\,B'}\frac{Q\left(B',\tilde{B}'\right)}{\Pi\left(B';B,\tilde{B},\tilde{B}'\right)}}_{\equiv\mathcal{E}'\left(B';B,\tilde{B},\tilde{B}'\right)}. \tag{16}$$

The left and right hand sides define functions equations $\mathcal{E}\left(B';B\right)$ and $\mathcal{E}'\left(B';B,\tilde{B},\tilde{B}'\right)$. At equality, these functions define the law of motion of debt. Because Q is discontinuous, there may be multiple solutions B'. The following result tells us that in a transition, the lowest value is valid.

Proposition 6. (Equilibrium Rates and Expenditures): *Consider a weakly monotone increasing sequence of spot borrowing lines $\tilde{B}_t \to \tilde{B}_{ss}$. For any $B_0 < B^h\left(\tilde{B}_{ss}\right)$, if an equilibrium exist, then $B_{t+1}=\mathcal{B}\left(B_t,\tilde{B}_t,\tilde{B}_{t+1}\right)$ where*

$$\mathcal{B}\left(B,\tilde{B},\tilde{B}'\right)=\max\left\{B^{\star}\left(\tilde{B}\right),\ \arg\min_{B'}\left\{\mathcal{E}\left(B';B,\tilde{B},\tilde{B}'\right)=\mathcal{E}'\left(B';B,\tilde{B},\tilde{B}'\right)\right\}\right\}.$$

Proposition 6 is key to describing the dynamics.[27]

Steady States. We use the subscript ss to denote steady states. In principle, workers could make both spot and chained expenditures in steady state. However, from the saver's problem we know that in steady state, $R_{ss}=\beta^{-1}$. To be consistent with the worker's law of motion for debt, the worker's marginal inflation must also equal one. Hence, in a steady state, the worker makes either only spot expenditures or only chained expenditures. We define an **undisrupted steady state** as a steady state with only spot expenditures, and $\mathcal{Y}_{ss}=1$. In turn, we define an **disrupted steady state** as a steady state where workers only make chained expenditures, and $\mathcal{Y}_{ss}<1$. The following corollary presents a condition that guarantees that the economy is in an undisrupted steady state.

[27] Implicitly, the proposition yields an algorithm to compute equilibria. Starting from, B_0, the sequence of debt generated in equilibrium is given by $\mathcal{B}\left(B,\tilde{B},\tilde{B}'\right)$, which is the smallest solution B' to the equation $\mathcal{E}\left(B';B,\tilde{B},\tilde{B}'\right)=\mathcal{E}'\left(B';B,\tilde{B},\tilde{B}'\right)$. For each B, we obtain B' and update the state accordingly.

Corollary 1. *Fix $\tilde{B}_t = \tilde{B}_{ss}$. For any $\tilde{B}_{ss} > 0$, the economy is in an undisrupted steady state for any $B_t \leq B^*_{ss}$.*

If $B_t \leq B^*_{ss}$ the worker can borrow (intra-period) more than his current income net of interests. If that is is the case, the economy is in an undisrupted steady state.[28]

Convergence Toward Undisrupted Steady States. Next, we describe the domain of attraction toward undisrupted steady states. The domain of attraction toward undisrupted steady states is the region for which the equation $\mathcal{E}(B';B) = \mathcal{E}'\left(B';B,\tilde{B},\tilde{B}'\right)$ has a solution where $B' < B$. This region has an upper bound:

$$B^*\left(\tilde{B}\right) \equiv \tilde{B}/\left(C^w\left(\tilde{B},\tilde{B}\right) + C^s\left(\tilde{B}\right)\right) \geq \tilde{B}.$$

This inequality follows from total consumption being less than total expenditures. We have the following.

Corollary 2. *Let $\tilde{B}_t = \tilde{B}_{ss}$. For any $B_0 < B^*\left(\tilde{B}_{ss}\right)$, $B_{t+1} < B_t$ if $B_t \in \left(B^\star(\beta^{-1},\tilde{B}_{ss}),\ B^*(\tilde{B}_{ss})\right)$ and $B_{t+1} = B_t$ if $B_t \leq B^\star(\beta^{-1},\tilde{B}_{ss},\tilde{B}_{ss})$. If, in addition, $\tilde{B}_{ss} > 1$ the economy converges toward an undisrupted steady state in finite time (and approaches zero debt if $\tilde{B}_{ss} \leq 1$).*

Corollary 2 describes the domain of attraction toward undisrupted steady states. If the worker holds debt between $\tilde{B}$ and $B^*\left(\tilde{B}\right)$ she repays her debts.[29] The deleverage will continue until she reaches a steady state debt once B^*_{ss}. An implication is that payments-chain crises are only temporary in the domain of attraction toward undisrupted steady states.

Figure 8 describes a transition toward an undisrupted steady state. The left panel of the figure plots different debt levels in the x-axis, holding $\tilde{B}$ fixed. The solid blue and dashed gray curves plot the functions $\mathcal{E}$ and $\mathcal{E}'$ correspondingly. The arrows in the figure trace the path of debt generated in equilibrium, following the sequence of solutions to B' implicit in Proposition 6. For initial conditions where $B_0 > B^*_{ss}$, the economy fails to converge. In that region, the only solution to $\mathcal{E} = \mathcal{E}'$ happens at $B' = B$. The middle panel plots the equilibrium interest rate, $R\left(B,\tilde{B},\tilde{B}'\right)$. The right panel highlights a

[28] Observe that if $\tilde{B}_{ss} \leq 1$, then $B^\star\left(\beta^{-1},\tilde{B}_{ss}\right) > 0$ in which case, no undisrupted steady state exists.

[29] Indeed, they will delever at the rate $\mathcal{R} < \beta^{-1}$ consistent with the condition in Proposition 4. In this region, equation (16) may have two solutions, but only the lowest value of B' is optimal.

gray area in the phase diagram, this is the area of hysteresis where a transition to the efficient steady state fails.

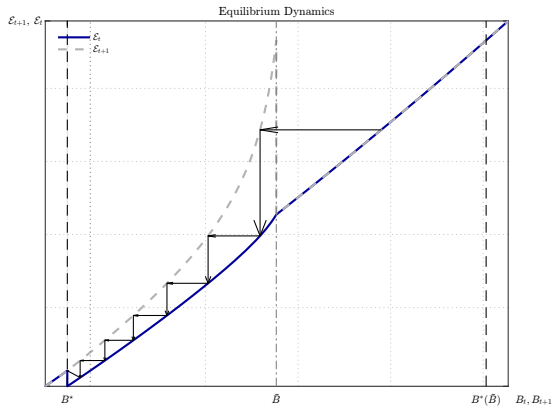
Proposition 6 and Corollary 2 help us interpret transitions when the exogenous borrowing limit $\left\{\tilde{B}_t\right\}$ is transitions toward a steady state. Indeed, we can interpret the equilibrium path as the response to a credit crunch event. We know that as long as $B_0 \leq B^*_{ss}$ and $\tilde{B}_{ss} > 1$, the economy converges to an undisrupted steady state. Figure 9 presents an example of a transition. The figure illustrates that there are different phases in a transition. In its extreme phase, at the beginning of the transition, TFP remains depressed at its lowest value because workers only spend by making chained orders and debt remains constant. In the smoothing phase, workers anticipate that they will be able to spend on spot goods in the subsequent period. To smooth consumption, they increase current chained expenditures taking in more debt. This jump in debt is followed by a recovery phase, where workers trade off consumption smoothing and against paying off debt to increase their credit lines and spend spot. The repayment phase corresponds to the last period with positive chained expenditures. Eventually, the economy converges to an efficient steady state with less debt than at the start of the transition. This pattern holds generically.

Debt-Overhang and Hysteresis of payments-chain Crises. we now study hysteresis provoked by debt overhang.

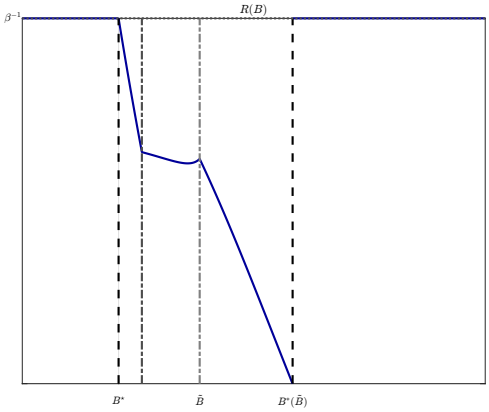
Corollary 3. (Hysteresis): *Let $\tilde{B}_t = \tilde{B}_{ss}$. For $B_t \geq B^h\left(\tilde{B}_{ss}\right)$, the economy is permanently in a disrupted steady state.*

Hysteresis occurs when debt is so large that the consumption sacrifice does not compensate the benefits deleveraging. Recall from Figure 7 that for sufficiently high debt levels, the worker's value function yields the same value as if $\tilde{B}_t = 0$ forever. When these value functions overlap, the economy experiences debt overhang; she prefers to stay put. The value functions are presented for the case where the rate is $R_{t+1} = \beta^{-1}$. Of course, in general equilibrium, the rate is endogenous. Figure 8 depicts the region of hysteresis in the gray area, and shows that the hysteresis region, equilibrium rates are indeed, $R_{t+1} = \beta^{-1}$.[30]

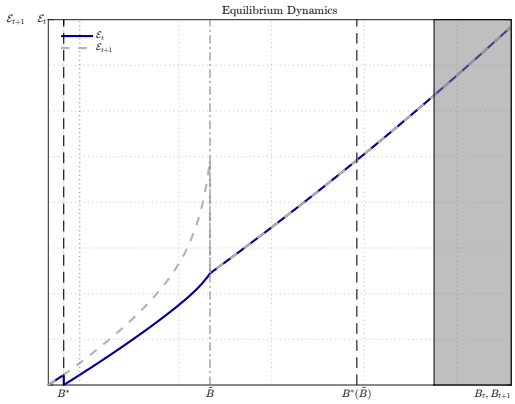
[30] For that rate, the only solution to the equilibrium condition $\mathcal{E} = \mathcal{E}'$ happens at $B' = B$.



(a) Transition Plot



(b) Equilibrium Rates



(c) Hysteresis Region

Figure 8: Equilibrium Phase Diagram and Hysteresis Region

Note: Figures are calculated using value function iteration: $\beta = 0.95$, $\delta = 0.9$, and $\tilde{B} = 0.2 \cdot \bar{B}$. Panels (a) and (c) are the phase diagram of B_t constructed by using $\mathcal{E}\left(B';B\right)$ and $\mathcal{E}'\left(B';B,\tilde{B},\tilde{B}'\right)$. Panel (b) plots $R\left(B,\tilde{B},\tilde{B}'\right)$ in the range of values $[B^\star - 1, B^* + 2]$. The shaded are in Panel (c) corresponds to the hysteresis region.

If debt starts in hysteresis region, it never falls. Excessive debt hampers the ability to make payments quickly, and the economy ends up in a permanent state of under-capacity production. With deterministic dynamics, hysteresis is a transient state. In a more general setting, a hysteresis-like region where agents don't delever, may be reached via shocks that provoke excess optimism or low discount factors. This form of debt overhang rationalizes several verbal descriptions of the Japanese lost decades or the aftermath of the Euro sovereign debt crisis.

Finally, we note that the domain of attraction of undisrupted steady states and hysteresis are disconnected. This implies that when debt falls between B^*_{ss} and B^h, a symmetric competitive equilibrium does not exist. Exploring multiplicity of equilibria or

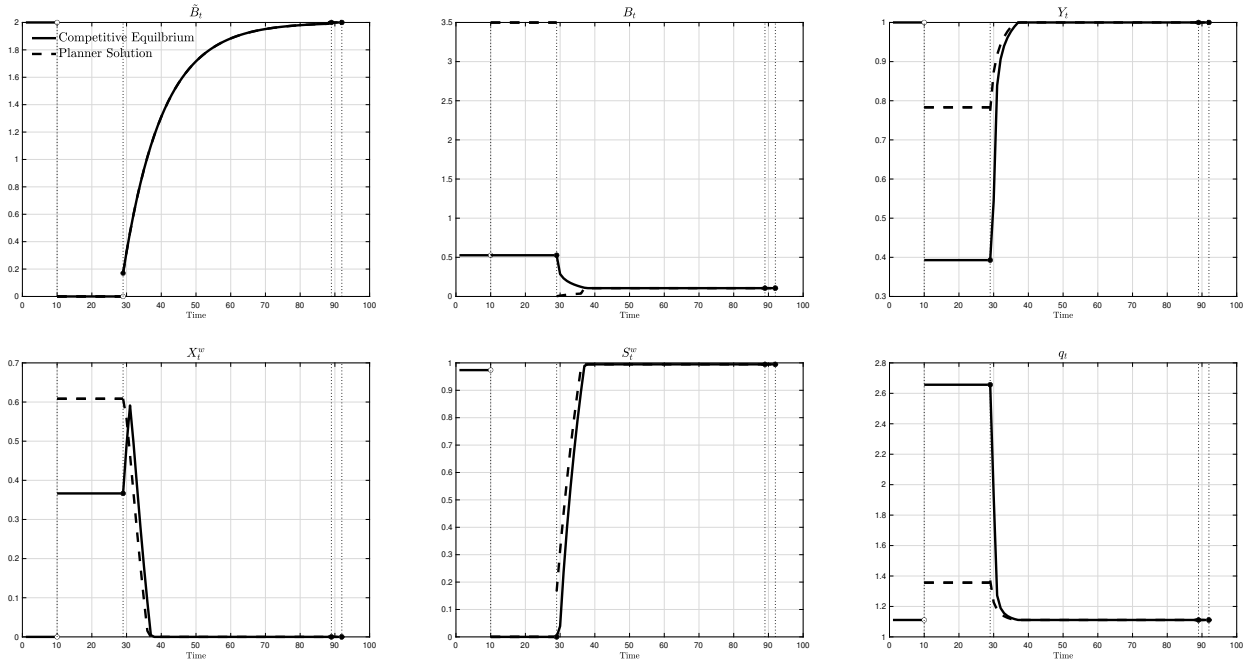


Figure 9: Efficient and Competitive Transition after a Smooth Credit Crunch.

Note: This figure reports a numerical example of a credit crunch episode. Figures are calculated using value function iteration: $\beta = 0.95$, $\delta = 0.9$, and $\tilde{B}_{ss} = 0.1 \cdot B^{\star}$. The crunch happens at period t=10. For $t \in \{11, 29\}$,$\tilde{B}_t = 0$. Then $\tilde{B}_t$ returns to steady state according to an AR(1) process with coefficient 0.9. The transition ends at t=90 when $\tilde{B}_t = \tilde{B}_{ss}$.

solutions with asymmetric behavior is left for future work.[31]

Discussion: Interpretation of Financial Constraints. Saver spot payments represent uses of internal liquid funds, i.e. deposits. The short-term funding limit, $\max\left\{\tilde{B}_t - B_t, 0\right\}$, represent 'credit lines", i.e. credit card limits, overdraft facilities, or supply chain finance facilities. Thus, worker spot payments are uses of those lines.[32] In turn, chained payments can be thought of as outstanding account receivables against goods to be delivered. In practice, many account receivables generate from contracts in which the product is not delivered until a fraction of the payment is anticipated. Thus, they rep-

[31] Note that in this middle region, at an individual level, workers would want to delever as seen in Figure 7. In general equilibrium, this would imply a rate below β^{-1}, for which there are no solutions to equation (16), as shown in Figure 8. The economy may possibly feature sunspot equilibria; a situation we do not consider in this paper.

[32] Other examples of facilities are Business Credit Lines, Standby Letters of Credit and Supplier Finance Programs. Under a Supplier Finance Program, the buyer wants to pay later, whereas suppliers request cash. Supplier Finance enables suppliers to be paid by banks against the receivables. Descriptions of these programs are offered by some of the largest financial institutions: J.P. Morgan Supplier Finance Facility or Citibank Supplier Finance Facility.

resent production agreements.

The SBL,$\tilde{S}$, is different from the debt limit, $\bar{S}$, the hard debt limits that is most prevalent in macro finance models. The $\tilde{B}$ limit applies only to short-term funding, but the worker has the option to roll over its repayment until the debt limit is reached. In the section, we studied contractions in the SBL, while keeping the debt limit constant. There is an economic motivation behind this modeling of credit crunch events: if credit is intermediate by banks, if a bank wants to cut back on credit, it may be convenient to tighten the SBL, but not necessarily to force workers to repay debt principals. If loan repayment is suddenly forced, it can trigger costly defaults.

Finally, it is worth discussing the misallocation of funding. Due to the big family assumption, the worker will receive labor income flows while there are still pending chained orders. This happens because the worker may receive payment on spot orders before it receives payments chained to chained expenditures. The big family assumption is present to avoid a distribution of ex-post outcomes while keeping the simplicity of the TFP function derived earlier. Chained orders would still induce delays even if the worker pool funds from different orders.[33] The option to pool funds will change the exact functional form of the TFP function but not its essential features.

5. Policy Implications of Payments-Chain Crises

5.1 Constrained Inefficiency

In Section 4, we expressed the chained expenditure ratio as a function of debt levels. In Section 3 we found that higher chained expenditure ratios lead to delays which provoke declines in measured TFP. In part, those losses result from the poor organization of payments chains, given a level of chained expenditures. Ideally, a planner would reorganize payments so that each chain is of equal length, but governments do not have that power. What governments can do is influence spending decisions. This sec-

[33] For the case where chained orders can be paid with any incoming payment, the chain length distribution can again be derived analytically. In that case, for $N^x > N^s$, the minimal chain length would be zero, the maximal chain length $\lceil N^x/N^s \rceil$, and the probability distribution uniform among chains with $n \leq \lfloor N^x/N^s \rfloor$. For $N^x \leq N^s$, the minimal chain is zero with probability $N^s - N^x$ and the maximal chain length 1.

tion studies such policies. Traffic regulation provides a useful analogy: governments cannot assign drivers into different lanes, but they can tax vehicles. The spirit of the exercise is not normative but, rather, to clarify the sources of constrained inefficiency.

Ideal Pareto Weights. To study constrained inefficiency, consider a transition in a competitive equilibrium that reaches an undisrupted steady-state debt level B_{ss} starting from a debt level B_0. We study a Ramsey planner problem with Pareto weights on workers, θ, which delivers the same steady-state debt level as the competitive equilibrium:

$$\frac{1-(1-\beta)B_{ss}}{(1-\beta)B_{ss}} = \frac{\theta}{1-\theta}. \tag{17}$$

Since the economy is efficient in an undisrupted steady state, any difference between the transition path of debt in a planner and competitive equilibrium solutions uncover a constrained inefficiency only during transitions.

The Ramsey Problem. Consider a sequence of debt taxes $\left\{\tau_t^k\right\}$, labor taxes $\left\{\tau_t^\ell\right\}$, and expenditure taxes $\{\tau_t^c\}$. The Ramsey Problem is:

Problem 4. (Ramsey Problem): *Given* B_0 *and* $\left\{\tilde{B}_t\right\}$*:*

$$\max_{\left\{\tau_{t+1}^k,\tau_t^c,\tau_{t+1}^\ell\right\}_{t\geq 0}} \sum_{t\geq 0} \beta^t \left[(1-\theta)\log\left(C_t^s\right) + \theta\log\left(C_t^w\right)\right],$$

subject to the (modified) saver budget constraint and optimality:

$$\left(1+\tau_{t+1}^k\right) R_{t+1}^{-1} B_{t+1} + \left(1+\tau_t^c\right) C_t^s = B_t,\ \forall t \geq 0,$$

to the (modified) worker budget constraint and optimality:

$$B_t + \left(1+\tau_t^c\right) E_t^w = R_{t+1}^{-1} B_{t+1} + 1 - \tau_t^\ell,\ \forall t \geq 0,$$

to the government's budget constraint:

$$\tau_{t+1}^k R_{t+1}^{-1} B_{t+1} + \tau_t^c \left(C_t^s + E_t^w\right) + \tau_{t+1}^\ell = 0,\ \forall t \geq 0,$$

and the structure of transactions: (i) optimal expenditures (11-12), (ii) total consumption (3), (iii) the optimality conditions of the worker and saver problems, (iv) the payments constraints (9), and (v) the shadow price, $q_t = \mathcal{A}\left(\mu_t\right)^{-1}$.

The Ramsey planner distorts the competitive equilibrium by using credit, labor, and expenditure taxes. This planner cannot distinguish between the two expenditure forms, but takes into account the agents' optimal behavior, their constraints, market clearing, respects the transactions technology, and satisfies budget balance.

The Primal Problem. we solve the Ramsey Problem by solving an equivalent Primal Problem.

Proposition 7. (Solution to Ramsey): *The allocations induced by a solution to the Ramsey Problem are the same allocations as the solution to the following problem:*

Problem 5. (Primal Planner Problem):*Taking* $\left\{\tilde{B}_t\right\}$ *as given:*

$$\max_{\{B_t\}_{t\geq 0}} \sum_{t\geq 0} \beta^t \mathcal{P}\left(B_t, \tilde{B}_t\right)$$

where

$$\mathcal{P}\left(B, \tilde{B}\right) \equiv (1-\theta)\log\left(C^s\left(B\right)\right) + \theta \log\left(S^w\left(B, \tilde{B}\right) + X^w\left(B, \tilde{B}\right)\right). \tag{18}$$

Let $\left\{B^p_{t+1}\right\}_{t\geq 0}$ *be a solution to the Primal Planner Problem. The solution to the Ramsey problem can be implemented setting* $(1+\tau^c_0) = B^p_0/B_0$ *and a sequence of debt taxes (a formula is given in the proof).*

The proposition asserts that the solution to the Ramsey Problem can obtained from the solution of a Primal Problem where the planner directly chooses the sequence of debt. This relation follows because the constraint set in the Primal Problem includes the constraints of the Ramsey Problem.[34] Hence, if a solution to the primal can be implemented with taxes, it solves the Ramsey Problem. The proposition shows that this is the case.

[34]This is immediate since market clearing in the asset market and the budget balance, implies, by Walras's law, that the resource constraint holds. In the implementation, capital taxes and a period-zero expenditure tax are required by the Ramsey planner to distort the evolution of debt and replicate the solution to the Primal Problem.

The implementation of the Primal Problem is possible because the Ramsey planner can set capital taxes to induce a desired path of debt $\{B_t\}$. Controlling B_t is key to the implementation. A key property of the environment is that once B_t is known, the time t allocation is determined: Current B_t determines the savers' expenditures through the constant expenditure rule of log utility. Since worker income is always one, and aggregate income equals aggregate expenditures, worker expenditures are given once we know saver expenditures. This is why the objective in the Primal Problem is separable in time (18).

Time separability also implies that expenditure taxes are redundant after $t = 0$. All that the Ramsey planner needs is a sequence of capital taxes to control the path of debt and the labor tax to redistribute the capital tax receipts toward the worker. The planner only needs expenditure taxes at $t = 0$ because debt is predetermined at $t = 0$. Because the Ramsey planner can implement a time-separable primal planner problem, an immediate lemma is that the optimal debt level at time t has a closed form that only depends on the SBL level, $\tilde{B}$, at that period:

Lemma 2. (Static Primal Problem): *The solution to the Primal Planner Problem* $B_t = B^p\left(\tilde{B}_t\right)$ *where* B^p *is the solution to the following static problem:*

$$\mathcal{P}^o\left(\tilde{B}\right) = \max_{B\in[0,\bar{B}]} \mathcal{P}\left(B, \tilde{B}\right).$$

The solution to this static problem is key to understand the inefficiencies of this environment.

Proposition 8. (Solution of the Primal Problem): *The solution to the static problem* B^p *has the following property:*

we. Efficient Allocation. *For* $\tilde{B} \geq \frac{1-\theta\beta}{1-\beta}$*, the planner's problem is unconstrained:* $\mathcal{A} = 1$ *and expenditures are is in steady state (17).*

II. Inefficient Allocations. *For* $\tilde{B} < \frac{1-\theta\beta}{1-\beta}$*, the planner constraints bind. The planner may or may not distort TFP and induce more or less debt than in steady state:*

II.a. Inefficient Insurance and Efficient Production. *For $\tilde{B} \in \left[\tilde{B}^i, \frac{1-\theta\beta}{1-\beta}\right)$, the planner induces fewer saver expenditures than in steady state, $B^p = B^\star\left(\tilde{B}\right) < B_{ss}$ but production is efficient $\mathcal{A} = 1$. A marginal decrease in debt at B^p increases efficiency.*

II.b. Inefficient Insurance and Inefficient Production *If $\tilde{B}^s < \tilde{B}^i$, production is inefficient, $\mathcal{A} < 1$.*

II.b.i Inefficient Insurance and Inefficient Production — Conflicting Case. *If $\tilde{B} \in \left[\tilde{B}^s, \tilde{B}^i\right)$, the planner induces fewer saver expenditures than in steady state, $B^p < B^\star\left(\tilde{B}\right) < B_{ss}$. B^p is the unique solution to*

$$\frac{1-(1-\beta)B}{(1-\beta)B} = \frac{\theta}{1-\theta}\frac{Q\left(B,\tilde{B}\right)}{q\left(B,\tilde{B}\right)}\left(\frac{q\left(B,\tilde{B}\right) - \beta\left(1+\epsilon_\mu^{\mathcal{A}}\left(\mu\left(B,\tilde{B}\right)\right)\right)}{1-\beta}\right). \tag{19}$$

In this interval, a marginal decrease in debt at B^p increases efficiency.

II.b.ii Inefficient Insurance and Inefficient Production — Reinforcing Case. *For $\tilde{B} \in \left[0, \tilde{B}^s\right]$, the planner's problem yields a constant lowest value. The planner induces more saver expenditures than in steady state. B^p is the unique solution to:*

$$\frac{1-(1-\beta)B}{(1-\beta)B} = \frac{\theta}{(1-\theta)}\left(1+\epsilon_\mu^{\mathcal{A}}\left(1-(1-\beta)B\right)\right). \tag{20}$$

In this interval, a marginal decrease in debt at B^p decreases efficiency.

The solution to the Primal Problem reveals a novel insight. Namely, during transitions away from payments-chain crises, debt may be inefficiently high or inefficiently low. The intuition behind this is that, for a given debt level, the planner can increase output by reducing or increasing debt. By reducing debt, the planner distributes wealth toward the worker and frees credit lines inducing greater spot expenditures. Alternatively, by increasing debt, the planner distributes wealth toward the saver also stimulating spot expenditures. Both distributive policies distort social insurance, but increase efficiency by speeding up production.

The analytic expression for B^p reveals how the planner allocates expenditures, balancing productive efficiency against social insurance differently, depending on whether $\tilde{B}$ falls in several intervals. There are four intervals of values of $\tilde{B}$ where the planner's solution is qualitatively different. These actual critical values depend on the threshold points $\frac{1+\theta\beta}{1-\beta}$ and $\left\{\tilde{B}^s, \tilde{B}^i, \right\}$ whose formulas are found in the proof.

The first interval of values is where $\tilde{B}$ is a above a threshold level above which the SBL constraint is not binding. In particular, if $\tilde{B} \geq \frac{1+\theta\beta}{1-\beta}$, the planner sets $B = B_{ss}$, and production is efficient. Furthermore, the ratio of marginal utilities equals the ratio of Pareto weights—efficient social insurance is satisfied condition (17).

If $\tilde{B}$ is below the efficiency threshold, $\frac{1+\theta\beta}{1-\beta}$, the planner distorts either social insurance or productive efficiency. The novelty is that the planner has two ways to increase productive efficiency: the planner can increase efficiency on the margin, either by distributing wealth toward the worker if $B < \tilde{B}$ or by distributing wealth toward the saver if $B > \tilde{B}$. To see this, observe that in the region where $B < \tilde{B}$, any further reduction in debt translates, on the margin, into more spot expenditures by the worker and, thus, increases efficiency. In the region where $B > \tilde{B}$, any increase in debt translates, on the margin, into more spot expenditures by the saver, also increasing efficiency. Because of this ambivalent nature, the planner's objective function $\mathcal{P}\left(B, \tilde{B}\right)$ is not concave in B, leading to bang-bang solutions as $\tilde{B}$ varies.

Figure 10 for describing the economics of the Primal Problem, in the regions away from efficiency. The left panel plots B^p for different values of $\tilde{B}$. In the interval of values where $\tilde{B}$ falls in a second interval, i.e. when $\tilde{B} \in [\tilde{B}^i, \frac{1+\theta\beta}{1-\beta})$, the planner solution induces productive inefficiency, $\mathcal{A} = 1$, but but the ratio of marginal utilities no longer the ratio of Pareto weights, as in (17). To understand why the planner does not distort production, observe that because $B^\star\left(\tilde{B}\right) < B_{ss}$, productive efficiency can only be achieved only by setting $B^p = B^\star\left(\tilde{B}\right) < B_{ss}$. Setting debt at $B^\star\left(\tilde{B}\right)$, which is less than its steady-state value, implies that saver expenditures are less than in steady-state. Thus, in this interval the planner sacrifices social insurance by redistributing wealth toward the worker to guarantee productive efficiency. The planner's solution is at a corner because the derivative of his objective is discontinuous at $B^\star\left(\tilde{B}\right)$. This discontinuity follows from the payments-chain network structure: the property that $\mathcal{A}(0) < 1$, which

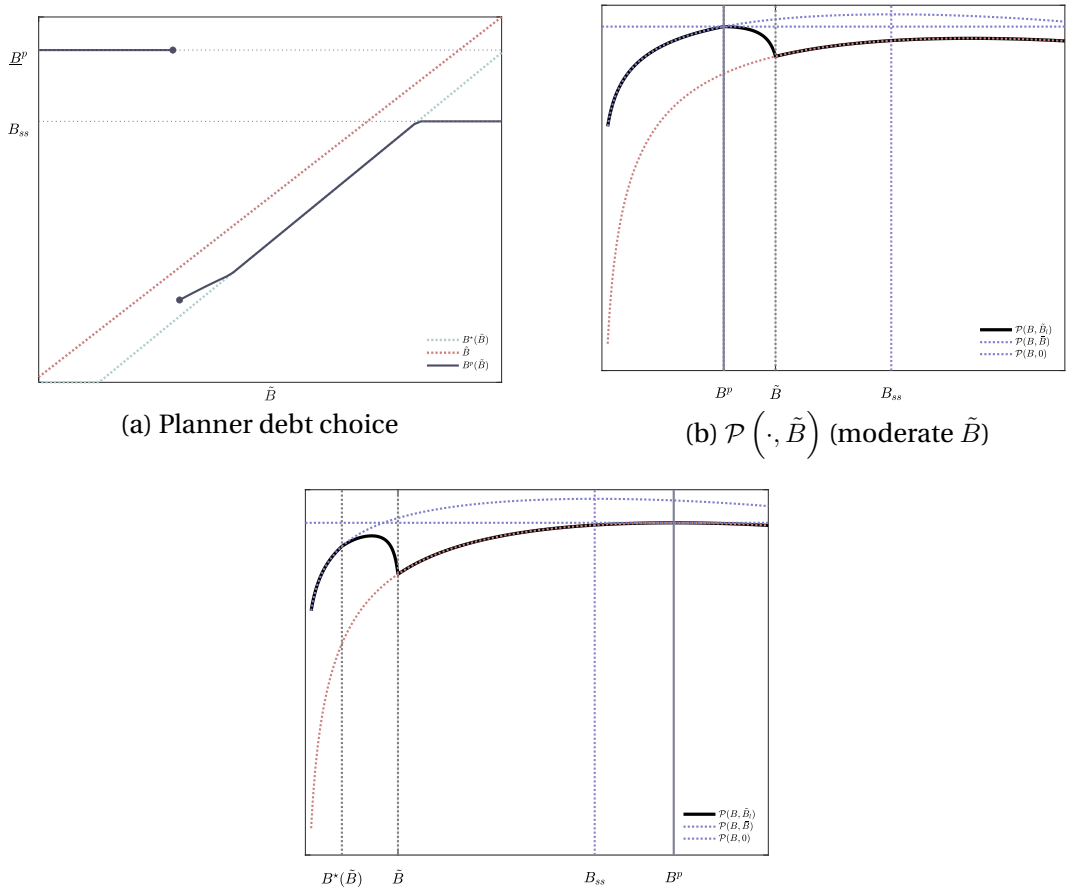
captures that even when the chained expenditure ratio is zero, and individual chained order features a production delay.

When $\tilde{B}$ is in a further tighter and falls in the third interval, $\tilde{B} \in \left[\tilde{B}^s, \tilde{B}^i\right)$, the Planner begins to sacrifice productive inefficiency. To maintain productive efficiency, for those levels of the SBL, the planner would have to redistribute even more wealth to the worker at the expense of saver expenditures. In that region, sacrificing social insurance only is not worth it. Hence, the planner sets debt above its efficient level, thus $B^\star\left(\tilde{B}\right) < B^p < \tilde{B}$. Therefore, in this interval, the planner still redistributes wealth (relative to steady state) toward the agent facing the financial constraint, but the planner does partially sacrifice productive efficiency. The planner balances productive efficiency and social insurance, as given by equation (19).

When $\tilde{B}$ falls below an extreme value, in the fourth interval, $\tilde{B} < \tilde{B}^s$, the nature of the planner's solution changes dramatically. Increasing productive efficiency by redistributing wealth toward the worker requires the planner to set debt below $\tilde{B}$. However, when $\tilde{B} < \tilde{B}^s$, setting debt so low would requires an extremely high sacrifice of saver consumption. Below that threshold, the planner prefers to redistribute wealth away from the worker, the constrained agent. Since the planner sets $B > \tilde{B}$, the planner induces only chained expenditures by the worker. Once the worker only makes chained expenditures, the SBL becomes irrelevant, so the planner chooses a constant debt level in this region. This debt level is higher than the unconstrained ideal debt level B_{ss} and given by Equation (20). Debt is higher than in steady state because the planner understands that a marginal increase in debt increases the wealth and, therefore, spot expenditures of savers. This increases productive efficiency.

The ambivalent nature of the planner's problem is germane to the nature payments-chain crises. In typical models with pecuniary externalities, a planner wants to rebalance wealth toward the financially constrained agents to increase productive efficiency. Thus, typically, social insurance and productive efficiency reinforce each other. Here, for extremely low values of $\tilde{B}$, the planner switches to a policy mix where social insurance and efficiency are in conflict. The middle and right panels of Figure 10 illustrates the change in the planner's strategy: In the middle panel, the planner prefers a value of debt where social insurance and efficiency complement. The right panel shows how

he switches strategy as the SBL is tighter.



(a) Planner debt choice

(b) $\mathcal{P}\left(\cdot,\tilde{B}\right)$ (moderate $\tilde{B}$)

(c) $\mathcal{P}\left(\cdot,\tilde{B}\right)$ (tight $\tilde{B}$)

Figure 10: Primal Planner Solution

Note: This figure reports a numerical example of the planner's solution. Figures are calculated using value function iteration: $\beta = 0.95$, $\delta = 0.9$, and $\theta = 0.75$. Panel (a) plots B^p as a function of values of $\tilde{B}$ in the range $[0.1, \theta + 0.1] * B^{star}$. Panel's (b) and (c) plot the objective of the planner for different values of B in the range $[0, 1.8]$. Panel (b) fixes $\tilde{B}$ at 0.15 and Panel (c) at 0.08.

Competitive Equilibrium vs. Efficient Transitions: An illustration. To illustrate how transitions during a payments-chain crisis are inefficient, Figure 9 overlays the planner solution to the competitive equilibrium described above. During the extreme phase of the crisis, the planner redistributes wealth toward savers to induce more spot expenditures. To do so, the planner must subsidize savings. From the outset, the policy seems draconian: the planner taxes the agent suffering most, the worker, to subsidize savers. By making savers wealthier, the planner induces more spot expenditures. The increase in saver wealth increases TFP and, actually, leads to an increase in worker consumption

despite that they are taxed more heavily. The increase in worker consumption comes about through the reduction of the price of chained consumption, which more than offsets the increment in labor taxes. A regressive policy from the outset, increases the welfare of both agents.

During the recovery phase, as $\tilde{B}$ increases and enters the extreme valued interval, the planner reverses the policy. In this region, the planner taxes savings to reduce debt. In doing so, the planner frees some spot-borrowing capacity, and the worker can then make some spot expenditures. As the credit standards are further relaxed, the planner increases debts slowly reaching the efficient steady state.

Though we considered the planner solution along a transition toward an undisrupted steady state, we also know that the planner solution described in Proposition 8 would not allow hysteresis in payments-chain crises.[35] A takeaway is that exiting the hysteresis region, may require policies that seem draconian from the outset.

5.2 Fiscal Policy and the Bocola Effect

Fiscal Policy: when the government pays matters. We now consider government expenditures. We distinguish between spot government expenditures and expenditures chained to future tax receipts. It turns out that the type of government expenditures matters. We call this effect the Bocola effect, because economist Luigi Bocola suggested this distinction.

To formulate the Bocola effect, we treat government expenditures as isomorphic to household expenditures: we label by G^s the spot government expenditures. We assume that the government must also borrow intra-period. In turn, the government can make chained expenditures, G^x. For that, we treat government taxes as income units isomorphic to the labor income of households. For simplicity, we assume that the government raises labor taxes and satisfies a balanced budget at the end of the period. Furthermore, the resources used by both forms of expenditures are wasted. We consider the following problem.

[35] Indeed, in the hysteresis region, $Q_t/q_t = 1$, so $B_t = B_{t+1}$ is the only solution to the debt accumulation in the competitive equilibrium, (16). In the planner's solution, that equation is altered by the term, $\left(1 - \epsilon^q_{\mu,t}\right) < 1$.

Problem 6. (Government Problem with Expenditures): *Given* $B_0 = B_{ss}$ *and* $\left\{\tilde{B}_t\right\}$*:*

$$\mathcal{P}_0^g = \max_{\left\{\tau_t^k, G_t^x, G_t^s\right\}_{t\geq 0}} \sum_{t\geq 0} \beta^t \left[(1-\theta)\log\left(C_t^s\right) + \theta \log\left(C_t^w\right)\right],$$

subject to the saver's optimal behavior (10), a worker budget constraint:

$$B_t + E_t^w = R_{t+1}^{-1} B_{t+1} + 1 - \tau_t^\ell,\ \forall t \geq 0$$

the worker's optimal behavior (14), (11-12), and (3), a government budget-balance constraint with expenditures:

$$\tau_t^\ell = G_t^x + G_t^s,\ \forall t \geq 0,$$

and the ratio of chained expenditures relative to total expenditures:

$$\mu_t = q_t X_t^w + G_t^x \text{ and } q_t = \mathcal{A}\left(\mu_t\right)^{-1}.$$

The problem is similar to the Ramsey Problem but this new problem includes government expenditures and excludes credit taxes. Instead of solving the problem, we compute government multipliers near no expenditures. These multipliers transparently showcase how the payment time of government expenditures matters:

Proposition 9. (Infinitesimal Government Multiplies): *Fix* $\{G^x, G^s\} = (0,0)$*. Consider an unexpected marginal increase in government expenditures of type* $g \in \{x,s\}$ *at time* t*. We have that:*

$$\frac{\partial \mathcal{P}_t^g}{\partial G^g} = \underbrace{\frac{\theta}{C^w}}_{\text{marginal ind. utility}} \times \underbrace{\frac{\partial C^w}{\partial G^g}}_{\text{multiplier}} \text{ for } g \in \{x,s\}.$$

The consumption responses are:

$$\frac{\partial C^w}{\partial G^x} = \begin{cases} -1 & B < B^\star\left(\tilde{B}\right) \\ -\mathcal{A}\left(\mu\right) & B > B^\star\left(\tilde{B}\right) \end{cases}, \qquad \frac{\partial C^w}{\partial G^s} = \begin{cases} -1 & B < B^\star\left(\tilde{B}\right) \\ -\mathcal{A}\left(\mu\right)\left(1+\epsilon_\mu^{\mathcal{A}}\right) & B > B^\star\left(\tilde{B}\right). \end{cases}$$

Finally, the output multipliers are:

$$\frac{\partial \mathcal{Y}}{\partial G^x} = \begin{cases} \mathcal{A}(\mu)\left(1+\epsilon_\mu^{\mathcal{A}}\right)-1 & B < B^\star\left(\tilde{B}\right) \\ \\ 0 & B > B^\star\left(\tilde{B}\right) \end{cases}, \quad \frac{\partial \mathcal{Y}}{\partial G^s} = \begin{cases} 0 & B < B^\star\left(\tilde{B}\right) \\ \\ 1-\mathcal{A}(\mu)\left(1+\epsilon_\mu^{\mathcal{A}}\right) & B > B^\star\left(\tilde{B}\right). \end{cases}$$

When $B \leq B^\star$, production is efficient. For such low levels of debt, Proposition 9 shows that any form of government expenditures is a waste. First, either form of government expenditure reduces worker expenditures one-for-one, leading to a reduction in welfare. In terms of output, if the expenditure is spot, the multiplier is zero as transfers spot expenditures from the private to the public sector. If in turn, government expenditures are chained, the income multiplier is negative because these government expenditures increase productive inefficiency, $\left(1-\mathcal{A}(\mu)\left(1+\epsilon_\mu^{\mathcal{A}}\right)\right) > 0$.

In a payments-chain crisis, when $B > B^\star$, the multipliers behave differently. Chained expenditures are again detrimental for welfare.[36] They carry a zero output multiplier because they transfer an inefficient source of expenditures from the private sector to the public sector. By contrast, if the government spends spot, it provokes a positive externality. This externality is captured by the elasticity $\epsilon_\mu^{\mathcal{A}}$. Spot government expenditures also crowd-out worker chained expenditures, but the income extracted from workers are spent upfront. Ultimately, this reduces the average chain length and increases TFP. In a deep crisis, spot government expenditures may even increase worker consumption. In particular, this occurs when $\epsilon_\mu^{\mathcal{A}} < -1$, a condition that may indeed occurs, as shown in the proof. In terms of the output multiplier of spot expenditures, it is always positive in a payments chain crisis. The multiplier is positive in deep recessions because the government substitutes inefficient private expenditures for efficient public expenditures. Welfare increases precisely when the multiplier is above one.

We learn a new lesson. In a payments-chain crisis, government expenditures can increase welfare but only when paid upfront during deep crises. The argument is different from the classic aggregate-demand stimulus arguments. In this setting, a government that subscribes to the idea that it can stimulate aggregate demand, simply by

[36] Government expenditures crowd-out worker chained expenditures one-for-one. Divided by $1/q$, this gives us the reduction in chained consumption of chained government expenditures.

spending without considering when it pays for these expenditures, would unwillingly reduce welfare and have no output effects. To produce positive effects, the government must spend by paying for things upfront.

Discussion: Fiscal Transfers, and Ricardian Equivalence. Because the labor is inelastic to labor taxes and saver expenditures are inelastic to capital taxation, the Ramsey planner is equivalent to one that can set transfers.[37] The exercise above therefore shows that Ricardian equivalence fails in this setting. For example, future transfers can have pervasive effects in the midst of a payments-chain crisis. The reason is that it may lead workers to spend more. If workers increase chained expenditures, this may reduce output.

6. Conclusion

There are many views regarding the nature of financial crises. A classic view is that during financial crises the economy suffers from coordination failures and demand externalities. Yet nothing makes financial crises particularly prone to coordination failures or demand externalities. In fact, the deepest and longest recessions—the Great Depression, Japan's "lost decades," and the Great Recession—were all triggered by disruptions in the financial system.

The contribution of the present paper is to interpret financial crises as arising from delays in payments. The core economic problem we highlight is that an inefficient distribution of funds causes payment delays that disrupt production. We show that crises of this form feature demand externalities and can lead to coordination failures. Although the policy recommendations resemble those in environments with price rigidity, we underscore the critical importance of accelerating transactions through appropriate transfers of funds.

In building the theory, we make two important simplifications. First, we assume that all transactions are bilateral and of identical size. Second, we assume that non–spot transactions require time to execute. Developing a general version of this model—with

[37] An equivalence holds as long as transfers cannot be immediately used for spot payments.

a richer sequence of transactions and heterogeneous delays—will be important for empirical applications. Nonetheless, we believe the lessons of this paper are robust to such extensions.

References

Alvarez, Fernando and Gadi Barlevy, "Mandatory disclosure and financial contagion," *Journal of Economic Theory*, 6 2021, *194*, 105237. 1

Barrot, Jean-Noel, "Trade Credit and Industry Dynamics: Evidence from Trucking Firms," *The Journal of Finance*, 2016, *71* (5), 1975–2016. 1

Biais, Bruno and Christian Gollier, "Trade Credit and Credit Rationing," *The Review of Financial Studies*, October 1997, *10* (4), 903–937. 3

Bigio, Saki and Jennifer La'O, "Financial Frictions in Production Networks," 2013. 1

Bloch, Francis and Matthew O Jackson, "Definitions of equilibrium in network formation games," *International Journal of Game Theory*, 2006, *34* (3), 305–318. 24

Bocola, Luigi, "Trade credit and financial amplification," 2022. Unpublished Manuscript. 1

Boissay, Frederic and Reint Gropp, "Trade Credit Defaults and Liquidity Provision by Firms," 2007. European Central Bank working paper series, 753. 1

Brunnermeier, Markus K. and Yuliy Sannikov, "The I-Theory of Money," 2017. 1

Cerdas, Ana María and Carlos Melegatti, *El sistema financiero costarricense en los últimos 25 años: sistema de pagos* number 2. In 'Monografía.', San José, Costa Rica: Academia de Centroamérica, 2014. 28 cm. 8

Costello, Anna Marie, "Credit Market Disruptions and Liquidity Spillover Effects in the Supply Chain," *Journal of Political Economy, forthcoming*, 2020. 1

Diamond, Peter A., "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 1982, *90*, 881–894. 1

Eggertsson, Gauti B. and Paul Krugman, "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach," *The Quarterly Journal of Economics*, July 2012, *127* (3), 1469–1513. _eprint: https://academic.oup.com/qje/article-pdf/127/3/1469/30457060/qjs023.pdf. 1

Elliott, Matthew, Benjamin Golub, and Matthew O. Jackson, "Financial Networks and Contagion," *American Economic Review*, 10 2014, *104*, 3115–53. 1

_ , _ , and Matthew V. Leduc, "Supply Network Formation and Fragility," *American Economic Review*, August 2022, *112* (8), 2701–47. 1

Fisher, Irving, *Stamp Scrip*, New York: Adelphi, 1933. 1

Freeman, Scott, "The Payments System, Liquidity, and Rediscounting," *The American Economic Review*, 1996, *86* (5), 1126–1138. Publisher: American Economic Association. 1

Green, Edward J., "Money and Debt in the Structure of Payments," *Federal Reserve Bank of Minneapolis Quaterly Review*, 1999, *23*, 13–29. 1

Guerrieri, Veronica and Guido Lorenzoni, "Liquidity and Trading Dynamics," *Econometrica*, 2009, *77* (6), 1751–1790. 1

_ and _ , "Credit Crises, Precautionary Savings, and the Liquidity Trap*," *The Quarterly Journal of Economics*, 2017, *132* (3), 1427–1467. 1

_ , _ , Ludwig Straub, and Ivan Werning, "Macroeconomic Implications of Covid-19: Can Negative Supply Shocks Cause Demand Shortages?," *American Economic Review*, 2022. 1

Hardy, Bryan, Ina Simonovska, and Felipe E. Saffie, "Economic Stabilizers in Emerging Markets:The Case for Trade Credit," 2022. 1

Jordà, Òscar, "Estimation and inference of impulse responses by local projections," *American economic review*, 2005, *95* (1), 161–182. 2, 1, 11

_ and Alan M Taylor, "Local projections," *Journal of Economic Literature*, 2025, *63* (1), 59–110. 2, 1, 11

Karabarbounis, Loukas, Gabriel Chodorow-Reich, and Rohan Kekre, "The Macroeconomics of the Greek Depression," 2021. 1

Kiyotaki, Nobuhiro and John Moore, "Credit Chains," 1997. Publisher: London School of Economics. 1

_ **and Randall Wright**, "On Money as a Medium of Exchange," *Journal of Political Economy*, 1989, *97* (4), 927–954. Publisher: The University of Chicago Press. 1

Kopytov, Alexandr, Bineet Mishra, Kristoffer Nimark, and Mathieu Taschereau-Dumouchel, "Endogenous Production Networks under Supply Chain Uncertainty," 2022. 1

Krugman, Paul, "Baby-Sitting the Economy," *Slate*, August 1998. 1

Lagos, Ricardo and Guillaume Rocheteau, "Liquidity in Asset Markets with Search Frictions," *Econometrica*, 2009, *77*, 403–426. 2

_ **and Randall Wright**, "Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 2005, *113* (3), 463–484. 1

_ **, Guillaume Rocheteau, and Pierre-Olivier Weill**, "Crises and Liquidity in OTC Markets," *Journal of Economic Theory*, 2011, *146*, 2169–2205. 2

La'O, Jennifer, "A Traffic Jam Theory of Recessions," 2015. 1

Li, Yiting, Guillaume Rocheteau, and Pierre-Olivier Weill, "Liquidity and the Threat of Fraudulent Assets," *Journal of Political Economy*, 2012, *120*, 815–846. 2

Lippi, Francesco, Stefania Ragni, and Nicholas Trachter, "Optimal monetary policy with heterogeneous money holdings," *Journal of Economic Theory*, 2015, *159*, 339 – 368. 1

Lucas, Robert E. and Nancy L. Stokey, "Money and Interest in a Cash-in-Advance Economy," *Econometrica*, 1987, *55* (3), 491–513. Publisher: The Econometric Society. 1, 4.1

Mendoza, Enrique G., "Sudden Stops, Financial Crises, and Leverage," *American Economic Review*, December 2010, *100* (5), 1941–66. 1

Meza, Felipe and Erwan Quintin, "Financial Crises: Still a Mystery," Technical Report Federal Reserve Bank of Dallas, Southwest Economics, Semptember/October 2005. 1

_ **and** _ , "Factor Utilization and the Real Impact of Financial Crises," *The B.E. Journal of Macroeconomics*, 2007, *7* (1). 1, 1

Nosal, Ed and Guillaume Rocheteau, *Money, Payment, and Liquidity*, MIT Press, 2011. 2

Oberfield, Ezra, "Productivity and misallocation during a crisis: Evidence from the Chilean crisis of 1982," *Review of Economic Dynamics*, 2013, *16* (1), 100–119. Special issue: Misallocation and Productivity. 1, 1

_ , "A Theory of Input-Output Architecture," *Econometrica*, 3 2018, *86* (2), 559–589. 1

Pratap, Sangeeta and Carlos Urrutia, "Financial frictions and total factor productivity: Accounting for the real effects of financial crises," *Review of Economic Dynamics*, 2012, *15* (3), 336–358. 1

Rocheteau, Guillaume, "Payment and Liquidity under Adverse Selection," *Journal of Monetary Economics*, 2011, *58* (191-205). 2

_ **, Tsz-Nga Wong, and Pierre-Olivier Weill**, "Working through the Distribution: Money in the Short and Long Run," May 2016. 1

Shi, Shouyong, "A Divisible Search Model of Fiat Money," *Econometrica*, 1997, *65*, 75–102. 2

Summers, Lawrence H., "Some Skeptical Observations on Real Business Cycle Theory," *Federal Reserve Bank of Minneapolis Quarterly Review*, 1986. 1

Taschereau-Dumouchel, Mathieu, "Cascades and Fluctuations in an Economy with an Endogenous Production Network," 2022. 1

Townsend, Robert M., "Models of Monetary Economies," 1980, pp. 265–303. 1

von Schedvin, Erik Jacobsen Tor;, "Trade Credit and the Propagation of Corporate Failure: An Empirical Analysis," *Econometrica*, July 2015, *83* (4), 1315–1371. 1

Woodford, Michael, "Effective Demand Failures and the Limits of Monetary Stabilization Policy," *forthcoming American Economic Review*, 2022. 1

Appendix (Not intended for publication)

Table of Contents for Appendix

1 Introduction **1**

2 Empirical Motivation **7**

3 Payments-Chains and Productivity **17**

3.1 Extensions 25

4 Payments-Chains in a Business-Cycle Model **29**

4.1 Environment 29

4.2 Characterization 33

5 Policy Implications of Payments-Chain Crises **44**

5.1 Constrained Inefficiency 44

5.2 Fiscal Policy and the Bocola Effect 52

6 Conclusion **55**

A Proofs of Section 3 **1**

A.1 Proof of Proposition 2 1

A.2 Related Results Used Elsewhere 7

B Model Extensions **10**

B.1 Markov Chain Representation 10

B.2 Truncation 15

B.3 Multinomial Case 18

B.4 Resampling 21

C Proofs of Section 4 **26**

C.1 Preliminary Observations 26

D Empirical Appendix **32**

D.1 Local Projection: 2005 Daily Data 32

D.2 Intensive Margin Regressions 32

D.3 Placebo Exercises 34

D.4 Distribution of Chain Sizes 34

D.5 Chained Payments and Prices 35

A. Proofs of Section 3

A.1 Proof of Proposition 2

Part 1. Derivation of TFP. In the body of the paper, we showed that for a chain of three orders, of which two are chained, the vector of production is $\{1, \delta, \delta^2\}$. The following induction argument generalizes: If the k-th node initiates production at time $1-\delta^{k-1}$, the delay from its inspection is $(1-\delta)\,\delta^{k-1}$, which added to previous delays leads to a transfer of funds only by time $1-\delta^{k-1}+(1-\delta)\,\delta^{k-1} = 1-\delta^k$. This, leaves δ^k time for production to the subsequent unit. Since we computed the delay for $k = 1, 2, \ldots$. the productions in a chain of length n are $\{1, \delta, \delta^2, \ldots, \delta^n\}$.

It follows that the average output for chained orders, that is excluding the output of the spot order in the chain, in a payments-chain of length n is

$$\bar{y}_n^x = \frac{1}{n}\sum_{m=1}^{n}\delta^m = \frac{\delta}{n}\left(\frac{1-\delta^n}{1-\delta}\right).$$

Proof. Recall that a chained order will necessarily fall in a chain with length $n \geq 1$. Thus, the p.m.f of lengths conditional on this event is

$$G^x(n;\mu) = \frac{(1-\mu)\,\mu^n}{\mu}.$$

Next, we turn to our goal of finding the expected output of a chained order:

$$\begin{aligned}
\mathbb{E}[\bar{y}^x] &= \sum_{n=1}^{\infty}\bar{y}_n^x G^x(n;\mu), \\
&= \sum_{n=1}^{\infty}\frac{(1-\mu)\,\mu^n}{\mu}\cdot\frac{\delta}{n}\left(\frac{1-\delta^n}{1-\delta}\right), \\
&= \frac{(1-\mu)}{\mu}\cdot\frac{\delta}{(1-\delta)}\cdot\sum_{n=1}^{\infty}\left(\frac{\mu^n}{n}-\frac{(\delta\mu)^n}{n}\right), \\
&= \frac{1-\mu}{\mu}\cdot\frac{\delta}{1-\delta}\cdot\ln\left(\frac{1-\delta\mu}{1-\mu}\right),
\end{aligned}$$

The last equality follows from:

$$\sum_{n=1}^{\infty} a^{n-1} = \frac{1}{1-a} \leftrightarrow$$
$$\sum_{n=1}^{\infty} \frac{a^n}{n} = \ln\left(\frac{1}{1-a}\right)$$

for $|a| < 1$, which can be shown by simply taking derivatives to both sides. A simple weak law of large numbers yields the desired result. Recall that we can have spot identities $i \in \mathcal{N}^s$ as unique identifiers for payment chains (spot orders for a one-to-one map with chains). Let $\ell(i)$ be the length of the payment chain that starts with spot order i. Given that for each $i \in \mathcal{N}^s$, $\ell(i) \sim i.i.d.\, G(\mu)$, then

$$\mathcal{A}(\mu;\delta) \equiv \operatorname*{plim}_{N\to\infty} \frac{1}{N^s} \sum_{i\in\mathcal{N}^s} \overline{y}^x_{\ell(i)} = \mathbb{E}\left[\overline{y}^x\right].$$

In words, the average output among chained orders converges to $\mathbb{E}\left[\overline{y}^x\right]$ as the network gets larger ($N \to \infty$). □

Next, we derive expected output—the limit as $N \to \infty$ is implicitly. The fraction of spot orders is $(1-\mu)$. Production is 1 in their case. The fraction of chained orders is μ, and they produce on average $\mathcal{A}(\mu)$. Thus, total output is:

$$\mathcal{Y}(\mu) = (1-\mu) + \mu\mathcal{A}(\mu).$$

Next, we obtain the derivative and limits of $\mathcal{Y}(\mu), \mathcal{A}(\mu)$.

Part 2. Limits. We first consider the limit as $\mu \to 0$:

$$\lim_{\mu\to 0} \mathcal{A}(\mu;\delta) = \frac{\delta}{(1-\delta)} \lim_{\mu\to 0} \left(\frac{1}{\mu} - 1\right) \cdot \ln\left(\frac{1-\delta\mu}{1-\mu}\right) = \lim_{\mu\to 0} \frac{\ln\left(\frac{1-\delta\mu}{1-\mu}\right)}{\mu}.$$

The last term is the ratio of two variables that converge to zero. Using L'Hospital's rule:

$$\lim_{\mu\to 0} \frac{\ln\left(\frac{1-\delta\mu}{1-\mu}\right)}{\mu} = \frac{\delta}{(1-\delta)} \frac{\lim_{\mu\to 0}\left(\frac{1}{1-\mu} - \frac{\delta}{1-\delta\mu}\right)}{1} = \delta.$$

where we used

$$\frac{\partial \ln\left(\frac{1-\delta\mu}{1-\mu}\right)}{\partial\mu} = \frac{1-\mu}{1-\delta\mu}\left(\frac{1-\delta\mu}{1-\mu}\right)\left(\frac{1}{1-\mu}-\frac{\delta}{1-\delta\mu}\right) = \left(\frac{1}{1-\mu}-\frac{\delta}{1-\delta\mu}\right).$$

For output, the limit is:

$$\lim_{\mu\to 0}\mathcal{Y}(\mu;\delta) = \lim_{\mu\to 0}(1-\mu)\lim_{\mu\to 0}\left(1+\frac{\delta}{1-\delta}\ln\left(\frac{1-\delta\mu}{1-\mu}\right)\right) = 1.$$

Next, we consider the limit as $\mu \to 1$:

$$\lim_{\mu\to 1}\mathcal{A}(\mu;\delta) = \frac{\delta}{(1-\delta)}\lim_{\mu\to 1}\left(\frac{1}{\mu}-1\right)\lim_{\mu\to 1}\ln\left(\frac{1-\delta\mu}{1-\mu}\right).$$

This is the product of numbers that go to 0 and infinity. Using L'Hospital's rule:

$$\lim_{\mu\to 1}\mathcal{A}(\mu;\delta) = \frac{\lim_{\mu\to 1}\left(-\frac{1}{\mu^2}\right)}{\lim_{\mu\to 1}\left(\frac{1}{1-\mu}-\frac{\delta}{1-\delta\mu}\right)} = 0.$$

For output, the limit is:

$$\lim_{\mu\to 1}\mathcal{Y}(\mu) = \lim_{\mu\to 1}(1-\mu) + \lim_{\mu\to 1}\mu\lim_{\mu\to 1}\mathcal{Y}^{x}(\mu) = 0.$$

Next, we consider the limit as limit as $\delta \to 0$:

$$\lim_{\delta\to 0}\mathcal{A}(\mu;\delta) = \left(\frac{1}{\mu}-1\right)\lim_{\delta\to 0}\frac{\delta}{(1-\delta)}\cdot\lim_{\delta\to 0}\ln\left(\frac{1-\delta\mu}{1-\mu}\right) = 0.$$

For output,

$$\lim_{\delta\to 0}\mathcal{Y}(\mu) = (1-\mu) + \mu\lim_{\delta\to 0}\mathcal{Y}^{x}\mathcal{A}(\mu;\delta) = (1-\mu).$$

Finally, we consider the limit as $\delta \to 1$:

$$\lim_{\delta\to 1}\mathcal{A}(\mu;\delta) = \left(\frac{1}{\mu}-1\right)\lim_{\delta\to 1}\delta\cdot\lim_{\delta\to 1}\frac{1}{(1-\delta)}\cdot\lim_{\delta\to 1}\ln\left(\frac{1-\delta\mu}{1-\mu}\right).$$

This derivative is of the ratio of two numbers that go to zero. Using L'Hospital's rule:

$$\lim_{\delta\to 1}\mathcal{A}(\mu;\delta)=\left(\frac{1}{\mu}-1\right)\frac{\lim_{\delta\to 1}\frac{1-\mu}{1-\delta\mu}\left(\frac{-\mu}{1-\mu}\right)}{-1}=\left(\frac{1-\mu}{\mu}\right)\left(\frac{\mu}{1-\mu}\right)=1.$$

For output,

$$\lim_{\delta\to 1}\mathcal{Y}(\mu)=(1-\mu)+\mu\lim_{\delta\to 1}\mathcal{Y}^{x}\mathcal{A}(\mu;\delta)=1.$$

This concludes the derivation of the limits of interest.

Part 3. Monotonicity. Next, we investigate the derivatives of $\mathcal{A}$ and $\mathcal{Y}$. We can write:

$$\mathcal{A}(\mu;\delta)=\left(\frac{1}{\mu}-1\right)\cdot\frac{\delta}{(1-\delta)}\cdot\ln\left(\frac{1-\delta\mu}{1-\mu}\right).$$

Thus,

$$\mathcal{A}_{\mu} = \frac{\delta}{(1-\delta)}\left(\left(-\frac{1}{\mu^{2}}\right)\cdot\ln\left(\frac{1-\delta\mu}{1-\mu}\right)+\left(\frac{1}{\mu}-1\right)\left(\frac{-\delta}{1-\delta\mu}+\frac{1}{1-\mu}\right)\right).$$

Factoring out $-1/\mu^{2}$:

$$\begin{aligned}\mathcal{A}_{\mu} &= -\frac{\delta}{(1-\delta)}\frac{1}{\mu^{2}}\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right)-\mu(1-\mu)\left(\frac{-\delta}{1-\delta\mu}+\frac{1}{1-\mu}\right)\right)\\ &= -\frac{\delta}{(1-\delta)}\frac{1}{\mu^{2}}\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right)-\mu\left(\frac{1-\delta}{1-\delta\mu}\right)\right).\end{aligned}$$

To show that the derivative is negative, we need to show that the term in the parenthesis is positive. Or likewise that

$$\ln(1-\delta\mu)-\left(\frac{\mu-\delta\mu}{1-\delta\mu}\right)>\ln(1-\mu).$$

A concave function $f(x)\equiv\log(1-x)$ satisfies:

$$f(x)+f'(x)|y-x|>f(y).$$

Let $x = \delta\mu$ and $y = \mu$. Because $\{\delta, \mu\} < 1$,

$$|y - x| = \mu - \delta\mu.$$

Thus, we have:

$$\ln(1-\delta\mu) + \underbrace{\left(-\frac{1}{1-\delta\mu}\right)(\mu - \delta\mu)}_{f'(x)|y-x|} > \ln(1-\mu).$$

which proves the desired inequality. Hence, $\mathcal{A}_\mu < 0$ for any $\mu > 0$. At $\mu = 0$, the derivative is zero.

We also obtain that:

$$\mathcal{Y}_\mu = -1 + \mu\mathcal{A}_\mu + \mathcal{A} = -\underbrace{(1-\mathcal{A})}_{>0} - \frac{\delta}{(1-\delta)}\frac{1}{\mu}\left(\underbrace{\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right) - \mu\left(\frac{1-\delta}{1-\delta\mu}\right)\right)}_{>0}\right) \text{ for } \mu > 0.$$

The derivative is also zero at $\mu = 0$.

Part 4. Concavity. Next we perform the convexity analysis. $\mathcal{A}_{\mu\mu}$ is

$$\frac{\delta}{(1-\delta)}\left[2\frac{1}{\mu^3}\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right) - \mu\left(\frac{1-\delta}{1-\delta\mu}\right)\right) - \frac{1}{\mu^2}\left(\frac{1}{1-\mu} - \frac{\delta}{1-\delta\mu} - \frac{1-\delta}{1-\delta\mu} - \mu\delta\frac{1-\delta}{(1-\delta\mu)^2}\right)\right]$$

After some algebraic manipulations, the second term in parenthesis simplifies to:

$$\frac{1}{1-\mu} - \frac{\delta}{1-\delta\mu} - \frac{1-\delta}{1-\delta\mu} - \mu\delta\frac{1-\delta}{(1-\delta\mu)^2} = \mu\frac{(\delta-1)^2}{(1-\mu)(1-\delta\mu)^2}.$$

Thus:

$$\mathcal{A}_{\mu\mu} = \frac{\delta}{(1-\delta)}\frac{1}{\mu^3}\left[2\ln\left(\frac{1-\delta\mu}{1-\mu}\right) - 2\mu\left(\frac{1-\delta}{1-\delta\mu}\right) - \frac{\mu^2(1-\delta)^2}{(1-\mu)(1-\delta\mu)^2}\right].$$

We can add the second and third terms to obtain:

$$\mathcal{A}_{\mu\mu} = \frac{\delta}{(1-\delta)}\frac{1}{\mu^3}\left[\ln\left(\frac{1-\delta\mu}{1-\mu}\right)^2 - \mu\left(\frac{1-\delta}{1-\delta\mu}\right)\left(\frac{2-\mu-3\delta\mu+2\delta\mu^2}{(1-\mu)(1-\delta\mu)}\right)\right].$$

This function is strictly concave if:

$$\ln\left(1-\delta\mu\right)^2 < \ln\left(1-\mu\right)^2 + \frac{1}{\left(1-\mu\right)^2}\left(\frac{\mu\left(1-\mu\right)\left(1-\delta\right)}{\left(1-\delta\mu\right)^2}\left(2-\mu+2\delta\mu^2-3\delta\mu\right)\right) \tag{21}$$

Let $F\left(x\right)=\ln\left(x\right)$. Set $x_0=\left(1-\mu\right)^2$ and $x_1=\left(1-\delta\mu\right)^2$ so that

$$x_0-x_1=-\mu\left(1-\delta\right)\left(2-\mu\left(1+\delta\right)\right).$$

By strict concavity of $\ln\left(x\right)$ we have

$$\ln\left(1-\delta\mu\right)^2 < \ln\left(1-\mu\right)^2 - \frac{1}{\left(1-\mu\right)^2}\mu\left(1-\delta\right)\left(2-\mu\left(1+\delta\right)\right)$$

so to prove that $\mathcal{A}_{\mu\mu}$ is strictly negative, we need to prove that the right hand side of the expression above is smaller than the condition needed for concavity, condition (21),

$$-\frac{1}{\left(1-\mu\right)^2}\mu\left(1-\delta\right)\left(2-\mu\left(1+\delta\right)\right) \leq \frac{1}{\left(1-\mu\right)^2}\left(\frac{\mu\left(1-\mu\right)\left(1-\delta\right)}{\left(1-\delta\mu\right)^2}\left(2-\mu+2\delta\mu^2-3\delta\mu\right)\right).$$

Cancelling common terms and rearranging, this condition is equivalent to:

$$-\left(1-\delta\mu\right)^2\left(2-\mu-\mu\delta\right) \leq \left(1-\mu\right)\left(2-\mu+2\delta\mu^2-3\delta\mu\right). \tag{22}$$

The term on the left is negative—and strictly negative for $\delta,\mu<1$.[38] Hence, the inequality above is verified as long as:

$$2 \geq \alpha\left(\mu,\delta\right) \equiv \mu - 2\delta\mu^2 + 3\delta\mu.$$

Hence, as long as

$$2 \geq \alpha^* = \max_{\{\mu,\delta\}\in[0,1]^2} \alpha\left(\mu,\delta\right)$$

the condition for concavity holds for all $\{\mu,\delta\}\in[0,1]^2$. We study this max function. Fix any μ. Since

$$\delta\left(3\mu-2\mu^2\right) \geq 0,$$

[38] This follows immediately because μ and δ are fractions.

the maximal value of α is achieved when $\delta = 1$. Hence, the objective is

$$\alpha^* = \max_{\{\mu,\delta\}\in[0,1]^2} \alpha(\mu,\delta) = \max_{\{\mu\}\in[0,1]} \alpha(\mu,1) = \max_{\{\mu\}\in[0,1]} 4\mu - 2\mu^2.$$

Maximizing the last expression over μ yields $\mu = 1$ as a solution and $\alpha^* = 2$ as the value. Hence, the inequality holds and guarantees (22). This suffices to prove concavity.

Next, we verify the concavity of total output. We have that $\mathcal{Y}_{\mu\mu}$ is

$$\begin{aligned}
=&\mathcal{A}_\mu + \mu\mathcal{A}_{\mu\mu} \\
=&-\frac{\delta}{(1-\delta)}\frac{1}{\mu^2}\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right) - \mu\left(\frac{1-\delta}{1-\delta\mu}\right)\right) + \frac{\delta}{(1-\delta)}\frac{2}{\mu^2}\mu\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right) - \mu\left(\frac{1-\delta}{1-\delta\mu}\right)\right) \\
&-\frac{\mu}{\mu^2}\left(\frac{1}{1-\mu} - \frac{\delta}{1-\delta\mu} - \frac{1-\delta}{1-\delta\mu} - \mu\delta\frac{1-\delta}{(1-\delta\mu)^2}\right) \\
=&\mathcal{A}_{\mu\mu} - \frac{\delta}{(1-\delta)}\frac{\mu}{\mu^3}\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right) - \mu\left(\frac{1-\delta}{1-\delta\mu}\right)\right) \\
<&0.
\end{aligned}$$

A.2 Related Results Used Elsewhere

In this section, we derived properties that are used later in the text.

Part 5. Inverse productivity. Now, we study the inverse of productivity. Let

$$q(\mu;\delta) = \mathcal{A}^{-1}(\mu;\delta).$$

Clearly, the function has the limits:

$$\lim_{\mu\to 0} q(\mu;\delta) = \delta^{-1} \text{ and } \lim_{\mu\to 1} q(\mu;\delta) = \infty \text{ and } \lim_{\delta\to 0} q(\mu;\delta) = \infty \text{ and } \lim_{\delta\to 1} q(\mu;\delta) = 1.$$

We also have that:

$$q_\mu = -\frac{\mathcal{A}_\mu}{\mathcal{A}^2} > 0.$$

We use the limit of the derivative of this function:

$$q_{\mu}(\mu) = \frac{\frac{\delta}{(1-\delta)}\frac{1}{\mu^2}\left(\ln\left(\frac{1-\delta\mu}{1-\mu}\right) - \mu\left(\frac{1-\delta}{1-\delta\mu}\right)\right)}{\left(\frac{(1-\mu)}{\mu}\cdot\frac{\delta}{(1-\delta)}\cdot\ln\left(\frac{1-\delta\mu}{1-\mu}\right)\right)^2}$$

Next, we check the convexity of function:

$$q_{\mu\mu} = -\frac{\mathcal{A}_{\mu\mu}}{\mathcal{A}^2} + \frac{\mathcal{A}_{\mu}}{\mathcal{A}^3} > 0.$$

Hence, q is convex in μ.

Part 6. Elasticity of $\mathcal{A}$. A useful object in later derivations is the elasticity of $\mathcal{A}$. Consider the derivative of

$$\mathcal{A}(\mu)\,\mu.$$

We have:

$$\mathcal{A}'(\mu)\,\mu + \mathcal{A}(\mu) = \mathcal{A}(\mu)\left[1+\epsilon_{\mu}^{\mathcal{A}}\right].$$

Recall that,

$$\mathcal{A}(\mu)\,\mu = (1-\mu)\left(\frac{\delta}{1-\delta}\ln\left(\frac{1-\delta\mu}{1-\mu}\right)\right).$$

Hence,

$$\begin{aligned}\mathcal{A}'(\mu)\,\mu + \mathcal{A}(\mu) &= \frac{\delta}{1-\delta}\left(-\ln\left(\frac{1-\delta\mu}{1-\mu}\right) + (1-\mu)\left(\frac{-\delta}{1-\delta\mu}+\frac{1}{1-\mu}\right)\right)\\ &= \frac{\delta}{1-\delta}\left(-\ln\left(\frac{1-\delta\mu}{1-\mu}\right) + \left(\frac{1-\delta}{1-\delta\mu}\right)\right).\end{aligned}$$

Hence, we obtain that:

$$\begin{aligned}\left[1+\epsilon_{\mu}^{\mathcal{A}}\right] &= \frac{\frac{\delta}{1-\delta}\left(-\ln\left(\frac{1-\delta\mu}{1-\mu}\right) + \left(\frac{1-\delta}{1-\delta\mu}\right)\right)}{\mathcal{A}(\mu)}\\ &= \frac{\mu}{1-\mu}\left(\frac{\frac{1-\delta}{1-\delta\mu}}{\ln\left(\frac{1-\delta\mu}{1-\mu}\right)} - 1\right).\end{aligned}$$

We are interested in the sign of $1+\epsilon_{\mu}^{\mathcal{A}}$ and its limits. We know $\epsilon_{\mu}^{\mathcal{A}} < 0$. Thus, the sign

of $1+\epsilon_\mu^{\mathcal{A}}$ is the sign of:

$$\left(\frac{1-\delta}{1-\delta\mu}\right)-\ln\left(\frac{1-\delta\mu}{1-\mu}\right).$$

The limits of the function that governs the sign are:

$$\lim_{\mu\to 0}\left(\frac{1-\delta}{1-\delta\mu}\right)-\ln\left(\frac{1-\delta\mu}{1-\mu}\right)=(1-\delta)>0,$$

and

$$\lim_{\mu\to 1}=\left(\frac{1-\delta}{1-\delta\mu}\right)-\ln\left(\frac{1-\delta\mu}{1-\mu}\right)=-\infty.$$

Since the function is continuous in μ, the sign is ambiguous.

Next, to establish monotonicity, notice that

$$\begin{aligned}\delta\frac{1-\delta}{(1-\delta\mu)^2}-\frac{1-\delta}{(1-\delta\mu)(1-\mu)} &= (1-\delta)\left(\frac{-\delta-1+\delta(1-\delta\mu)}{(1-\delta\mu)^2(1-\mu)}\right)\\ &= -(1-\delta)\left(\frac{1+\delta^2\mu}{(1-\delta\mu)^2(1-\mu)}\right)\\ &< 0.\end{aligned}$$

Hence, there's a unique crossing point where the function $1+\epsilon_\mu^{\mathcal{A}}$ is negative.

Finally, we compute relevant limits. Firs, we compute:

$$\lim_{\mu\to 0}1+\epsilon_\mu^{\mathcal{A}}=(1-\delta)\lim_{\mu\to 0}\frac{\mu}{\ln\left(\frac{1-\delta\mu}{1-\mu}\right)}=\frac{1-\delta}{\lim_{\mu\to 0}\frac{1-\mu}{1-\delta\mu}\frac{1-\delta\mu}{1-\mu}\left(\frac{-\delta}{1-\delta\mu}-\frac{-1}{1-\mu}\right)}=1.$$

where the first equality are the surviving terms after taking limits, the second equality follows from L'Hospital's rule. For the limit $\mu\to 1$, by L'Hospital's rule:

$$\lim_{\mu\to 1}\frac{1}{1-\mu}\left(\frac{1}{\ln\left(\frac{1-\delta\mu}{1-\mu}\right)}\right)=\lim_{\mu\to 1}\frac{-\frac{1}{1-\mu^2}}{\frac{1-\delta}{(1-\delta\mu)(1-\mu)}}=\lim_{\mu\to 1}-\frac{(1-\delta\mu)}{(1-\mu)(1-\delta)}.$$

Adding terms the missing term:

$$\begin{aligned}\lim_{\mu\to 1}\left[1+\epsilon_{\mu}^{\mathcal{A}}\right] &= \lim_{\mu\to 1}-\frac{(1-\delta\mu)}{(1-\mu)(1-\delta)}-\frac{\mu}{1-\mu}\\ &= \lim_{\mu\to 1}\frac{-1+2\delta\mu-\mu}{(1-\mu)(1-\delta)}\\ &= -2\lim_{\mu\to 1}\frac{1}{(1-\mu)}\\ &= -\infty.\end{aligned}$$

We finally compute the elasticity. From

$$\frac{\mathcal{A}'(\mu)}{\mathcal{A}(\mu)}\mu<0,$$

we know that the elasticity $\epsilon_{\mu}^{\mathcal{A}}$ is decreasing. Hence, $1+\epsilon_{\mu}^{\mathcal{A}}$ starts at zero and falls continuously until diverging at $\mu\to 1$.

B. Model Extensions

B.1 Markov Chain Representation

- μ_x fraction of chained
- $\mu_s=1-\mu_x$ fraction of spot
- the column vector is μ with the first entry μ

Payments chain have a Markov chain representation:

- It behaves as a history of states
- $S^t=\{s,x,s,x,s\}$

Thus, must be possible to write it as a Markov Chain with transition matrix

$$Q=\begin{bmatrix} p_{ss} & 1-p_{xx}\\ 1-p_{ss} & p_{xx}\end{bmatrix}.$$

Since Q is a probability matrix, all the information is contained in one row. Then, we must have that populations are maintained:

$$\mu = Q\mu.$$

We have special restrictions.

- iid case, considered in current version:

$$p_{ss} = \mu_s, p_{xx} = \mu_x.$$

Thus:

$$Q^{iid} = \begin{bmatrix} \mu_s & \mu_s \\ \mu_x & \mu_x \end{bmatrix}$$

The, clearly:

$$\mu = Q^{iid}\mu$$

since:

$$\begin{bmatrix} \mu_s \\ \mu_x \end{bmatrix} = \begin{bmatrix} \mu_s & \mu_s \\ \mu_x & \mu_x \end{bmatrix} \begin{bmatrix} \mu_s \\ \mu_x \end{bmatrix} = \begin{bmatrix} \mu_s (\mu_s + \mu_x) \\ \mu_x (\mu_s + \mu_x) \end{bmatrix}.$$

- Deviations from the iid case:

$$p_{ss} = \mu_s\kappa_s, p_{xx} = \mu_x\kappa_x.$$

Thus:

$$Q = \begin{bmatrix} \mu_s\kappa_s & 1 - \mu_x\kappa_x \\ 1 - \mu_s\kappa_s & \mu_x\kappa_x \end{bmatrix}$$

Which we conveniently write:

$$Q = \begin{bmatrix} \mu_s + \mu_s(\kappa_s - 1) & 1 - \mu_x - \mu_x(\kappa_x - 1) \\ 1 - \mu_s - \mu_s(\kappa_s - 1) & \mu_s + \mu_x(\kappa_x - 1) \end{bmatrix} = \left(Q^{iid} + \tilde{Q}\right)$$

where

$$\tilde{Q} = \begin{bmatrix} \mu_s(\kappa_s - 1) & -\mu_x(\kappa_x - 1) \\ -\mu_s(\kappa_s - 1) & \mu_x(\kappa_x - 1) \end{bmatrix}$$

- Next, we have that:

$$\mu = Q\mu = \left(Q^{iid} + \tilde{Q}\right)\mu$$

since

$$\mu = Q^{iid}\mu$$

then:

$$0 = \tilde{Q}\mu$$

$$\mu_s^2(\kappa_s - 1) = \mu_x^2(\kappa_x - 1)$$

Hence:

$$\mu_s^2(\kappa_s - 1) = \mu_x^2(\kappa_x - 1).$$

Thus, if

$$\mu_s(\kappa_s - 1) = \frac{\mu_x}{\mu_s}\mu_x(\kappa_x - 1).$$

Define:

$$q = \mu_s(\kappa_s - 1).$$

Then:

$$Q = \begin{bmatrix} \mu_s + q & 1 - \mu_x - \frac{\mu_s}{\mu_x}q \\ 1 - \mu_s - q & \mu_x + \frac{\mu_s}{\mu_x}q \end{bmatrix},$$

or more convenient to do change of variables:

$$\frac{\mu_s}{\mu_x} q = \rho.$$

- Bounded probabilities:
  - if $\mu_s > 1/2 \rightarrow q \leq 1 - \mu_s, q \geq -\mu_s$
  - if $\mu_s > 1/2 \rightarrow q \leq 1 - \mu_x, q \geq -\mu_x$
  - coupling: $q \leq 1 - \max\{\mu_x, \mu_s\}, q \geq -\max\{\mu_x, \mu_s\}$.
- This allows for a correlated structure.
- Moments:
  - fraction of spots in a chain:
  $$1 - \mu_s - q$$
  - fraction of spots alone:
  $$\mu_s + q$$
  - distribution of chain length, conditional on chain of size at least one:
  $$F(n) = \left(1 - \mu_x - \frac{\mu_x}{\mu_s} q\right) \left(\mu_x + \frac{\mu_x}{\mu_s} q\right)^{n-1}$$
  - Then:
  $$F(n) = (1 - (\mu_x + \rho)) (\mu_x + \rho)^{n-1}.$$
  That, is, chain distribution is Geometric with probability $p = \mu_x + \rho$.
- Output of chained orders in chain of length n

$$y^x(n) = \delta\left(1 + \delta + \delta^2 ... \delta^{n-1}\right) = \delta\left(\frac{1}{1-\delta} - \frac{\delta^n}{1-\delta}\right),$$

for $n \geq 1$. Clearly,

$$y^x(1) = \delta$$

- Average output in chain of length n is:

$$\bar{y}^x(n) = \frac{y(n)}{n}.$$

- Integrating the first term with respect to the Geometric distribution:

$$\mathbb{E}\left[\bar{y}^x(n) \,|n \geq 1\right] = \frac{1}{p}\frac{\delta}{1-\delta}\sum_{n=1}^{\infty}\frac{1}{n}\left(1-\delta^n\right)F(n).$$

- Then, average output is given by:

$$\mathbb{E}\left[\bar{y}^x(n) \,|n \geq 1\right] = \frac{\delta}{p}\frac{1-p}{1-\delta}\sum_{n=1}^{\infty}\frac{1}{n}\left(p^n-(\delta p)^n\right).$$

  We use that:

$$\vartheta(p) = \sum_{n=1}^{\infty}\frac{p^n}{n} = -\log(1-p),$$

  which is shown by taking derivatives on both sides:

$$\vartheta'(p) = \sum_{n=1}^{\infty}p^{n-1} = \frac{1}{1-p},$$

  and integrating back evaluating at $p=0$ to obtain the constant to be equal to zero. Then, we have that:

$$\begin{aligned}\mathbb{E}\left[\bar{y}^x(n) \,|n \geq 1\right] &= -\delta\frac{1-p}{1-\delta}\log(1-p) + \delta\frac{1-p}{1-\delta}\log(1-\delta p)\\ &= \delta\frac{1-p}{1-\delta}\log\left(\frac{1-\delta p}{1-p}\right).\end{aligned}$$

- This is the average output of chained orders conditional on falling in a chain.
- A very similar formula to the one employed in the paper, except that we have only consider production
- Overall production among spot is:

$$\mathbb{E}\left[y^s(n) \,|n \geq 0\right] = 1$$

Thus, overall production is:

$$\mathbb{E}\left[y^s\left(n\right)|n\geq 0\right]\left(1-\mu^x\right)+\mu^x\mathbb{E}\left[\bar{y}^x\left(n\right)|n\geq 1\right]$$

$$=\left(1-\mu^x\right)+\mu^x\delta\frac{1-p}{1-\delta}\log\left(\frac{1-\delta p}{1-p}\right)$$

$$=\left(1-\mu^x\right)+\mu^x\frac{1-\mu_x+\rho}{\mu_x+\rho}\frac{\delta}{1-\delta}\log\left(\frac{1-\delta\left(\mu_x+\rho\right)}{1-\mu_x+\rho}\right).$$

- Using the notation in the paper:

$$\left(1-\mu\right)+\mu\frac{1-\mu+\rho}{\mu+\rho}\frac{\delta}{1-\delta}\log\left(\frac{1-\delta\left(\mu+\rho\right)}{1-\mu+\rho}\right).$$

B.2 Truncation

- We verified that the formula generalized to the case with correlation across chained orders and spot orders.
- Simplified notation:

$$\mu_x=\mu$$

- Now, we investigate the role of truncation.
- The chain length is as before, distributed

$$F\left(n\right)=\frac{\left(1-p\right)}{p}\left(p\right)^n,$$

  which integrates to 1.
- Trunctation states that if n is above a threshold value η, the chained order is dropped. Thus, production is given by:

$$y^x\left(n\right)=\delta\left(1+\delta+\delta^2...\delta^{n-1}\right)=\delta\left(\frac{1}{1-\delta}-\frac{\delta^n}{1-\delta}\right),\text{ if }n\leq\eta,$$

and otherwise:

$$y^x(n) = \delta\left(1+\delta+\delta^2...\delta^{n-1}\right) = \delta\left(\frac{1}{1-\delta}-\frac{\delta^\eta}{1-\delta}\right) = \delta\left(\frac{1}{1-\delta}-\frac{\delta^n}{1-\delta}-\frac{\delta^\eta}{1-\delta}\left(1-\delta^{n-\eta}\right)\right).$$

Thus, with truncation,

$$\bar{y}^x(n) = \frac{y(n)}{n} = \frac{\delta}{n}\left(\frac{1}{1-\delta}-\frac{\delta^n}{1-\delta}\right) - \bar{\gamma}(n)$$

where

$$\bar{\gamma}(n) = \begin{cases} 0 & \text{if } n \leq \eta \\ \\ \frac{1}{n}\frac{\delta^\eta}{1-\delta}\left(1-\delta^{n-\eta}\right) & \text{if } n > \eta. \end{cases}$$

Thus, from our earlier derivation:

$$\mathbb{E}\left[\bar{y}^x(n)\,|n \geq 1\right] = \frac{1-\mu_x+\rho}{\mu_x+\rho}\frac{\delta}{1-\delta}\log\left(\frac{1-\delta(\mu_x+\rho)}{1-\mu_x+\rho}\right) - \mathbb{E}\left[\bar{\gamma}(n)\,|n \geq 1\right].$$

We compute:

$$\mathbb{E}\left[\bar{\gamma}(n)\,|n \geq 1\right] = \mathbb{E}\left[\frac{1}{n}\frac{\delta^\eta}{1-\delta}\left(1-\delta^{n-\eta}\right)|n \geq \eta\right]\Pr(n \geq \eta).$$

- Working by pieces:

$$\begin{aligned} \Pr(n \geq \eta) &= \sum_{n=\eta}^{\infty} F(n) \\ &= \sum_{n=\eta}^{\infty} \frac{(1-p)}{p}(p)^n \\ &= \frac{(1-p)}{p}p^\eta\left(\sum_{n=\eta}^{\infty}(p)^{n-\eta}\right) \\ &= \frac{(1-p)}{p}p^\eta\left(\sum_{j=0}^{\infty}(p)^{j}\right) \\ &= \frac{(1-p)}{p}\frac{p^\eta}{(1-p)} = p^{\eta-1}. \end{aligned}$$

- Next, we compute:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\frac{\delta^{\eta}}{1-\delta}\left(1-\delta^{n-\eta}\right)|n\geq\eta\right] &= \sum_{n=\eta}^{\infty}\frac{1}{n}\frac{\delta^{\eta}}{1-\delta}\left(1-\delta^{n-\eta}\right)\frac{F(n)}{\Pr(n\geq\eta)} \\
&= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\sum_{n=\eta}^{\infty}\frac{1}{n}\left(1-\delta^{n-\eta}\right)(p)^{n-\eta} \\
&= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\sum_{n=\eta+1}^{\infty}\frac{1}{n}\left(1-\delta^{n-\eta}\right)(p)^{n-\eta} \\
&= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\sum_{j=1}^{\infty}\frac{1}{\eta+j}\left(1-\delta^{j}\right)(p)^{j} \\
&= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\sum_{j=1}^{\infty}\frac{1}{\eta+j}\left(p^{j}-(p\delta)^{j}\right) \\
&= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\left(\frac{1}{p^{\eta}}\sum_{j=1}^{\infty}\frac{1}{\eta+j}p^{j+\eta}-\frac{1}{(p\delta)^{\eta}}\sum_{j=1}^{\infty}\frac{1}{\eta+j}(p\delta)^{j+\eta}\right) \\
&= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\left(\frac{1}{p^{\eta}}\sum_{j=1}^{\infty}\frac{1}{\eta+j}p^{j+\eta}-\frac{1}{(p\delta)^{\eta}}\sum_{j=1}^{\infty}\frac{1}{\eta+j}(p\delta)^{j+\eta}\right) \\
&= \frac{(1-p)}{p}\frac{\delta}{1-\delta}\left(\frac{\delta^{\eta}}{\delta p^{\eta}}\sum_{j=0}^{\infty}\frac{1}{\eta+j}p^{j+\eta}-\frac{1}{\delta(p)^{\eta}}\sum_{j=0}^{\infty}\frac{1}{\eta+j}(p\delta)^{j+\eta}\right) \\
&= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\left(\Phi(p,1,\eta)-\Phi(p\delta,1,\eta)\right)
\end{aligned}$$

The term:

$$\Phi(z,1,\eta)=\sum_{j=0}^{\infty}\frac{z^{j}}{\eta+j}=\int_{0}^{\infty}\frac{\exp(-\eta t)}{1-z\exp(-t)}dt,$$

is the Lerch transcendent. Notice that for $\eta=1$:

$$\begin{aligned}
\int_{0}^{\infty}\frac{\exp(-t)}{1-z\exp(-t)}dt &= \int_{0}^{\infty}\frac{1}{\exp(t)-z}dt=\frac{1}{z}\log(1-z\exp(-t))\,|_{t=0}^{t=\infty} \\
&= -\frac{1}{z}\log(1-p)=\frac{1}{z}\sum_{k=1}^{\infty}\frac{z^{k}}{k}=\frac{1}{z}\sum_{j=0}^{\infty}\frac{z^{j+1}}{1+j}=\frac{1}{z}\sum_{j=1}^{\infty}\frac{z^{j}}{1+j}.
\end{aligned}$$

- and thus:

$$= \frac{(1-p)}{p}\frac{\delta^{\eta}}{1-\delta}\left(\frac{\delta^{\eta}}{\delta p^{\eta}}\Phi(p,1,\eta)-\frac{1}{\delta p^{\eta}}\Phi(p\delta,1,\eta)\right)$$

- This formula allows us to generalize to the case with truncation:

$$\mathbb{E}\left[\frac{1}{n}\frac{\delta^{\eta}}{1-\delta}\left(1-\delta^{n-\eta}\right)|n\geq\eta\right]p^{\eta-1} = \frac{(1-p)}{p}\frac{\delta}{1-\delta}\left(\frac{\delta^{\eta}}{\delta p}\Phi\left(p,1,\eta\right)-\frac{1}{\delta p}\Phi\left(p\delta,1,\eta\right)\right)$$

Thus, we obtain:

$$(1-\mu^{x})+\mu^{x}\frac{1-\mu_{x}+\rho}{\mu_{x}+\rho}\frac{\delta}{1-\delta}\left(\log\left(\frac{1-\delta\left(\mu_{x}+\rho\right)}{1-\mu_{x}+\rho}\right)-\left(\frac{\delta^{\eta}}{\delta p}\Phi\left(p,1,\eta\right)-\frac{1}{\delta p}\Phi\left(p\delta,1,\eta\right)\right)\right).$$

In the special case of $\eta\to\infty,\rho=0$ we have the original formula.

B.3 Multinomial Case

Now consider the situation where instead of experiencing delays, a particular transaction is executed but the receiver fails to receives the funds with probability $1-q$ without a delay. Otherwise, it faces the same delay structure as above. This probability is completely random.

Given a chain of length n, the chances that there are k delays is:

$$M\left(k,n\right)=\left(\begin{array}{c}n\\k\end{array}\right)q^{k}\left(1-q\right)^{n-k}.$$

While we can compute average number of delays, they have to be ordered: meaning, delays occur starting from k-th unit. Delays have to be sequential.

Producition is:

$$y\left(k,n\right)=n-k+\delta\frac{1-\delta^{k}}{1-\delta}.$$

Then, we integrate across all k:

$$\sum_{k=0}^{n}y\left(k,n\right)\times M\left(k,n\right)=(n-k)\sum_{k=0}^{n}M\left(k,n\right)+\frac{\delta}{1-\delta}\sum_{k=0}^{n}\frac{1}{1-\delta}M\left(k,n\right)-\frac{\delta}{1-\delta}\sum_{k=0}^{n}\delta^{k}\left(\begin{array}{c}n\\k\end{array}\right)q^{k}\left(1-q\right)^{n-k}$$

But we know that the multinomial integrates to 1:

$$\sum_{k=0}^{n} \left(\begin{array}{c} n \\ \\ k \end{array} \right) q^k (1-q)^{n-k} = 1.$$

We also know that its expectation is:

$$\sum_{k=0}^{n} k \left(\begin{array}{c} n \\ \\ k \end{array} \right) q^k (1-q)^{n-k} = \sum_{k=0}^{n} k \frac{n!}{(n-k)!k!} q^k (1-q)^{n-k}$$

but by property of factorials:

$$n-k! = n-k \times n-k-1 \times n-k-2...$$

$$n-(k-1)! = n+1-k \times n-k-1 \times n-k-2...$$

$$\left(\begin{array}{c} n \\ \\ k \end{array} \right) = \frac{n}{k} \times \left(\begin{array}{c} n-1 \\ \\ k-1 \end{array} \right)$$

$$\begin{aligned}
\sum_{k=1}^{n} k \left(\begin{array}{c} n \\ \\ k \end{array} \right) q^k (1-q)^{n-k} &= n \sum_{k=1}^{n} \left(\begin{array}{c} n-1 \\ \\ k-1 \end{array} \right) q^k (1-q)^{n-k} \\
&= nq \left(\sum_{k=1}^{n-1} \left(\begin{array}{c} n-1 \\ \\ k-1 \end{array} \right) q^{k-1} (1-q)^{n-k} \right) \\
&= nq \left(\sum_{k=1}^{n-1} \left(\begin{array}{c} n-1 \\ \\ k-1 \end{array} \right) q^{k-1} (1-q)^{n-1-(k-1)} \right)
\end{aligned}$$

change of variables guarantees:

$$nq.$$

Hence:

$$\sum_{k=0}^{n} kM\left(k,n\right) = qn.$$

Thus:

$$\left(n-k\right)\sum_{k=0}^{n} M\left(k,n\right) = \left(1-q\right)n.$$

And, moreover:

$$\frac{1}{1-\delta}\sum_{k=0}^{n} M\left(k,n\right) = \frac{1}{1-\delta}.$$

Finally, we have that:

$$\sum_{k=0}^{n} \delta^{k} \left(\begin{array}{c} n \\ \\ k \end{array} \right) q^{k}\left(1-q\right)^{n-k} = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ \\ k \end{array} \right) \left(\delta q\right)^{k}\left(1-q\right)^{n-k}$$

which by the binomial theorem equals:

$$\left(1-q+\delta q\right)^{n}$$

Hence, all in all, production is:

$$\mathbb{E}\left[y\left(k,n\right)|n\right] = \left(1-q\right)n + \delta\frac{1-\left(1-q+\delta q\right)^{n}}{1-\delta}.$$

Thus, for $q=0$ we have:

$$\mathbb{E}\left[y\left(k,n\right)|n\right] = n$$

and when $q=1$

$$\mathbb{E}\left[y\left(k,n\right)|n\right] = \delta\frac{1-\delta^{n}}{1-\delta}.$$

corresponding to the formula above for chained production. Then, average production is:

$$\mathbb{E}\left[\bar{y}\left(k,n\right)|n\right] = \left(1-q\right) + \frac{1}{n}\frac{\delta-\delta\left(1-q+\delta q\right)^{n}}{1-\delta}.$$

Thus:

$$\mathbb{E}\left[\mathbb{E}\left[\bar{y}\left(k,n\right)|n\right]\right]=\left(1-q\right)+\sum_{n=1}^{\infty}\frac{1}{n}\frac{\delta-\delta\left(1-q+\delta q\right)^{n}}{1-\delta}F\left(n,p\right)$$

$$\sum_{n=1}^{\infty}\frac{1}{n}\frac{\delta-\delta\left(1-q+\delta q\right)^{n}}{1-\delta}F\left(n,p\right)=\frac{1-p}{p}\frac{\delta}{1-\delta}\left(\log\left(1-p\left(1-q+\delta q\right)\right)-\log\left(1-p\right)\right).$$

Thus, we obtain a generalized formula:

$$\mathbb{E}\left[\bar{y}^{x}\right]=\left(1-q\right)+\frac{1-p}{p}\frac{\delta}{1-\delta}\left(\log\left(1-p\left(1-q+\delta q\right)\right)-\log\left(1-p\right)\right).$$

Which again, collapses to the general formula.

B.4 Resampling

- Suppose that after observing their own draw, of place in a Chain, chained types can try to resample.
- Potentially there is a cost, but for now assume there is not, and only assume that types above $n>\eta$, instead of dropping out, resample.
- If they resample, they contact another chain at random and they are placed somewhere in the chain, their own chain shrinks to size η.
- They can again match with a chain of size η, which would force another resampling.
- The process repeats itself indefinitely until there are no further resampling.
- The question is, to what distribution does the process converge too?
- We've already seen that the distribution of chain sizes will be Geometric:
  - $F\left(n,p\right)$ with $\left(1-p\right)p^{n}$ be the distribution of chain size with $n\geq 0$
- Let $z\in\{0,1,...,\infty\}$ be the number of rounds.
  - it is better to write this as a continuous-time process so we avoid problems with integers

  - matches occur at rate intensity λ

- Let $G_0(n) = F(n,p)$ be the initial distribution
- After the initial distribution, the fraction

$$\mu_0 = \sum_{n>\eta} G_0(n)$$

will seek to resample. They randomly match with some other agent. A match is successful if it falls within a chain of length less than η, because otherwise we ran into the same problem again. The available chains are:

$$A_0 = \sum_{n=0}^{\eta-1} G_0(n)$$

out of the pool:

$$1 - \mu_0 = \sum_{n=0}^{\eta} G_0(n).$$

Thus, a match is useful with probability:

$$p_0 = \frac{\alpha_0}{1-\mu_0} \to p(z) = \frac{\sum_{n=0}^{\eta-1} G_0(n)}{\sum_{n=0}^{\eta} G_0(n)}$$

and

$$p(n,z) = \frac{G_z(n)}{\sum_{j=0}^{\eta} G_z(j)}.$$

- Thus, we can think of the mass of unsuccessful matches as shrinking at rate:

$$\dot{\mu}(z) = -\lambda p(z)(1-\mu(z))\mu(z).$$

If matches occur, chains add one more type.

- Thus, they satisfy the law of motion:

  - inflow probabilities are:

– for $n = 0$

$$\dot{G}(0,z) = -\lambda p(0,z)\mu(z)(1-\mu(z)) = -\lambda G(0,z)\mu(z).$$

and for $n \in \{1, \eta - 1\}$

$$\dot{G}(n,z) = -\lambda\left(G(n,z)\mu(z) - 2G(n-1,z)\mu(z)\right)$$

and for $n = \eta$

$$\dot{G}(\eta,z) = \lambda 2G(\eta-1,z)\mu(z).$$

- Now, we verify things by adding masses:

$$\sum_{n=0}^{\eta}\dot{G}(n,z) = \lambda G(0,z)\mu(z) + \lambda G(1,z)\mu(z) + G(1,z)\mu(z)... + G(\eta-1,z)\mu(z)$$

which accounts for the preservation of mass probability mass since:

$$\lambda G(0,z)\mu(z)+\lambda G(1,z)\mu(z)+G(1,z)\mu(z)...+G(\eta-1,z)\mu(z) = \lambda p(z)(1-\mu(z))\mu(z)$$

Hence:

$$\sum_{n=0}^{\eta}\dot{G}(n,z) + \dot{\mu} = 0.$$

- Next, we try to find a way to characterize the evolution of the system. The key challenge is the presence of η. In matrix form we can write:

$$\mathbf{G}(z) = \begin{bmatrix} G(1,z) \\ G(2,z) \\ \\ \\ G(\eta,z) \end{bmatrix}$$

$$\dot{\mathbf{G}} = -\lambda \mathbf{A} \mathbf{G}(z) \mu(z)$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & & \cdots & 0 \\ -2 & 1 & 0 & & & \\ 0 & -2 & 1 & & & \\ & & & \ddots & & \\ \vdots & & & -2 & 1 & \vdots \\ 0 & & & & -2 & 0 \end{bmatrix}.$$

- We write the system as in non-linear form as:

$$\dot{\mathbf{G}} = -\lambda \mathbf{A}(z) \mathbf{G}(z)$$

$$\mathbf{A}(z) = \mathbf{A}\mu(z)$$

which in Matrix form solves:

$$\mathbf{G}(z) = \exp\left(-\lambda \int_0^z \mathbf{A}(\tau) d\tau\right) \mathbf{G}(0),$$

where

$$\mu(z) = -\lambda \left(1 - \frac{G(\eta, z)}{\sum_{n=0}^{\infty} G(n, z)}\right) \mu(z) (1 - \mu(z)).$$

- The solution to $\mu(z)$ can also be written as a time-varying problem:

$$\dot{\mu}(z) = -\lambda(z) \mu(z) (1 - \mu(z))$$

where

$$\lambda(z) = \lambda \left(1 - \frac{G(\eta, z)}{\sum_{n=0}^{\infty} G(n, z)}\right)$$
$$\dot{\mu}(z) = -\lambda(z) \mu(z) (1 - \mu(z)).$$

Guess and verify that:

$$\mu(z) = \frac{1}{1+\exp\left(\int_0^z \lambda(\tau)\,d\tau\right)}.$$

Derivative:

$$\dot{\mu}(z) = -\lambda(\tau)\frac{\exp\left(\int_0^z \lambda(\tau)\,d\tau\right)}{\left(1+\exp\left(\int_0^z \lambda(\tau)\,d\tau\right)\right)^2}$$

$$-\lambda(\tau)\mu(z)(1-\mu(z)) = -\lambda(\tau)\frac{1}{1+\exp\left(\int_0^z \lambda(\tau)\,d\tau\right)}\left(1-\frac{1}{1+\exp\left(\int_0^z \lambda(\tau)\,d\tau\right)}\right),$$

proving the solution.

- Thus, we need to solve the system:

$$\mathbf{G}(z) = \mathbf{G}(0)\exp\left(-\mathbf{A}\int_0^z \lambda\mu(z)\,d\tau\right)$$

where$k$ is some constant.

$$\mu(z) = \frac{1}{1+\exp\left(\int_0^z \lambda(\tau)\,d\tau\right)}.$$

- Thus, suppose that the process does indeed converge in finite time.

$$\mathbf{G}(\infty) = \exp(-k\mathbf{A})\,\mathbf{G}(0)$$

- Then:

$$\mathbf{e}\mathbf{G}(\infty) = 1 = \mathbf{e}\exp(-k\mathbf{A})\,\mathbf{G}(0).$$

Thus,

$$\exp\left(-\lambda\int_0^z \mathbf{A}(\tau)\,d\tau\right).$$

Hence, we look for k such that

$$1 = \mathbf{e}\exp(-k\mathbf{A})\,\mathbf{G}(0).$$

If we find such a k we have the solution.

C. Proofs of Section 4

C.1 Preliminary Observations

We begin with a set of identities that are convenient to proof the main results.

Average Price: Definitions and Identities. Recall that the average price of a worker as it's expenditures relative to total consumption: $Q \equiv E^w/C^w$. The saver's expenditure is shown to be $E^s = (1-\beta)\,B$. By the income expenditure identity:

$$E^w = 1 - (1-\beta)\,B. \tag{23}$$

Total consumption by the worker, given its optimal expenditures, is given by:

$$C^w = \frac{E^w - S^w}{q} + S^w = \frac{E^w}{q} - \left(\frac{1}{q} - 1\right) S^w.$$

where

$$S^w = \min\left\{\max\left\{0, \tilde{B} - B\right\}, 1 - (1-\beta)\,B\right\},$$

$X^w = \left(E^w - S^w\right)/q$.

Thus, we have that:

$$Q = \left(\frac{1}{q} - \left(\frac{1}{q} - 1\right)\frac{S^w}{E^w}\right)^{-1} = \left(\frac{1}{q}\left(1 - \frac{S^w}{E^w}\right) + \frac{S^w}{E^w}\right)^{-1}.$$

Also, we have that:

$$Q = \frac{E^w - S^w}{C^w} + \frac{S}{C^w} = q\frac{X^w}{C^w} + \frac{S}{C^w}.$$

The last two proof the following result.

Lemma 3. *Q is average of the goods prices weighted by the worker's consumption shares and Q is the harmonic mean of the goods prices weighted by the expenditure share.*

Marginal Expenditure and Borrowing Prices. For convenience, we define again the two prices that enter in marginal decisions. First, we define the *marginal expenditure*

price:

$$\tilde{q}_t^e \equiv q_t \underset{\sim}{\overset{\prec}{\approx}}_{[B_t \geq B^\star(\tilde{B}_t)]} + \left(1 - \underset{\sim}{\overset{\prec}{\approx}}_{[B_t \geq B^\star(\tilde{B}_t)]}\right).$$

Next, we define the *marginal borrowing price*:

$$\tilde{q}_{t+1}^b \equiv q_{t+1} \underset{\sim}{\overset{\prec}{\approx}}_{[\tilde{B}_{t+1} \leq B_{t+1}]} + \left(1 - \underset{\sim}{\overset{\prec}{\approx}}_{[\tilde{B}_{t+1} \leq B_{t+1}]}\right).$$

Analysis of the Marginal Expenditure Price. Next, we describe the behavior of $\tilde{q}_t^e$ in equilibrium. We have that $\tilde{q}_t^e = \tilde{q}^e\left(B_t, \tilde{B}_t\right)$ where:

$$\tilde{q}^e\left(B, \tilde{B}\right) \equiv q\left(\mu\left(B, \tilde{B}\right)\right) \underset{\sim}{\overset{\prec}{\approx}}_{[B \geq B^\star(\tilde{B})]} + \left(1 - \underset{\sim}{\overset{\prec}{\approx}}_{[B \geq B^\star(\tilde{B})]}\right).$$

In this expression, we are using:

$$\mu\left(B, \tilde{B}\right) = 1 - (1-\beta)B - \min\left\{\max\left\{0, \tilde{B} - B\right\}, 1 - (1-\beta)B\right\}.$$

Observe that

$$\mu\left(B, \tilde{B}\right) = \begin{cases} 0 & B < B^\star\left(\tilde{B}\right), \\ 1 + \beta B - \tilde{B} & B \in \left[B^\star\left(\tilde{B}\right), \tilde{B}\right], \\ 1 - (1-\beta)B & B > \tilde{B}. \end{cases}$$

$\mu\left(B, \tilde{B}\right)$ is continuous, starts at zero and increases up to $B = \tilde{B}$, starting from that point, the function is decreasing. Since q_t is monotone in μ_t, the function $\tilde{q}^e\left(B, \tilde{B}\right)$ must follow the same pattern in the interior of the middle segment of teh function and in the last segment. Also notice that when $B < B^\star\left(\tilde{B}\right)$, $q^e = 1$. Next, we have that since μ is continuous, μ goes to 0 from above as $B \downarrow B^\star\left(\tilde{B}\right)$ so we can write,

$$\lim_{B \downarrow B^\star(\tilde{B})} \tilde{q}^e\left(B, \tilde{B}\right) = \lim_{\mu \downarrow 0} q(\mu) \cdot 1 = \delta^{-1}$$

where we used the fact that $\lim_{\mu \downarrow 0} \mathcal{A}(\mu) = \delta$. Hence, the function $\tilde{q}^e$ is discontinuous at $B = B^\star$ because $\lim_{B \uparrow B^\star\left(\tilde{B}\right)} \tilde{q}^e = 1 \neq \delta^{-1}$, the function is also not monotonic. Then[39]

$$\tilde{q}^e\left(B, \tilde{B}\right) = \begin{cases} 1 & B < B^\star\left(\tilde{B}\right) \\ q & B \in \left[B^\star\left(\tilde{B}\right), \tilde{B}\right] \\ q & B > \tilde{B}. \end{cases}$$

Next, we are interested in the behavior of $\frac{Q}{q^e}$, for reasons that become clear in the main text. We have that for $B < B^\star\left(\tilde{B}\right)$, since $q = 1$, it must be that $\frac{Q}{q^e} = 1$. For $B \geq \tilde{B}$ also $q^e = Q = q$. Therefore, $\frac{Q}{q^e} = 1$. In the middle range of values, we have that:

$$\frac{Q}{q^e} = \frac{1}{q} \cdot \frac{q}{\left(1 - \frac{S^w}{E^w}\right) + q\frac{S^w}{E^w}} = \frac{1}{\left(1 - \frac{S^w}{E^w}\right) + q\frac{S^w}{E^w}}.$$

Thus, we have the following formula:

$$\frac{Q}{q^e} = \begin{cases} 1 & B < B^\star\left(\tilde{B}\right) \\ \frac{1}{1 - \frac{\min\left\{\max\left\{0, \tilde{B} - B\right\}, 1 - (1-\beta)B\right\}}{1-(1-\beta)B} + q\left(\mu\left(B, \tilde{B}\right)\right)\frac{\min\left\{\max\left\{0, \tilde{B} - B\right\}, 1 - (1-\beta)B\right\}}{1-(1-\beta)B}} & B \in \left[B^\star\left(\tilde{B}\right), \tilde{B}\right] \\ 1 & B > \tilde{B}. \end{cases}$$

Since at $B = \tilde{B}$ we have $\frac{S^w}{E^w} = 0$, the function is continuous at that point. However,

$$B \downarrow B^\star\left(\tilde{B}\right) \implies \frac{S^w}{E^w} \downarrow 1, \text{ and } q \downarrow \delta^{-1} > 1,$$

at that point. Thus, $\frac{Q}{q^e} = \delta$ at $B = B^\star\left(\tilde{B}\right)$ (and at its right limit) and $Q/q^e = 1$ at the left limit of this point. Namely, the function Q/q^e is discontinuous at $B^\star$.

Analysis of the Marginal Borrowing Price. Next, we investigate the behavior of the marginal borrowing price. Recall that this is the price of consumption at which the worker trades off future consumption when he borrows marginally. Let B_{t+1} be the

[39] In fact, the value of q^e at $B = B^\star$ will not be used since the function is discontinuous at that point and optimality conditions will be characterized via the right and left limit. However, we choose to define it according to the intuition presented lines above.

debt level the worker chooses today for next period and $\tilde{B}_{t+1}$ the next period's SBL. We have that $\tilde{q}^b_{t+1} = \tilde{q}^b\left(B_{t+1}, \tilde{B}_{t+1}\right)$ where:

$$\tilde{q}^b\left(B, \tilde{B}\right) \equiv q\left(\mu\left(B, \tilde{B}\right)\right) \underset{\approx}{\prec}_{[B \geq \tilde{B}]} + \left(1 - \underset{\approx}{\prec}_{[B \geq \tilde{B}]}\right).$$

Then,

$$\tilde{q}^b\left(B, \tilde{B}\right) = \begin{cases} 1 & B < B^\star\left(\tilde{B}\right), \\ 1 & B \in \left[B^\star\left(\tilde{B}\right), \tilde{B}\right], \\ q & B > \tilde{B}. \end{cases}$$

We have observed that for $B = B^\star\left(\tilde{B}\right)$, all consumption is spot and thus $q = 1$. However, at $B = \tilde{B}$, the function features a discontinuity since:

$$\lim_{B \downarrow \tilde{B}} \tilde{q}^b\left(B, \tilde{B}\right) = q\left(\mu\left(\tilde{B}, \tilde{B}\right)\right) = q\left(1 - (1-\beta)\,\tilde{B}\right) > 1.$$

Next we investigate the behavior of $\frac{Q}{q^b}$.

For any $B < B^\star\left(\tilde{B}\right)$, both q^b and Q must equal 1, thus, $\frac{Q}{q^b} = 1$. Then, we have that for $B \in \left[B^\star\left(\tilde{B}\right), \tilde{B}\right]$, it must the case that $\frac{Q}{q^b} = Q$ because $q^b = 1$. Finally, when $B > \tilde{B}$, $\tilde{q}^b\left(B, \tilde{B}\right) = Q = q$. Thus, we have:

$$\frac{Q}{\tilde{q}^b} = \begin{cases} 1 & B < B^\star\left(\tilde{B}\right), \\ Q & B \in \left[B^\star\left(\tilde{B}\right), \tilde{B}\right], \\ 1 & B > \tilde{B}, \end{cases}$$

where

$$Q = \left(\frac{1}{q}\left(1 - \frac{S^w}{E^w}\right) + \frac{S^w}{E^w}\right)^{-1}$$

and $\lim_{B \downarrow B^\star(\tilde{B})} Q = 1$.

Average Price Elasticity. Next, we derive the elasticity of the average price with respect to total debt, in equilibrium—i.e., after replacing $E^w = 1 - (1-\beta) B$. This elasticity is critical for the Ramsey policy analysis.

We have that the average price is the harmonic mean:

$$Q = \frac{1}{\frac{1}{q}\left(1 - \frac{S^w}{E^w}\right) + \frac{S^w}{E^w}}.$$

where both the price q, E^w, and S^w are functions of B.

Thus, we have that:

$$\frac{\partial Q}{\partial B} = -Q \cdot \frac{\left(1 - \frac{S^w}{E^w}\right) \frac{\partial}{\partial B}\left[\frac{1}{q}\right] - \left(\frac{1}{q} - 1\right) \frac{\partial}{\partial B}\left[\frac{S^w}{E^w}\right]}{\left(\frac{1}{q}\left(1 - \frac{S^w}{E^w}\right) + \frac{S^w}{E^w}\right)}.$$

The numerator has two additional derivatives. The first one is:

$$\frac{\partial}{\partial B}\left[\frac{S^w}{E^w}\right] = \begin{cases} 0 & B < B^\star\left(\tilde{B}\right) \\ \left[\frac{\tilde{B}-B}{1-(1-\beta)B}\right]\left(\frac{(1-\beta)}{1-(1-\beta)B} - \frac{1}{\tilde{B}-B}\right) & B \in [B^\star\left(\tilde{B}\right), \tilde{B}) \\ 0 & B \geq \tilde{B} \end{cases}$$

where the term in the intermediate region follows from:

$$\frac{\partial}{\partial B}\left[\frac{S^w}{E^w}\right] = \frac{S^w}{E^w}\left(\frac{1}{S^w}\frac{\partial}{\partial B}\left[S^w\right] - \frac{1}{E^w}\frac{\partial}{\partial B}\left[E^w\right]\right),$$

but then for $B \in [B^\star\left(\tilde{B}\right), \tilde{B})$

$$\frac{\partial}{\partial B}\left[S^w\right] = -1,$$

and

$$\frac{\partial}{\partial B}\left[E^w\right] = -\left(1-\beta\right).$$

Hence:

$$\frac{\partial}{\partial B}\left[\frac{S^w}{E^w}\right] = \left[\frac{\tilde{B}-B}{1-(1-\beta)B}\right]\left(\frac{(1-\beta)}{1-(1-\beta)B} - \frac{1}{\tilde{B}-B}\right) < 0.$$

We evaluate the limits of this function. Clearly, at

$$\frac{\partial}{\partial B}\left[\frac{S^w}{E^w}\right]\Big|_{B\uparrow\tilde{B}} = -\frac{1}{1-(1-\beta)\,\tilde{B}} < -1,$$

and at

$$\begin{aligned}
\frac{\partial}{\partial B}\left[\frac{S^w}{E^w}\right]\Big|_{B\downarrow B^\star} &= \frac{(1-\beta)\left(\tilde{B}-B^\star\left(\tilde{B}\right)\right)}{\left(1-(1-\beta)\,B^\star\left(\tilde{B}\right)\right)^2} - \frac{1}{\left(1-(1-\beta)\,B^\star\left(\tilde{B}\right)\right)} \\
&= \frac{(1-\beta)\left(\tilde{B}-B^\star\left(\tilde{B}\right)\right) - \left(1-(1-\beta)\,B^\star\left(\tilde{B}\right)\right)}{\left(1-(1-\beta)\,B^\star\left(\tilde{B}\right)\right)^2} \\
&= \frac{(1-\beta)\,\tilde{B}-1}{\left(1-(1-\beta)\,B^\star\left(\tilde{B}\right)\right)^2} \\
&= -1.
\end{aligned}$$

The last steps follows from the definition of $B^\star$. Hence, this derivative is discontinuous. For the second derivative of interest, recall that:

$$\frac{\partial}{\partial B}[q] = \frac{\partial}{\partial \mu}[q]\cdot\frac{\partial \mu}{\partial B},$$

and

$$\frac{\partial \mu}{\partial B} = \frac{\partial}{\partial B}[E^w] - \frac{\partial}{\partial B}[S^w]$$

since $\mu = qX = E^w - S^w$. Then, we have that:

$$\frac{\partial}{\partial B}[q] = q_\mu \mu \frac{\frac{\partial}{\partial B}[E^w] - \frac{\partial}{\partial B}[S^w]}{E^w}.$$

We can express this as:

$$\frac{\partial}{\partial B}\left[q^{-1}\right] = \frac{1}{q}\epsilon^q_\mu \frac{\frac{\partial}{\partial B}[E^w] - \frac{\partial}{\partial B}[S^w]}{E^w - S^w}.$$

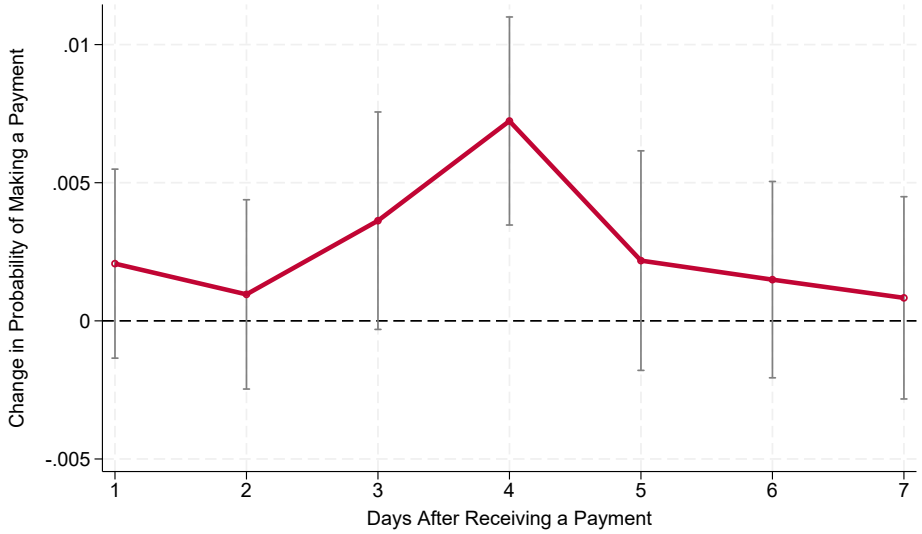
Hence, we obtain:

$$\frac{\partial Q}{\partial B}\frac{1}{Q} = -Q \cdot \left\{ \left(1 - \frac{S^w}{E^w}\right) \frac{1}{q} \epsilon_{\mu}^{q} \frac{\frac{\partial}{\partial B}\left[E^w\right] - \frac{\partial}{\partial B}\left[S^w\right]}{E^w - S^w} - \left(\frac{1}{q} - 1\right) \frac{S^w}{E^w} \left[\frac{\frac{\partial}{\partial B}\left[S^w\right]}{S^w} - \frac{\frac{\partial}{\partial B}\left[E^w\right]}{E^w} \right] \right\}.$$

D. Empirical Appendix

D.1 Local Projection: 2005 Daily Data

Figure 11: Triggered Payments: Local Projection



Notes: The figure shows a local projection Jordà (2005); Jordà and Taylor (2025), where the dependent and independent variable include 1 lag, as determined by a VAR-based optimal lag selection criteria. Data is daily and spans 2005.

D.2 Intensive Margin Regressions

We also study the intensive margin relationship between payments made and received. We follow equation (1), but now considering the (inverse hyperbolic sine of the) amounts sent and received by firm j. Results are presented in Table 3 consistent with those of the extensive margins; firms become more likely to make payments once they receive money transfers. The elasticity across all firms in column (1) implies that a 1% increase in the amounts received by a firm on a given day increases the amounts sent the next day by approximately 0.04%.

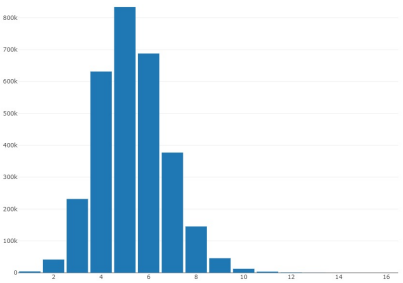
Table 3: Triggered Payments Across Firms—Intensive Margin

Dependent variable: Payments made by firm j at time t (asinh)

	(1)	(2)	(3)	(4)	(5)	(6)
$\ln amount\ received_{j,t-1}$	0.038	0.018	0.051	0.050		0.038
(*asinh*)	(.0005)***	(.0009)***	(.0009)***	(.0008)***		(.0005)***
$Small_j \times \ln amount\ received_{j,t-1}$		0.021				
		(0.001)***				
$\frac{Assets_j}{Sales_j} \times \ln amount\ received_{j,t-1}$			-0.023			
			(0.001)***			
$\frac{Liabilities_j}{Sales_j} \times \ln amount\ received_{j,t-1}$				-0.023		
				(0.001)***		
$\ln amount\ received_{j,t-5}$					0.013	0.012
(*asinh*)					(.0004)***	(.0003)***
Observations	10,628,428	6,936,593	7,793,996	7,793,996	10,468,698	10,468,698
Adjusted R^2	0.501	0.487	0.424	0.493	0.500	0.501
$t, j \times$ d FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The independent variable in column (1) is the total amount (in colones) received by firm j from another firm within the last 24 hours, transformed using the inverse hyperbolic sine function. This variable is interacted with firm j's size, assets-to-sales and liabilities-to-sales in columns (2), (3), and (4) respectively. Columns (5) and (6) add an independent variable equal to the total amount (in colones) received by firm j from another firm 5 days before time t, transformed using the inverse hyperbolic sine function. Robust standard errors, clustered by firm, are in parentheses. We include time and firm×weekday fixed-effects. Data is daily and spans 2024.

Figure 12: Distribution of Chain Sizes



Notes: The figure shows the distribution of chained sizes based on a 5% random sample of spot firms in the economy used as initial nods. Data is collapsed for December 2024 to construct the chains.

D.3 Placebo Exercises

Table 4: Placebos—Receiving Payments Right After Making a Transfer

Dependent variable: Payment made by firm j at time t

	A. Extensive margin				B. Intensive margin (*asinh*)		
	$h=1$	$h=2$	$h=3$		$h=1$	$h=2$	$h=3$
	(1)	(2)	(3)		(4)	(5)	(6)
$received_{j,t+h}$	-0.001	-0.003	-0.007	$\ln amount$	-0.0005	-0.002	0.00003
	(0.0004)***	(0.0004)***	(0.0004)*	$received_{j,t+h}$	(0.0004)	(0.0004)***	(0.0004)
Observations	10.5M	10.5M	10.5M		10.5M	10.5M	10.5M
Adjusted R^2	0.468	0.468	0.468		0.501	0.501	0.501
$t, j \times$ d FE	Yes	Yes	Yes		Yes	Yes	Yes

Notes: In panel A, the independent variables are indicators equal to one if firm j received a payment in $t+h$. In panel B, the independent variables are the total amounts (in colones) received by firm j from another firm in $t+h$, transformed using the inverse hyperbolic sine function. Robust standard errors, clustered by firm, are in parentheses. We include time and firm×weekday fixed-effects. Data is daily and spans 2024.

D.4 Distribution of Chain Sizes

D.5 Chained Payments and Prices

We consider the following regression to explore the role of deferred payments on prices for chained firms on transaction v:

$$\ln Price_{vpjt} = \delta + \gamma\, Deferred_{vpjt} + \lambda_p + \lambda_j + \lambda_t + \varepsilon_{vpjt},$$

where $Deferred_{vpjt} = 1$ if sale v has a deferred paymen as opposed to being paid in full immediately—and zero otherwise, and we include product, firm, and time fixed effects. We can also interact $Deferred_{vpjt}$ with an indicator, $Chained_{kt} = 1$, which depends on whether firm k—which made payment v to firm j—was classified as chained.

Column (1) of Table 5 shows that prices for a given product are 12% higher for transactions with a deferred payment. Most relevant for our purposes, Column (2) shows this effect is larger for firms classified as chained, indicating that there is a higher price charged for a given quantity of good p if the firm buying the product is chained. Column (3) further breaks down transactions depending on the type of both the receiver (j) and sender (k) of the payment. Reassuringly, a higher price is charged only if the payment is made by a chained firm, regardless of the receiver's type.

Table 5: Chained Payments and Prices

Dependent variable: Price of transaction t for good p sold by firm j to firm k at time t (asinh)

	(1)	(2)	(3)
$Deferred_{vpjt}$	0.101	0.073	0.028
	(0.018)***	(0.017)***	(0.007)***
$Chained_{kt} \times Deferred_{vpjt}$		0.038	
		(0.006)***	
$Chained_{jt}$&$Chained_{kt} \times Deferred_{vpjt}$			0.037
			(0.007)***
$Spot_{jt}$&$Chained_{kt} \times Deferred_{vpjt}$			0.049
			(0.007)***
$Spot_{jt}$&$Spot_{kt} \times Deferredt_{vpjt}$			0.008
			(0.006)
Observations	50,836,448	50,836,448	25,459,682
Adjusted R^2	0.670	0.670	0.699
p, j, t FE	Yes	Yes	Yes

Notes: The independent variable in column (1) is an indicator equal to 1 if transaction v;s payment was deferred, as opposed to having an immediate payment. This variable is interacted with a dummy that depends on firm k's classification as chained or spot in column (2). Column (3) further decomposes transactions depending on whether firm j and k were either chained or spot at time t; the omitted category having firm j as chained and firm k as spot. Robust standard errors, clustered by firm, are in parentheses. We include product, firm, and time fixed-effects. Data is by sale and spans 2024.