Author:
Liu, Yilei

Title:
Essay on money and credit with limited commitment

Essay on Money and Credit with Limited Commitment

By

Yilei Liu

School of Economics
University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Social Science and Law.

April 2022

Abstract

This thesis consists of three chapters.

In Chapter 1 we construct a monetary model based on Lagos and Wright (2005) where unsecured credit and money are used as means-of-payments, and we analyse how the cost and the quality of record-keeping technology affect welfare. Specifically, monitoring agents' debt repayment is costly but is essential to the use of unsecured credit because of limited commitment. To finance this cost, fees on credit transactions are imposed, and the maximum credit limit that is incentive compatible depends on such fees and monitoring level. Alternatively, the use of money avoids such costs. A higher credit limit does not necessarily improve welfare, especially when the limit is high: the benefit from increased trade surpluses from a higher credit limit is offset by the increased cost of monitoring to achieve the improvement. Moreover, under the optimal arrangement, optimal credit limit decreases with the marginal cost of monitoring. When the cost is sufficiently low, a pure credit equilibrium is optimal. When the marginal cost is high, it is optimal to have a pure-currency economy. But when the cost is at an intermediate level, we show that credit is sustainable but not socially optimal. In this range, the implementable credit limit leads to a higher trade surplus than in a pure monetary economy, but owing to the cost of operating the record-keeping system, social welfare in credit equilibrium is lower than the welfare in a pure monetary equilibrium. In addition we show that there can be a non-monotonic relationship between the optimal record-keeping level and the optimal credit limit.

In Chapter 2 we construct a monetary model with both formal and informal sectors to study optimal monetary policy and fiscal policy under needs for public goods. In the formal market, agents have access to financial services but are subject to taxation. In the informal market, agents can only use cash to trade but can avoid taxation. We show that it is optimal to have higher inflation for economies with larger informal sectors, a pattern we see in developing countries. We capture two effects of monetary policy on social welfare: high inflation decreases equilibrium real balances therefore decreases the trade surplus in the informal sector. However, it increases the seigniorage revenue and contributes to the public expenditure. We show that the optimal monetary policy depends on the size of the informal sector. Specifically, the inflationary policy is socially optimal only when the informal sector is larger than a certain threshold; otherwise, it is

optimal to choose a deflationary policy. The reason for this is that large-sized informal sector decreases the governments' tax capacity in the formal sector, but it provides a large tax base for seigniorage, so it is optimal to rely more on high inflation and seigniorage revenue to provide public goods. In addition, we show that better credit conditions can alleviate the tax burden on seignorage and decrease the optimal inflation rate.

In Chapter 3 we present a micro-founded model with bank deposits used as the means of payment. Banks obtain returns from productive assets with a cost and an economy of scale but are subject to limited commitment, and the assets held are used as collateral for deposit issuance. In the absence of government intervention, the private sector may not provide sufficient liquidity if the productive asset is scarce. As a result, interventions, such as interest on reserves and reserve requirements, may increase welfare by improving liquidity supply. A high interest on reserves makes banks profitable for holding excess reserves which are fully pledgeable, and the required reserve requirements, either proportional or fixed, directly increase banks' pledgeable assets holding. The optimal design of banking regulations and monetary policy depends on the asset productivity level and the fiscal requirement in an economy. We explicitly take the public financing aspect of such policies into account. If the central bank has little fiscal burden, then the optimal policy does not use inflation but uses interest on reserves, as well as reserve requirements to increase liquidity provision, with the optimal interest rate on excess reserve procyclical. If the central bank has to contribute to the fiscal revenue, then it is optimal to have high inflation and reserve requirements, but only to generate seigniorage revenues, and at optimal central bank does not use IOER.

Dedication and acknowledgements

First and foremost, I am extremely grateful to my supervisor Tai-Wei Hu. I could not have undertaken this journey without his assistance and dedicated involvement. His rigorous academic attitude and insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. His enthusiasm, energy and deep understanding of the area of Monetary Economics always inspired me in all the time of my academic research and daily life. Being his first Ph.D. student and working with him were the luckiest things during my Ph.D.

I would also like to thank many faculty members from UOB who have dedicated their time to help me improve my work during meetings and seminars, including Arpad Abraham, Erlend Berg, Sek-yu Choi, Pawel Doligalski, Katerina Raoukka, Gregory Jolivet, David Pacini, In-Uck Park, Alireza Sepahsalari, Hans Sievertsen, Hakki Yazaci, Yanos Zylberberg and many others.

Special thanks also go to Guillaume Rocheteau and Asgerdur Petursdottir for their useful comments and suggestions for my research. I would also like to express my sincere gratitude to Jonathan Temple for his supervision in the first year of my Ph.D. His advice at the beginning of my research has been extremely valuable.

A big thank you to all my Ph.D. colleagues and friends for all the discussions, coffees, and laughs we shared. It is their kind help and support that have made my study and life in the UK a wonderful time.

Last, but not least, my deepest gratitude goes to Yichen Ma and my parents for their encouragement, unconditional love, and support. I would also like to acknowledge my sponsor, the China Scholarship Council.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

Table of Contents

List of Tables

List of Figures

Chapter 1

Costly Credit in Credit Economies

1.1 Introduction

Though technological progress in recent decades has greatly enhanced record keeping, a crucial ingredient to sustain credit arrangements (Kocherlakota and Wallace, 1998), it does not eliminate the basic friction that hinders credit transactions, namely, limited commitment of economic agents. Moreover, the cost of establishing a modern credit system which allows for accurate updates of transaction data is still non-trivial. In contrast, an economy can avoid this costly credit arrangement by using money as the only payment instrument. Indeed, in many economies today, most transactions are accomplished with cash alone.[1] These observations naturally lead to two research questions when we consider endogenous record-keeping and enforcement: first, when is it feasible to adopt the technology and sustain a credit economy? Second, when the use of credit is sustainable, would it be socially optimal to use it?

In this paper, we propose a monetary model based on Lagos and Wright (2005) with endogenous liquidity needs and introduce a costly and incomplete record-keeping technology to answer these questions. Specifically, we endogenize the level of record-keeping and study the socially optimal use of payment instruments when agents cannot commit to repay their future obligations. Previous literature, as Bethune et al. (2018) have shown that in an environment with complete record-keeping and limited commitment, a pure credit economy is better than a pure currency economy in terms of welfare, a

[1]See The 2020 McKinsey Global Payment Report for the cash usages by country, where in emerging economies, such as Argentina, India, Indonesia, Malaysia and Mexico, cash accounts for an average 80 percent of total transaction volume. See also the Fifth Global Payment System Survey by the World Bank for the cashless transaction per capita according to country income levels.

result consistent with Kocherlakota (1998) in more general settings.[2] The gain in welfare is that as agents bear the opportunity cost of holding fiat money and thereby bring an inefficient level of liquidity to trade, credit arrangement relaxes this liquidity constraint and hence improves gains from trade. Thus, when the cost of record-keeping is zero, it is optimal to use credit. But once we introduce the cost of record-keeping, where better enforcement requires more real resources from society, this comparison will then depend on the marginal cost of providing a better quality record-keeping service.

Indeed, by endogenizing the record-keeping technology, we capture a non-trivial trade-off that is missing in the literature: while record-keeping of higher quality directly increases liquidity provided by the credit system, its cost requires real resources which will affect liquidity provision indirectly. Unsurprisingly, credit equilibrium is sustainable only when the marginal cost of improving record-keeping is not too high. However, for a range of intermediate marginal costs, we find that there can be a non-monotonic relationship between credit limits and welfare. Moreover, for a range of such marginal costs, a pure monetary economy generates higher welfare than a credit economy, even though a higher level of liquidity can be sustained in a credit economy.

Following Hu et al. (2009) and Andolfatto (2013), we assume that the government has limited coercion power, and the cost of record-keeping has to be borne by private agents in a way that this burden also affects liquidity provision due to limited commitment. Our model shares the standard environment according to Lagos-Wright, where agents meet bilaterally in a decentralized market and use money and/or credit to trade. This decentralized market is followed by a centralized market, where agents repay obligations and rebalance money holdings. Following Kocherlakota and Wallace (1998) and Bethune et al. (2015), we introduce a record-keeping system which is subject to stochastic record-updating; that is, it may fail to record a default when it happens and it is costly to lower the rate of such failures.

Coupled with the limited commitment friction, the accuracy of the record-keeping technology implies a level of self-enforcement that is linked to the incentive-compatible credit limit. To finance a more accurate record-keeping system and hence a higher level of enforcement, agents who access the technology in order to issue credit have to pay a fee used to finance the cost of operating the record-keeping technology. Thus, in principle, a higher level of enforcement does not necessarily imply a higher debt limit, as the cost of better record-keeping technology also enters into the incentive constraints for repayments.

[2]Bethune et al. (2018) show this by assuming a fixed bargaining protocol across the pure currency and the pure credit economies.

For low marginal costs of record-keeping, however, we show that it is always monotonic, but this may not be true for higher marginal costs.

Our main result characterizes the endogenous record-keeping technology for a given marginal cost, and how the optimal means-of-payments and liquidity provision varies with that cost. For a given marginal cost of improving record-keeping, we find that the welfare can be non-monotonic in the credit limit. In particular, for low credit limit below the amount of liquidity that money alone can provide, higher credit limit would only require higher social cost for better enforcement without increasing the overall liquidity, and hence would lower welfare. This result is a natural consequence of the result in Gu et al. (2016) when we add cost to the record-keeping system. However, while welfare would increase with credit limit after that, it might decrease again when the credit limit is high. The intuition is that the marginal benefit from an increase in credit limit becomes smaller as the limit increases, but the marginal cost that is required remains constant or becomes higher. As a result, the optimal credit limit may not be the highest implementable one.

Our second main result then characterizes the regions under which credit is better than money and vice versa. When the marginal cost of enhancing enforcement is low, it is optimal to use credit and optimal credit limit decreases as such marginal cost increases. When such cost is very high, money dominates; in particular, agents would not even be willing to access the credit given the choice as it requires a high credit fee to finance the credit system. By contrast, the optimal arrangement in the intermediate range of such marginal costs is more subtle. In such range, the credit limit is high but the opportunity of holding money is also high, leading to the fact that though agents prefer to use pure credit, the welfare from the pure credit equilibrium is, however lower than the welfare from a pure monetary equilibrium. The inconsistency comes from the fact that when agents compare the opportunity cost of holding money and the cost of accessing credit, the former cost does not enter the social welfare but the later cost does. As a result, when the liquidity provided by money is not too far from the credit economy, a pure currency economy in fact delivers higher welfare than a credit economy. However, if the government provides the record-keeping technology, all agents prefer to access it, making money no value, so it might not be the optimal policy in terms of social welfare.

Related literature

This paper is related to a growing literature that studies endogenous credit limits by taking limited commitment and enforcement seriously. Conceptually, our formulation of endogenous credit limits follows that in Kehoe and Levine (1993) and Kocherlakota and

Wallace (1998), and is similar to those taking the same inheritance to the Lagos-Wright (2005) framework for tractability, including Sanches and Williamson (2010), Chiu et al. (2018), and Bethune et al. (2018), among others.

The paper is also related to the literature that discusses the optimal use of means-of-payments that follows Kocherlakota (1998). We have a similar model setup to Gu et al. (2016), but in their paper the record-keeping technology is complete and costless. Their findings suggest that credit is neutral in terms of both allocation and welfare when money is valued, which is a special case of our results by letting the marginal cost of enforcement be zero. Other relevant papers include Bethun et al. (2018); whose focus is more on the non-stationary credit equilibrium that is welfare-improving, and Araujo and Hu (2018); in which the access to money and credit is exogenous.

Other papers emphasize the imperfection of record-keeping technology. In Kocherlakota and Wallace (1998) and Bethune et al. (2015), the record-keeping technology is incomplete and it is shown that better quality of record-keeping always improves credit condition because it is costless. For the same reason, more credit always has a welfare improving effect. Still others have introduced an exogenous cost to access credit but they assume perfect commitment therefore credit is unconstrained, as in Wang et al. (2020) and Dong and Huangfu (2021). In contrast to these papers, we characterize both the incompleteness and the cost of record-keeping technology and endogenize it by taking the limited-commitment friction seriously.

Another related paper is Lotz and Zhang (2016): in their work, the cost of accessing credit is paid by sellers. This leads to complementarities between buyers and sellers and hence multiple equilibria issue; moreover, it generates inefficiency from bargaining-related holdup problems. Instead, we let buyers pay the credit cost and have all the bargaining power, by means of which we have more tractable equilibrium analysis and can focus more on endogenizing the record-keeping technology and welfare analysis.

Earlier papers indicating that credit is not always welfare-improving includes Chiu et al. (2018). The authors capture a general equilibrium effect of better access to credit as it increases credit user's consumption, but also drives up the equilibrium price, which harms the money users. In our work here, agents endogenously choose the means-of-payment and the negative effect of credit comes from the cost of enforcement and limited commitment friction.

1.2 Environment

The baseline model is based on Lagos and Wright (2005). Time is discrete and has an infinite horizon. The economy is populated by two continuum sets of agents, labelled *buyers* and *sellers*. Every period is divided into two stages. In the first stage there is a decentralized market (DM), where buyers and sellers meet in pairwise meetings, and only buyers desire to consume and only sellers can produce the DM good. The probability that a buyer meets a seller (and vice versa) is $\sigma \in (0, 1]$ and the buyer makes a take-it-or-leave-it-offer to the seller. In the second stage there is a centralized market where all agents meet. All agents can consume and produce a single good, called the CM good, with linear preferences and a linear technology to produce. Both DM goods and CM goods are perfectly divisible and nonstorable.

The instantaneous utility functions of buyers and sellers are given by

$$U^b(y, x) = u(y) + x \text{ and } U^s(y, x) = -y + x,$$

respectively, where y is the amount of DM consumption (production) and x is the amount of CM consumption (negative number is interpreted as production). The utility function is twice continuously differentiable with $u'(\cdot) > 0$, $u''(\cdot) < 0$, $u'(0) = +\infty$, $u'(\infty) = 0$ and $u(0) = 0$. Let y^* denote the consumption level that maximizes the match surplus $u(y) - y$, which is determined by $u'(y^*) = 1$. The cost function for the seller is assumed to be linear for expositional simplicity, and it has no bearing on our results. The discount factor is β, and denote the discount rate as $r = 1/\beta - 1$.

The government operates a financial system with a costly and imperfect monitoring technology. The buyer can access the financial system in the DM by incurring a credit fee χ (in terms of the CM goods) that is due in the CM.[3] The monitoring technology records the buyer's credit transaction in the DM, payment of the credit fee and repayment of debt up to a limit D in the coming CM. The technology updates the buyer's status to be either of good standing (g) or bad standing (b), depending on the buyer's payments. The status of a buyer is observable to all sellers, but the status is updated imperfectly. Any repayment beyond the credit limit D cannot be recognised and recorded. Moreover, if a buyer of status g who incurs debt d in the DM repays at least $\min\{D, d\}$ to the seller and the credit fee χ to the government in the CM, he remains in status g. If he fails to repay $\min\{D, d\}$ to the seller or the credit fee χ in the CM, his status will be updated

[3]Of course we can let sellers pay the cost. But this will cause holdup problem and complementary, as discussed in Lotz and Zhang (2016).

to b with probability $\epsilon \in [0, 1]$.[4]

The probability of successful record-updating ϵ measures the enforcement level, or the quality of record-keeping.[5] It is endogenously determined and is costly to implement. In particular, the government needs to spend $\xi v(\epsilon)$ units of CM good (in per buyer terms) to sustain ϵ, and $v(\cdot)$ is a convex cost function. We start with a linear cost function $\xi v(\epsilon) = \xi\epsilon$ and generalize the main results under any convex cost function in the last section. With linear cost function, the parameter ξ is the marginal cost of monitoring and enforcement. Such a cost will depend on the technology for reviewing transactions and updating records, but it will also depend on the difficulty to overcome rent-seeking opportunities that would allow a defaulter to avoid a bad record. The government uses the credit fee, imposed on buyers who use credit, to finance this cost.[6] As mentioned above, the credit fee is denoted by χ, and it occurs when the buyer issues credit in a meeting, but is payable only in the following CM.[7]

There is only one asset, fiat money, which is intrinsically useless but is perfectly durable and recognizable. Let M_t denote the aggregate money supply at the beginning of period t, and in the first period, M_0 is given to the buyers evenly. In the baseline model we assume money supply is constant over time, and we discuss the case where money grows at a constant rate in the last section. There is a competitive market for money and CM goods in the CM and we express real balances as z (in terms of the CM goods). We consider stationary equilibrium where the aggregate real balance is constant over time.

[4]Alternatively, we can let buyers pay the credit fee ex ante before accessing credit; by doing so, the record-updating has to be done also ex ante, which means the technology labels a buyer who does not pay the fee directly as *bad*, before the buyer enters DM and uses credit. Because otherwise, buyers could always avoid the tax without leaving any record and being punished. The two versions, paying credit fee ex post or ex ante, do not influence the incentive constraint and the main results.

[5]This formulation is commonly used in literature to measure the quality of the record-updating, but it has other interpretations. Kocherlakota and Wallace (1998) interpret the probability as a measure of the average lag between transaction updating. In Bethune et al. (2015), it is a measure of the sophistication of the financial system. In Sanches and Williamson (2010), it represents the fraction of sellers that have monitoring potential.

[6]Of course, this can be done through a private sector as well; the main point here is that the monitoring is costly and there is a need for budget-balancing. We assume that this is done by the benevolent government to avoid any further agency problem.

[7]We following Araujo and Hu (2018) and interpret the cost of credit as the social resources invested to the record-keeping system, including to operate the credit card system and to screen, monitor and reveal information. This social cost is financed by taxation called credit fee and agents can always default on it. There are other ways to model costly credit. For example, in Wang et al. (2020) and Dong and Huangfu (2021), buyers incur a fixed or proportional utility cost in the decentralized market while using credit. In Lotz and Zhang (2016), sellers randomly incur heterogeneous utility cost and the cost is zero for a fraction of sellers. Such utility cost allows for heterogeneity but avoid the payment issue that could indirectly impact the incentive constraint and therefore credit limit, which is our main focus in this paper.

In a DM meeting a buyer can always use cash to finance his consumption. Or he can choose to issue credit but subject to the credit fee and repayment constraint. If a buyer defaults on his obligation and being detected and labeled as *bad*, the punishment is permanent autarky. In literature, the more commonly used punishment of being caught is permanently exclusion from the credit system. Our punishment may be too strict as a *bad* buyer could always use money if money has value in the equilibrium. We adopt this punishment as it simplifies the incentive constraint and does not affect the welfare ranking when money and credit coexist. Moreover, it does not affect any analysis in the pure credit equilibrium as money has no value and being excluded from using credit implies autarky. [8]

1.3 Equilibrium Analysis

In this section, we begin with a partial equilibrium analysis for the endogenously given credit limit D and credit fee χ. We characterize different stationary equilibrium according to the use of different means-of-payment.

In particular, we show that a monetary equilibrium ($z > 0$) exists only when the credit limit is not too high, and there is a threshold for the credit limit above which money is dominated by credit and a pure credit equilibrium exists. When credit limit is below that threshold, depending on the value of credit limit and credit fee, either a coexistence equilibrium exists where buyers use both money and credit in the decentralized trade, or a pure monetary equilibrium exists where only money is used even though credit is provided. For the non-monetary equilibrium ($z = 0$), we show that there is another threshold of credit limit above which buyers access credit and a pure credit equilibrium exists, otherwise buyers do not trade in the decentralized market even though credit is provided, and the economy is in an autarky equilibrium.

In section 1.3.1, we fully characterize the range of (D, χ) such that different equilibrium exists. We start from a pure credit equilibrium where money has no value. Then we introduce the use of money and discuss the range of (D, χ) such that money is accepted and thus has value while credit is provided. Then we show that when a monetary and a non-monetary equilibrium both exist, the buyers' ex ante payoff in the monetary

[8] A micro-foundation explanation of such punishment is that the trading protocol depends on the buyer's status, as in Araujo and Hu (2018). Specifically, a buyer with a good standing makes a take-it-or-leave-it offer to the seller, but the seller makes a take-it-or-leave-it to a buyer with a bad standing. They show that such punishment in fact a feature of the optimal trading protocol to enhance incentive-compatibility.

equilibrium is always higher than it is in the non-monetary equilibrium. Later on we will show that the social welfare in the monetary equilibrium is also higher than it is in the non-monetary equilibrium when both equilibrium exist.

In section 1.3.2, we move on to endogenize the credit limit. For given monitoring technology ϵ and credit fee χ, we find the set of credit limit that satisfies the buyers' incentives to repay and a balanced government budget. Then we discuss the relationship between the highest implementable credit limit and the environment rate.

1.3.1 Equilibrium under a given credit limit

Here we assume an exogenously given credit limit, $D \leqslant y^*$, and we assume that buyers always repay their debts and credit fees in the CM if they choose to access the financial system in the DM trade. In the final subsection we shall consider the incentive problem of repayment and endogenize this limit D.

For a given D, and assuming that buyers always repay their debts up to D, a stationary equilibrium consists of equilibrium real balance, z, that each buyer holds when leaving the CM and equilibrium DM trade, (y, p, d), where $p \leqslant z$ is the equilibrium money transfer (denominated in CM-goods terms) from the buyer to the seller, $d \leqslant D$ is the equilibrium credit (denominated in CM-goods terms) that each buyer issues to seller and y is the equilibrium output. Further, equilibrium requires that z be the optimal money holding for each buyer, and, whenever $d > 0$, buyers are willing to access the financial system in the DM and repay the debt d and pay the credit fee χ. Finally, sellers are willing to produce the amount y in exchange for the payment $p + d$.

In section 1.3.1.1, we discuss the non-monetary equilibrium ($z = 0$) for given credit limit D and credit fee χ. In this equilibrium, money has no valued and hence no agents accept money as means-of-payment. If buyers decide to access the financial system $d > 0$, we call it a pure credit equilibrium, otherwise when $d = 0$, we call it an autarky equilibrium. In subsection 1.3.1.2 we move on to discuss the range of given (D, χ) such that a monetary equilibrium exists ($z > 0$). In this equilibrium, money has value χ and is accepted as means-of-payment. If buyers do not issue credit $d = 0$, we call it a pure monetary equilibrium. Otherwise, if buyers use credit $d > 0$, we call it a coexistence equilibrium.

1.3.1.1 Non-monetary equilibrium

We start form a non-monetary equilibrium when money has no value and credit is the only means-of-payment. To do so, we denote the value function for a buyer who accessed the financial system in the previous DM and enters CM with debt d by $\hat{W}(d)$.

The standard Lagos-Wright argument implies that the CM value function is linear on liquidity condition d, that is, $\hat{W}(d) = -d - \chi + \hat{W}(0)$. We assume that in the equilibrium buyers can promise to repay the debt d up to the threshold D and the credit fee χ and remains a *good*-type in the record keeping system. Denote the buyer's continuation value in the DM as $\hat{V}$ that solves:

$$\hat{V} = \max_{d \le D} \{\sigma[u(\hat{y}) + \hat{W}(d)] + (1-\sigma)\hat{W}(0)\}, \tag{1.1}$$

subject to seller's participation constraint $\hat{y} \le d$. With probability σ the buyer meets the seller the the buyer makes an take-it-or-leave-it offer $(\hat{y}, d)$. The seller, if agree, produces goods $\hat{y}$ with linear cost $-\hat{y}$ and the buyers consume the goods in the DM. Then the match breaks and the buyer and the seller enter the CM to settle the debt. It is straightforward that the buyer uses all the liquidity $d = D$ to consume DM goods and hence $\hat{y} = D$. When there is no meeting between the buyer and the seller, no trade happens in the DM.

We denote $\hat{\Psi}$ as the buyer's value of accessing the record-keeping technology which is $\hat{\Psi} = \sigma[u(D) - D - \chi]$. Note that if the buyer chooses to not access finance, the equilibrium is in autarky.

For the given credit limit $D \leqslant y^*$ and credit fee χ, the next lemma describes the buyer's decision to access the financial system and the stationary equilibrium allocations.

Lemma 1.3.1. *Let $\hat{\chi} \equiv u(y^*) - y^*$. A non-monetary equilibrium always exists with $z = 0$, and we characterize the stationary equilibrium below with $y = d$.*

1. *If $\chi > \hat{\chi}$, buyers do not use credit in DM with $d = 0$ and the economy is in autarky.*
2. *Otherwise, there exists $\hat{D}(\chi)$, determined by $\sigma[u(\hat{D}) - \hat{D} - \chi] = 0$, such that*
   - *(2.1) if $D \geqslant \hat{D}(\chi)$, in equilibrium buyers only use credit $d = D$ in the DM;*
   - *(2.2) if $D < \hat{D}(\chi)$, $d = 0$ in the DM and the economy is in autarky.*

In Lemma 1.3.1, when the credit fee is higher than the first-best gain from trade $(\chi > \hat{\chi})$, buyers do not access the financial system for any credit limit. As money has no value, the DM trade is not active.

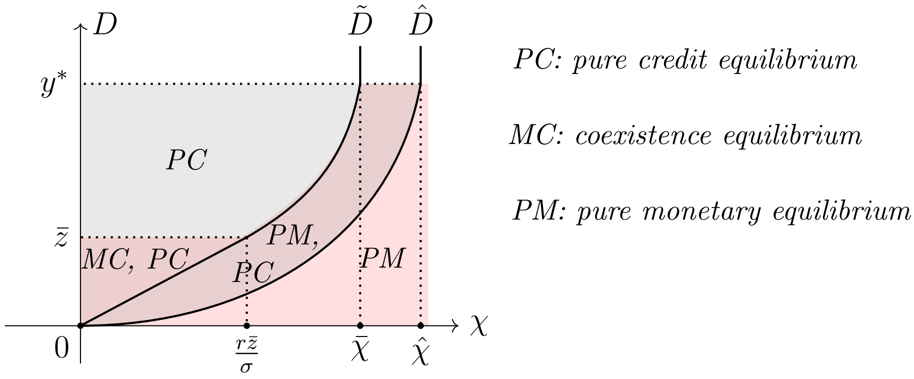

Figure 1.1: Stationary equilibrium for given credit limit and credit fee

When the credit fee is lower than that threshold, the buyer's decision to access the financial system depends on the credit limit. In Lemma 1.3.1 (2.1), when the credit limit is higher than a threshold $\hat{D}$, buyers use credit to trade in the DM, and a pure credit equilibrium exists. Otherwise in Lemma 1.3.1 (2.2), when the credit limit is low, the gain from trade cannot compensate for the cost of using credit, so buyers do not access the financial system and do not trade in the DM, regardless of the credit limit.

We summarize the results in Lemma 1.3.1 in Figure 3.1. In the grey area ($D \geqslant \hat{D}$) a pure credit equilibrium (PC) exists. When the credit fee increases, the buyer access credit only if the credit limit also goes up. Otherwise there is no credit equilibrium.

1.3.1.2 Monetary equilibrium

We move on to the case when both money and credit exist. Denote the value function for a buyer who did not access the financial system in the previous DM and enters CM with z real balances by $W^m(z)$, and denote a buyer who accessed the financial system in DM and enters CM with z real balance and d debt by $W^c(z,d)$. We assume that sellers do not carry money across periods and sell all their money holding accumulated from the DM in the coming CM, an assumption that is with no loss of generality, and hence we only consider buyers' problems. The buyers CM values follows:

$$W^m(z) = z + W^m(0) \text{ and } W^c(z,d) = z - d - \chi + W^m(0). \tag{1.2}$$

Note that $W^c(0,0) = W^m(0) - \chi$, which reflects the fact that the buyer is obligated to pay the credit fee χ, independent of his debt issuance. Note also that $W^m(0)$ solves the CM problem for a buyer who enters the CM with no money and did not access the financial system, and, by linearity, this represents the typical CM problem for the buyer, regardless of whether he has accessed the financial system in the previous DM.

Now we move to the DM problem. Denote the DM value function for a buyer with real balance z by $V(z)$, and it satisfies

$$V(z) = \sigma \max\{V^m(z), V^c(z)\} + (1-\sigma)W^m(z), \tag{1.3}$$

where $V^m(z)$ is the continuation value without accessing the financial system and solves

$$V^m(z) = \max_{y^m \leqslant p, p \leqslant z} \big\{u(y^m) + W^m(z-p)\big\}, \tag{1.4}$$

in which (y^m, p) is the offer made to the seller, subject to the liquidity constraint $p \leqslant z$, and the seller participation constraint $y^m \leqslant p$; and where $V^c(z)$ is the continuation value by accessing the financial system and solves

$$V^c(z) = \max_{y^c \leqslant p+d, p \leqslant z, d \leqslant D} \big\{u(y^c) + W^c(z-p, d)\big\}, \tag{1.5}$$

in which (y^c, p, d) is the offer made to the seller, subject to the liquidity constraint $p \leqslant z$, the credit limit $d \leqslant D$ and the seller participation constraint $y^c \leqslant p + d$. If a buyer is not matched with a seller, the buyer takes real balance z to the CM without debt or tax obligation.

By (1.2) and (1.4), the consumption of a buyer who only uses money in the DM is determined by $y^m = y^m(z) = y^*$ if $z \geqslant y^*$, and $y^m = y^m(z) = z$ otherwise. Similarly, by (1.2) and (1.5), the consumption of a buyer who accesses the financial system in DM is determined by $y^c = y^c(z, D) = y^*$ if $z + D \geqslant y^*$, and $y^c = y^c(z, D) = z + D$ otherwise. In both cases buyers will use all available payment capacity in DM to consume, until the first-best level of output is achieved.

Now we consider the CM problem, which is given by

$$W^m(0) = \max_z \big\{ -z + \beta V(z)\big\}. \tag{1.6}$$

When choosing his money holding, the buyer anticipates his DM decision to access the financial system or not (which is embedded in the Bellman equation for V in (1.3)), and plans accordingly. To characterize the optimal decision, define

$$\Psi^m \equiv \max_z \big\{ -rz + \sigma[u(y^m(z)) - y^m(z)]\big\}, \tag{1.7}$$

the solution to which is denoted by $\bar{z}$, and it solves $u'(\bar{z}) = \frac{\sigma+r}{\sigma}$, and define

$$\Psi^c(D) \equiv \max_z \big\{ -rz + \sigma[u(y^c(z+D)) - y^c(z+D) - \chi]\big\}, \tag{1.8}$$

with solution $z = 0$ if $D > \bar{z}$; $z = \bar{z} - D$ otherwise. The following lemma shows that the CM problem can be solved by comparing Ψ^m and $\Psi^c(D)$, and it summarizes equilibrium allocations and means-of-payments as well.

Lemma 1.3.2. *Let $\bar{\chi} \equiv \frac{\sigma[u(y^*)-y^*]-\Psi^m}{\sigma}$. We characterize the monetary equilibrium with the highest z below, in which $y = z + d$.*

1. *If $\chi > \bar{\chi}$, then in equilibrium real balance $z = \bar{z}$ and buyers do not use credit in the DM.*

2. *Otherwise, there exists $\tilde{D} = \tilde{D}(\chi)$, determined by $\Psi^m = \Psi^c(\tilde{D})$ such that*

   (2.1) if $D < \tilde{D}$, in equilibrium $z = \bar{z}$ and buyers do not use credit in the DM;

   (2.2) if $D \in [\tilde{D}, \max\{\bar{z}, \tilde{D}\})$, in equilibrium $z = \bar{z} - D$ and $d = D$ in the DM;

   (2.3) if $D \geqslant \max\{\bar{z}, \tilde{D}\}$, there is no monetary equilibrium.

Intuitively, the buyer's decision to access the financial system depends on the potential gain from trade achievable, which depends on the credit limit, D, and the cost of accessing it, namely the credit fee χ. Since the benefit is bounded above by the first-best gain from trade, a very high fee for credit trades will deter the buyer from accessing the financial system. Lemma 1.3.2(1) characterizes the precise upper bound $\bar{\chi}$, above which buyers do not access the financial system, regardless of the credit limit. In this case, there is a pure monetary equilibrium, and the equilibrium real balance is $z = \bar{z}$.

When χ is below that threshold, Lemma 1.3.2(2) shows that buyers' access to the financial system will depend on the credit limit. The critical value is given by the threshold $\tilde{D}$, which is determined by $\Psi^c(D) = \Psi^m$, and hence this result also shows that the buyer solves $\max\{\Psi^c(D),\ \Psi^m\}$ to determine whether to use credit or not. When the credit limit is below $\tilde{D}$ (which is a function of χ), Lemma 1.3.2 (2.1) states that the buyer will not issue credit but use money alone and there is a pure monetary equilibrium. Otherwise, Lemma 1.3.2 (2.2) states that the buyer will use credit, and supplement it with cash if the credit limit is not high enough , and there exists a coexistence equilibrium. Finally, when $D \geqslant \max\{\tilde{D}, \bar{z}\}$, there is no monetary equilibrium since using credit gives the buyer a higher interim payoff than using money.

In Figure 3.1, we use the red area ($D < \max\{\bar{z}, \tilde{D}\}$) to summarise the results in Lemma 1.3.1. Note that for given credit limit, when the credit fee is low, a pure credit equilibrium exists and when the cost goes up, a pure monetary equilibrium exists. There is a range of credit fees such that a money and credit equilibrium coexists with a pure credit equilibrium, a finding similar to what is found in Wang et al. (2020) under fixed cost, except that the credit is unconstrained in their model due to perfect commitment, which could be seen as a special case where D approaches y^*.

The next lemma compares the buyer's ex ante payoff, which is also the equilibrium social welfare, when multiple equilibria exist. Note that, however, a full discussion of welfare requires the government budget to be balanced. Nevertheless, the following results hold for any χ, and hence can be easily transferred to the case where we introduce government budget constraint.

Lemma 1.3.3. *Monetary equilibrium and pure credit equilibrium coexist when $D \in [\hat{D}, \max\{\bar{z}, \tilde{D}\})$, and the buyer's ex ante payoff in the monetary equilibrium is always higher than in the pure credit equilibrium.*

Lemma 1.3.3 implies that whenever the buyer chooses to use money when credit is provided, the allocation achieved by using money is always better than without it. Because the buyer faces an opportunity of cost of holding money, hence the trade surplus from using money has to be high to encourage the buyer to use money. Specifically, when $D \in [\tilde{D}, \bar{z})$, the credit limit is lower than the amount of real balances in a pure monetary equilibrium, hence the buyer uses money to compensate for the lack of credit and achieve a higher trade surplus in the coexistence equilibrium. When $D \in [\hat{D}, \tilde{D})$, the credit limit is lower than the threshold $\tilde{D}$. Lemma 1.3.2 implies that in this range, the buyer chooses to not access credit but use money as the only means-of-payment and therefore, the ex ante payoff from using money is even higher than in the pure monetary equilibrium.

Before we move on to endogenize the credit limits and credit fee, it is useful to define the set of credit limit and select the stationary equilibrium that we will focus on. Lemma 1.3.3 indicates that when $D \in [\hat{D}, \max\{\bar{z}, \tilde{D}\})$, a pure credit equilibrium is never socially optimal. Therefore we do not consider them in the following analysis. We denote the credit limit that is of interest as $D \in [\tilde{D}, y^*]$, where when $D < \bar{z}$, we focus on the coexistence equilibrium, as described in Figure 1.2.

1.3.2 Endogenous credit limit and incentive compatibility

Here we endogenize the credit limit by considering the buyer's incentive to repay, taking ϵ (the monitoring technology) and χ as given. We analyze equilibrium conditions taking in to account the buyers' repayment decisions, as well as government budget constraints related to the financing of the cost of the monitoring technology ϵ. In particular, for a candidate credit limit D and enforcement level ϵ, the pair is said to be *implementable* if

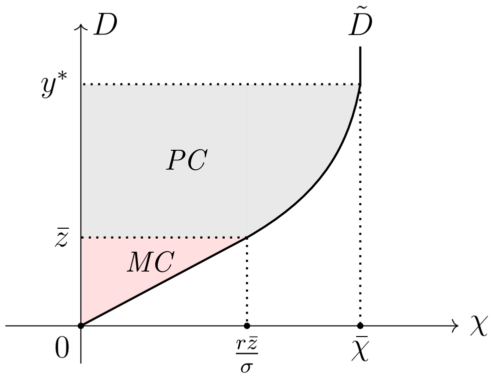

Figure 1.2: The set of credit limit $D \in [\tilde{D}, y^*]$

buyers are willing to access the financial system and repay up to D plus the credit fee, and if the credit fees collected are sufficient to finance the cost associated with ϵ, given the equilibrium allocation characterized by Lemma 1.3.2 and Lemma 1.3.1.

In the previous section we have characterized the optimal behaviour for a buyer with good standing. Note that since for a buyer with bad standing, it is the seller who makes the TIOLI offer, and since the cost of holding money across periods is positive, the continuation value for such a buyer is zero.

For incentive compatibility, consider a buyer with good standing in the CM, with a loan equal to the credit limit $D \leqslant y^*$ from the previous DM and the credit fee χ. If the buyer repays, his continuation value is $W^c(z, D) = z - D - \chi + W^m(0)$. Alternatively, the buyer could default. Since the record updating technology does not depend on the size of the default up to the credit limit plus the credit fee, it is optimal to default on the total obligation if the buyer chooses to default on any amount. By doing so, with probability ϵ the buyer's standing will be changed to bad, and, as we have seen, this implies a zero continuation value. Thus, to ensure that such a buyer repays his obligation, we need

$$W^c(z, D) \geqslant z + (1 - \epsilon)W^m(0), \tag{1.9}$$

where the left-side is the continuation value if the buyer repays, and the right-side is the continuation value otherwise: z is from the buyer selling his money holding and with probability $1 - \epsilon$ he is not caught and can continue as a buyer with good standing. As mentioned, we focus on the case where $D \geqslant \tilde{D}(\chi)$, hence $\Psi^c(D) \geqslant \Psi^m$, and we can rewrite the condition as

$$r(D + \chi) \leqslant \epsilon \Psi^c(D). \tag{1.10}$$

The government uses the credit fees to finance the cost of monitoring, and only buyers who access the financial system pay those fees. Assume then that all buyers with access use the system, and the government faces the following budget constraint:

$$\sigma\chi = \xi\epsilon. \tag{1.11}$$

In principle, budget constraint would only require the left-side of (1.11) to be no less than the right-side; we only consider equality because it is welfare-maximizing and simplifies the notation.

To summarize, the pair (D, ϵ) is implementable if, for χ satisfying (1.11), the incentive constraint (1.10) holds, and if the credit equilibrium exists, as in Lemma 1.3.2 and Lemma 1.3.1. Now, as we focus on the set $D \in [\tilde{D}(\chi), y^*]$, by (1.11) we can express $\tilde{D}$ as a function of (ϵ, ξ) and hence we use the notation $\tilde{D}(\epsilon; \xi)$ from this point. The following lemma characterizes implementable (D, ϵ) that $D \in [\tilde{D}, y^*]$.

Lemma 1.3.4. *For a given marginal cost of enforcement, ξ, the pair (D, ϵ) is implementable with $D \geqslant \tilde{D}(\frac{\xi\epsilon}{\sigma}) \equiv \tilde{D}(\epsilon; \xi)$ if and only if*

$$\begin{cases} \Psi^m + rD - \xi\epsilon \geqslant \frac{r}{\epsilon}D + \frac{r\xi}{\sigma} & \text{if } D \in [\tilde{D}, \max\{\bar{z}, \tilde{D}\}); \\ \sigma[u(D) - D] - \xi\epsilon \geqslant \frac{r}{\epsilon}D + \frac{r\xi}{\sigma} & \text{if } D \in [\max\{\bar{z}, \tilde{D}\}, y^*); \\ \sigma[u(y^*) - y^*] - \xi\epsilon \geqslant \frac{r}{\epsilon}D + \frac{r\xi}{\sigma} & \text{if } D = y^*. \end{cases} \tag{1.12}$$

Moreover, for each pair (ϵ, ξ) there is a threshold $\bar{D} = \bar{D}(\epsilon; \xi)$ such that D satisfies (1.12) if and only if $D \in [\tilde{D}, \bar{D}]$ (which might be empty in case $\tilde{D} > \bar{D}$).

Given Lemma 1.3.4, we can then fully characterize implementable (D, ϵ) when $D \in [\tilde{D}, y^*]$.

Proposition 1.3.1.

1. *For all $\xi > \frac{\sigma}{\sigma+r}\Psi^m$, there is no implementable (D, ϵ) such that $D \geqslant \tilde{D}(\epsilon; \xi)$.*
2. *For all $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, there exists a threshold $\bar{\epsilon}(\xi) > 0$ such that for any $\epsilon \in [0, \bar{\epsilon}(\xi)]$, $\tilde{D}(\epsilon; \xi) \leqslant \bar{D}(\epsilon; \xi)$. Moreover, in this range,*

   (2.a) $\frac{\mathrm{d}\bar{D}(\epsilon;\xi)}{\mathrm{d}\epsilon} \geqslant 0$ and $\frac{\mathrm{d}\bar{D}(\epsilon;\xi)}{\mathrm{d}\xi} \leqslant 0$, with strict inequality if $\bar{D} < y^*$;

   (2.b) $\bar{D}(0; \xi) = 0$, $\bar{D}(\bar{\epsilon}(\xi); \xi) \geqslant \bar{z}$ and the inequality is strict for all $\xi < \frac{\sigma}{\sigma+r}\Psi^m$.

Proposition 1.3.1 fully characterizes implementable (D, ϵ) when $D \in [\tilde{D}, y^*]$. Part (1) above shows that when the marginal cost of enforcement is too high ($\xi > \frac{\sigma}{\sigma+r}\Psi^m$), there is no implementable (D, ϵ) such that $D \geqslant \tilde{D}$. Otherwise, for any enforcement rate ϵ below the threshold $\bar{\epsilon}$, Lemma 1.3.4 implies that any $D \in [\tilde{D}, \bar{D}]$ is implementable under ϵ and Proposition 1.3.1 (2) implies that the range is not empty. As typical of the credit economies with endogenous credit limits, this implies that there is a continuum of credit limits that are incentive compatible when ξ is low.[9] Here we show more: there is also a range of enforcement rates that are both incentive feasible and fiscally feasible as enforcement is costly.

Our main focus, however, is on optimal arrangements that maximize social welfare. For any given $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$ and $\epsilon \leqslant \bar{\epsilon}(\xi)$, as we will later see, the highest implementable credit limit is $\bar{D} = \bar{D}(\epsilon;\xi)$. Proposition 1.3.1 (2.a) then shows that $\bar{D}$ is an increasing function of ϵ for any $\epsilon \in [0, \bar{\epsilon}(\xi)]$. Intuitively, an increasing in the enforcement rate ϵ influences the buyer's repayment decision through two effects. The first effect is an increase in the probability of being caught and of losing all the trade surplus, which encourages buyers to repay. This then allows for a higher credit limit, which also encourages buyers to access the financial system. However, the second effect is an increase in the credit fee, χ, which makes accessing the financial system more expensive and makes it more tempting to default as the credit fee is also part of the buyer obligation. It turns out that the first effect is stronger, and the reason for this is that, to make the arrangement incentive compatible, the marginal cost ξ cannot be too high to begin with and this limits the magnitude of the second effect. In contrast, an increase in ξ (but holding ϵ at constant) only has the second effect, and it decreases $\bar{D}$.

Proposition 1.3.1 (2.b) describes two boundary conditions of $\bar{D}(\epsilon;\xi)$. Clearly, when $\epsilon = 0$, buyers face no penalty by defaulting, and, anticipating this, sellers would never issue any credit, $\bar{D}(0;\xi) = 0$. When the enforcement is the highest, $\epsilon = \bar{\epsilon}(\xi)$, the credit limit exceeds $\bar{z}$ and hence the buyer does not carry cash. Since $\bar{D}$ is continuous in ϵ, this also implies that there is a cutoff point, $\underline{\epsilon}(\xi)$, below which the credit limit $\bar{D}$ is low, and hence the buyer carries cash to compliment the difference between $\bar{z}$ and $\bar{D}$, but above which the buyer only uses credit and does not carry cash. See Figure 1.3 for a depiction of the relationship between $\bar{D}$ and ϵ, and the ranges for the optimal portfolio of the buyer.

1.4 Welfare Analysis and Optimal Policy

Here we study the optimal scheme from a social planner's perspective, who takes ξ as given and chooses ϵ and D, subject to implementability. We do this in two steps. First, for a given ξ, we trace the welfare change as D varies and show that it is not necessarily optimal to implement the highest possible credit limit. This is also related to

[9] See, for example, Bethune et al. (2018) for a general discussion on this point.

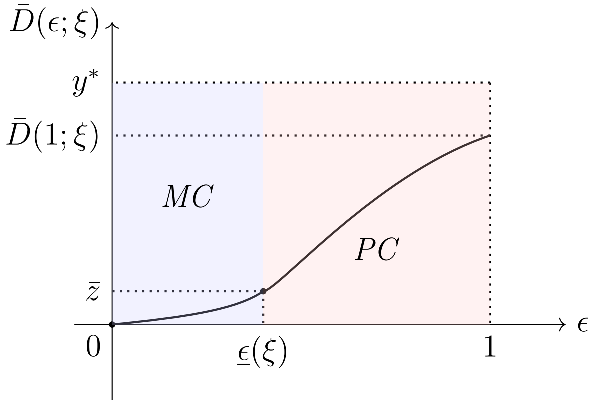

Figure 1.3: The highest implementable credit limit $\bar{D}$ for a given enforcement rate ϵ

the non-monotonicity noted in the introduction. Second, we show that the optimal credit limit decreases with ξ, and it collapses to zero for higher ξ's even though a pure credit equilibrium is implementable.

We first define welfare. A stationary allocation in our economy consists of only the enforcement rate, ϵ, and the level of DM trade, y. Given the allocation (ϵ, y), the corresponding welfare is

$$\mathcal{W}(\epsilon, y) = \sigma[u(y) - y] - \xi\epsilon. \tag{1.13}$$

In the following analysis we also maintain budget balance according to (1.11).

1.4.1 Social welfare in credit equilibrium

To discuss welfare in arrangements where credit is used, we first consider the alternative arrangement where credit is not used at all. If the social planner decides not to use credit, it is then optimal to choose $\epsilon = 0$. In that case, welfare is given by

$$\mathcal{W}^m = \sigma[u(\bar{z}) - \bar{z}], \tag{1.14}$$

where $\bar{z}$ solves (1.7), the equilibrium real balance holding in a pure monetary equilibrium where credit is not used. Since the pure monetary equilibrium is implementable for any ξ by setting $\epsilon = 0$, $\mathcal{W}^m$ serves as a lower bound to the optimal welfare. Moreover, it will not be optimal to use credit unless it can achieve a higher welfare than $\mathcal{W}^m$. Finally, since a credit equilibrium is implementable with $D \geqslant \tilde{D}$ only if $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, we only need to consider that range.

For $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, from Proposition 1.3.1, a pair (ϵ, D) is incentive compatible if and only if $\epsilon \leqslant \bar{\epsilon}(\xi)$ and $D \in [\tilde{D}(\epsilon;\xi), \bar{D}(\epsilon;\xi)]$. Given an implementable pair, the equilibrium

DM trade is given by $y^c = y^c(z+D)$ with z the solution to (1.8); that is, $y^c = \max\{\bar{z}, D\}$. Thus, to maximize welfare among credit equilibria, the problem becomes

$$\begin{aligned} \max_{(D,\epsilon)} \mathcal{W}^c(D,\epsilon) \equiv &\sigma\left[u\left(\max\{\bar{z}, D\}\right) - \max\{\bar{z}, D\}\right] - \epsilon\xi, \\ &\text{s.t. } \tilde{D}(\epsilon;\xi) \leqslant D \leqslant \bar{D}(\epsilon;\xi). \end{aligned} \tag{1.15}$$

Now, since y^c increases in D and the welfare increases in y^c (note that D and hence y^c is always below or equal to y^*), it is optimal to choose $D = \bar{D}$ for any given ϵ. As a result, we can reduce the problem to the choice of ϵ. However, Proposition 1.3.1 (2.a) implies that there is a one-to-one relationship between $\bar{D}$ and ϵ as long as $\bar{D} < y^*$. Now, once $\bar{D}$ reaches y^*, a higher ϵ does not increase the first term in (1.15) but only decreases the total welfare by increasing the cost in the second term. Thus we may restrict our choice of ϵ so that there is always one-for-one correspondence between $\bar{D}$ and ϵ (so, if $\bar{D} = y^*$ is considered, we just choose the lowest ϵ so that $\bar{D}(\epsilon;\xi) = y^*$).

Now let $\bar{\bar{D}}(\xi) \equiv \bar{D}(\bar{\epsilon}(\xi);\xi)$, the highest $\bar{D}$ implementable under the marginal cost of enforcement ξ. We can then define the inverse of $\bar{D}(\cdot;\xi)$ for the given ξ, and take $\epsilon(D;\xi)$ to be the smallest ϵ so that $\bar{D}(\epsilon;\xi) = D$. Note that the function $\epsilon(D;\xi)$ in $[0, \bar{\bar{D}}]$ is well-defined and continuous as $\bar{D}$ is strictly increasing up to $\bar{D} = y^*$ by Proposition 1.3.1 (2.a) and (2.b). To solve (1.15), it is then further reduced to

$$\max_{D\in[0,\bar{\bar{D}}(\xi)]} \mathcal{W}^c[D, \epsilon(D;\xi)]. \tag{1.16}$$

In this maximization problem, any choice of D is implicitly already the optimal one for the enforcement level $\epsilon(D;\xi)$. Since $\mathcal{W}^c(D,\epsilon)$ is increasing in D and decreasing in ϵ, equation (1.16) highlights the essential trade-off involved in the optimal credit limit when the level of enforcement is endogenous: a higher D increases the trade surplus in the first term in (1.15), but it also requires a higher ϵ that increases the cost in the second term. The next proposition shows that this is a non-trivial trade-off.

Proposition 1.4.1. *Let $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$ be given.*

1. *For any $D < \bar{z}$, $\mathcal{W}^c[D, \epsilon(D;\xi)]$ strictly decreases with D.*

2. *Suppose that $u(y) = \frac{y^\alpha}{\alpha}$ with $\alpha \in (0,1)$. If $r < \sigma\frac{1-\alpha}{\alpha}$, then $\mathcal{W}^c[D, \epsilon(D;\xi)]$ first increases with D but then decreases with D for $D \in [\bar{z}, \bar{\bar{D}}]$.*

Proposition 1.4.1 shows a non-monotonic relationship between credit limits and social welfare. According to Proposition 1.4.1 (1), it is never optimal to choose a $D < \bar{z}$. Indeed, for implementable D in that range, in equilibrium the buyer uses both money and credit, but equilibrium DM trade is not affected by the credit limit and stays at $\bar{z}$ as D increases. However, higher D still requires higher ϵ and hence welfare strictly decreases with D. For higher value of D, Proposition 1.4.1 (2) gives a sufficient condition for the welfare to be non-monotonic in D. Indeed, a straightforward differentiation yields

$$\frac{\mathrm{d}\mathcal{W}^c}{\mathrm{d}D} = \sigma[u'(D) - 1] - \xi\frac{\mathrm{d}\epsilon(D;\xi)}{\mathrm{d}D}. \tag{1.17}$$

Since the first term decreases with D because of the concavity of u, when D is relatively high, the second term dominates the first. The condition that r is relatively small ensures that this range exists.

We demonstrate this reserve U-shaped effect of increase in the credit limit in Figure (1.4), where $r = 0.08$, $\sigma = 0.25$, $\alpha = 0.75$ and $\xi = 0.005$. Thus the first-best trade outcome is $y^* = 1$ with the first-best social welfare $\mathcal{W}^* = 0.0833$, and the DM trade in a pure monetary equilibrium is $\bar{z} = 0.3293$ with welfare $\mathcal{W}^m = 0.0625$. In a pure credit equilibrium the highest implementable pair is $(D, \epsilon) = (0.958, 1)$. The curve represents the social welfare that obtains when credit is used. In the coexistence equilibrium, social welfare strictly decreases with D, as in Proposition (1.4.1)(1). In the pure credit equilibrium, social welfare first increases and then decreases with D, and reaches the highest value when $(D, \epsilon) = (0.9208, 0.9596)$. The highest social welfare is 0.0783, accounting for 94.03% of the first-best social welfare, and improves 25.28% from a pure monetary social welfare.

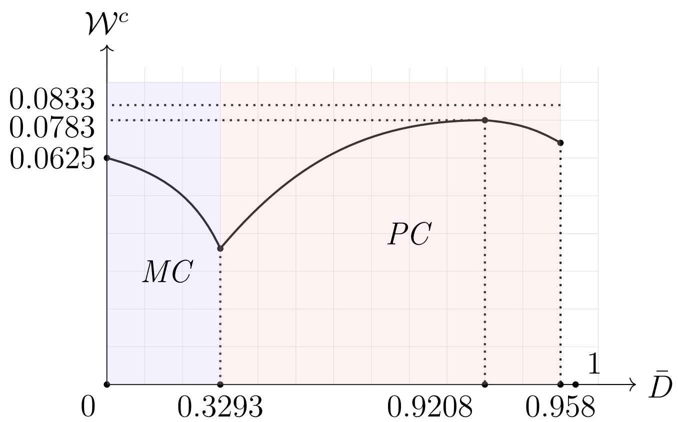

Figure 1.4: The social welfare for given highest implementable credit limit

1.4.2 Cost of enforcement and optimal policy

Now we turn to the optimal arrangement, and examine how changes in the marginal cost of enforcement, ξ, affect the optimal means-of-payments. First we give a proposition to characterize the optimal means-of-payment to be used.

Proposition 1.4.2. *Let $\xi < \frac{\sigma}{\sigma+r}\Psi^m$ be given. Denote the optimal welfare in a credit equilibrium by $\mathcal{W}^c(\xi)$, then there exists $\hat{\xi} < \frac{\sigma}{\sigma+r}\Psi^m$ such that*

$$\mathcal{W}^c(\xi) \geqslant \mathcal{W}^m \text{ if and only if } \xi \in [0, \hat{\xi}]. \tag{1.18}$$

Hence, it is optimal to implement a pure credit equilibrium for $\xi \in [0, \hat{\xi}]$ and to implement a pure monetary equilibrium for $\xi > \hat{\xi}$. Moreover, $\mathcal{W}^c(\xi)$ strictly deceases with ξ.

Proposition 1.4.2 shows that the solution to the optimal welfare under credit arrangement always exists. Moreover, it shows that there is a threshold for the marginal cost of enforcement, $\hat{\xi}$, below which a pure credit equilibrium is optimal and above which a pure monetary equilibrium is optimal. Note that above the threshold $\hat{\xi}$ a pure credit equilibrium is implementable but dominated by a pure monetary equilibrium. The intuition for this result is simple. When ξ is close to the threshold $\frac{\sigma}{\sigma+r}\Psi^m$, although a pure credit equilibrium is implementable, the maximal credit limit D is close to $\bar{z}$ and hence the trading surplus in credit trade is not significantly than the monetary trade in a pure monetary equilibrium. However, the credit economy requires a cost of enforcement that is not needed in a monetary equilibrium, and hence welfare in the credit economy must be lower. Our next result states that for ξ of relatively small value, the optimal credit limit decreases with the marginal cost of enforcement.

Proposition 1.4.3. *There exists a threshold $\tilde{\xi} \in (0, \hat{\xi}]$ such that for all $\xi < \tilde{\xi}$, the optimal credit limit, denoted by $D^* = D^*(\xi)$, decreases with ξ. If $u(y) = \frac{y^\alpha}{\alpha}$ with $\alpha \in (0,1)$, then $\tilde{\xi} = \hat{\xi}$ and D^* strictly decreases with ξ for all $\xi \in [0, \hat{\xi}]$.*

Proposition 1.4.3 shows that the optimal credit limit decreases with ξ, at least for a range of small ξ's; and, if u is CRRA with $u(0) = 0$, then this holds for all ξ's under which it is optimal to use credit. In Figure 1.5 we depict the optimal credit limit, D^*, and the optimal social welfare, $\mathcal{W}$, both as functions of ξ. As can be seen from the right panel, around $\xi = \hat{\xi}$ there is a kink to the optimal welfare, which results from the discontinuity in the amount of liquidity in the economy. Indeed, the left panel shows that for any $\xi < \hat{\xi}$, the optimal credit limit is bounded away from $\bar{z}$, but that for $\xi > \hat{\xi}$ the real balance in the pure currency economy is constant at $\bar{z}$. As a result, the output level in the DM will have a discontinuous increase when the marginal cost of enforcement decreases below $\hat{\xi}$. Note that this last result does not depend on the functional form of the utility function u, as $D^*(\xi) > \bar{z}$ for all $\xi < \hat{\xi}$.

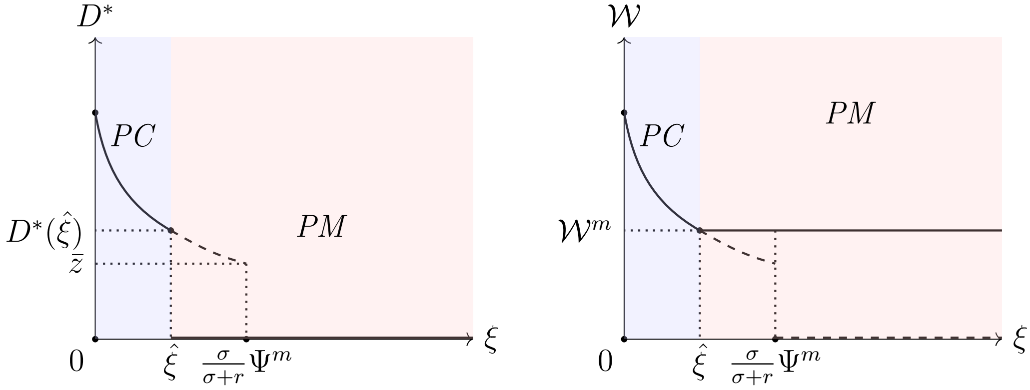

Figure 1.5: Optimal credit limit and social welfare

1.4.3 Convex cost of enforcement

We remark here that while we have assumed a linear enforcement cost, our results can be extended to a convex cost case, $\xi v(\epsilon)$ with $v'(\cdot) > 0$, $v''(\cdot) > 0$, $v(0) = 0$, $v'(0) = 0$ and $v(1) = 1$. The analysis then follows the same path as the linear cost case, and we highlight only a few key results. We will show that the monotonic relationship between the highest implementable credit limit and enforcement rate holds only when the cost parameter ξ is sufficiently small, and, in this range, the welfare in pure credit equilibrium decreases with ξ. Interestingly, with convex cost, there can be a non-monotonic relationship between the cost parameter and the optimal enforcement level.

With government budget $\sigma\chi = \xi v(\epsilon)$, the incentive constraint, as equation (1.10), is given by

$$r(D + \frac{\xi v(\epsilon)}{\sigma}) \leqslant \epsilon \Psi^c(D), \tag{1.19}$$

and the next proposition gives a sufficient condition for implementability:

Proposition 1.4.4. *When* $\xi \leqslant \min\{\frac{1}{v'(1)}\frac{\sigma}{\sigma+r}\Psi^m, r\bar{z}\}$, *for any* $\epsilon \in [0,1]$, $D \in [\tilde{D}(\epsilon;\xi), \bar{D}(\epsilon;\xi)]$ *is implementable and is not empty, with* $\frac{\mathrm{d}\bar{D}(\epsilon;\xi)}{\mathrm{d}\epsilon} \geqslant 0$.

Note that the influence of enforcement rate on the implementable credit limit depends on the magnitude of three opposite effects: on the right side of equation (1.19), a higher

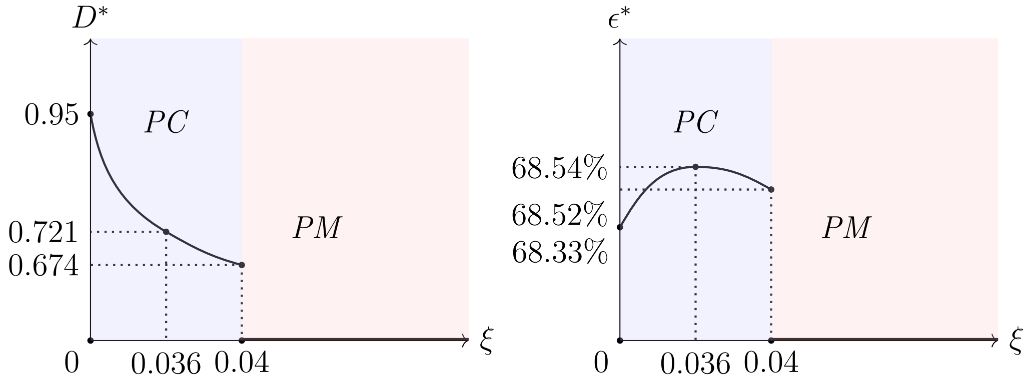

Figure 1.6: Optimal credit limit and enforcement rate

enforcement rate increases the probability of being caught, which extensively increases the cost of default. But the continuation value of accessing credit decreases as credit is more expensive, which affects the intensive margin of default. On the left side, a higher enforcement rate also increases the benefit from default, which in turn discourages buyers from using credit. In Proposition 1.4.4, we show that only when enforcement is not overly costly (ξ is sufficiently small or $v'(\epsilon)$ is low), does better enforcement lead to a higher implementable credit limit.

Given ξ is sufficiently small, the optimal welfare and credit limit satisfies:

Proposition 1.4.5. *Let $\xi \leqslant \min\{\frac{1}{v'(1)}\frac{\sigma}{\sigma+r}\Psi^m, r\bar{z}\}$, the optimal welfare in the pure credit equilibrium strictly decreases with ξ: $\frac{d\mathcal{W}^c(\xi)}{d\xi} < 0$ with $\mathcal{W}^c(0) > \mathcal{W}^m$. Moreover, optimal credit limit $D^*(\xi)$ decreases with ξ for sufficiently small value of ξ.*

The proof of Proposition 1.4.5 is same as for Proposition 1.4.1 and 1.4.3. Interestingly, with convex cost, the optimal credit limit still decreases with the marginal cost when it is small, but there can be a non-monotonic relationship between the cost parameter ξ and the optimal enforcement level: $\epsilon^* = \epsilon(D^*(\xi); \xi)$:

$$\frac{d\epsilon(D^*(\xi);\xi)}{d\xi} = \frac{d\epsilon}{dD^*}\frac{dD^*}{d\xi} + \frac{d\epsilon}{d\xi}, \tag{1.20}$$

where the first term is negative and captures an indirect effect of ξ on ϵ^*: when ξ increases, the optimal credit limit decreases, so the enforcement rate needed to implement a lower credit limit also decreases. However, the second term is positive and captures a direct effect: as ξ increases, credit fees increase and agents are less likely to use credit. So for a given credit limit, a higher enforcement rate is required to implement it.

Therefore, the relationship between ϵ^* and ξ, and also ϵ^* and D^* can be non-monotonic in some cases. In Figure 1.6 we depict such an example. Here we have $u(y) = \frac{(y+b)^\alpha - b^\alpha}{\alpha}$

with $\alpha = 0.25, \sigma = 0.18, r = 0.15, b = 0.005$, and the cost of enforcement is given by $\xi\epsilon^m$ with $m = 1.27$. When $0 \leqslant \xi < 0.036$, the optimal credit limit increases with the enforcement rate. But when $0.036 \leqslant \xi < 0.04$, the optimal credit limit decreases with the enforcement rate. When $\xi \geqslant 0.04$, the social welfare in a credit equilibrium is lower then the welfare in a pure monetary equilibrium. When $\xi = 0$, the optimal enforcement rate is 68.33%, and it increases to 68.54%, and then decreases to 68.52%, as ξ increases to 0.036 and 0.04.

1.4.4 Monetary policy

We show here that our main results hold when the money supply grows at a constant rate $\gamma \geqslant 1$ and $M_{t+1} = \gamma M_t$. Note that the underlying friction that the government has no coercion power to tax agents implies that any deflationary monetary policy is not feasible. The monetary policy sets the lower bound of welfare that can be achieved in the economy, but it does not influence either the implementability or the optimal policy when a non-monetary equilibrium is optimal, which is when the cost parameter ξ is sufficiently low. The next proposition gives a sufficient condition such that the main results hold, where i denotes the nominal interest rate with $i = \gamma(1+r) - 1 \geqslant r$, and $\bar{z}(i)$ and $\Psi^m(i)$ denote the equilibrium real balances and value of using money given monetary policy i respectively:

Proposition 1.4.6. *When* $\xi \leqslant \min\{\frac{\sigma}{\sigma+r}\Psi^m(i), i\bar{z}(i)\}$, *for any* $\epsilon \in [0,1]$, $D \in [\tilde{D}(\epsilon;\xi,i), \bar{D}(\epsilon;\xi)]$ *is implementable and is not empty, with* $\frac{d\bar{D}(\epsilon;\xi)}{d\epsilon} > 0$. *The optimal welfare in the pure credit equilibrium and optimal credit limit are* $\mathcal{W}^c(\xi)$ *and* $D^*(\xi)$, *and are independent with* i.

Note that monetary policy i does not affect the upper bound of the implementable credit limit $\bar{D}(\epsilon;\xi)$ in a pure credit equilibrium, and therefore does not affect the results in the welfare analysis when a pure credit equilibrium is optimal. The monetary policy i does, however, influence the implementable credit limit in a coexistence equilibrium, but this equilibrium is never a socially optimal choice. Moreover, the monetary policy i also influences the threshold of credit limit below which the buyer prefers money, even though credit is provided, $\tilde{D}(\epsilon;\xi,i)$. However in the welfare analysis, we only consider the highest implementable credit limit that would maximize social welfare.

1.5 Concluding Remarks

We have developed a model of endogenous use of unsecured credit and money, but under costly enforcement. We obtained three main results. First, the use of credit can be sustained in an equilibrium that is both incentive compatible and fiscally feasible only if the marginal cost of enforcement is not overly high. Second, even when it is sustainable, a high enforcement rate may not necessarily be the optimal one, consistent with the empirical finding of the reverse U-shaped relationship between financial development and economic development; moreover, for a range of marginal cost of enforcement, it is optimal to use money alone while it is incentive compatible and fiscally feasible to sustain the use of credit. Third, there can be a non-monotonic relationship between the optimal credit limit and enforcement levels. These results suggest that looking at the debt-GDP ratio alone or at indexes for institution qualities, which are typically proxies for efficiency of enforcement, as the main guidance for policy recommendations on different use of means-of-payments, can be misleading.

Although the optimal arrangement in our model is either a pure-currency economy or a pure-credit economy, other features could potentially be included in our baseline model to obtain coexistence, such as a two-stage DM structure, as in Araujo and Hu (2018), with different costs of using the monitoring technology across meetings between the two DM stages. Our results can be useful as they point out the basic trade-offs, both for individuals and for the society as a whole.

1.6 Appendix: Proofs of Lemmas and Propositions

Proof of Lemma 1.3.2

We prove Lemma 1.3.2 by three steps:

1. There exists $\underline{z} \geqslant 0$ such that, if $\underline{z} > 0$, then $V^m(z) > V^c(z)$ if and only if $z > \underline{z}$; if $\underline{z} = 0$, then $V^m(z) \geqslant V^c(z)$ for all z with strict inequality for all $z > 0$.
2. $\Psi^c(D) \geqslant \Psi^m$ if and only if $\chi \leqslant \bar{\chi}$ and $D \geqslant \tilde{D}$, where $\tilde{D}$ is the unique solution to $\Psi^c(D) = \Psi^m$.
3. The buyer chooses to access the financial system if and only if $\Psi^c(D) \geqslant \Psi^m$.

(1) Let

$$f(z; D, \chi) = V^c(z) - V^m(z) = \sigma[u(y^c(z, D) - y^c(z, D)] - \sigma\chi - \sigma[u(y^m(z)) - y^m(z)]$$

denote the difference between the buyer's DM continuation value when either accessing the financial system or not. It then follows that $f(z; D, \chi)$ strictly decreases with z for $z \leqslant y^*$, with $f(0; D, \chi) = \sigma[u(D) - D] - \sigma\chi$ and $f(y^*; D, \chi) = -\sigma\chi < 0$. We consider two cases. First, suppose that $\chi \leqslant u(D) - D$. Then, there exists a unique $\underline{z}(D, \chi) \in [0, y^*]$ such that $f(\underline{z}; D, \chi) = 0$. Second, suppose that $\chi > u(D) - D$. Then, let $\underline{z} = 0$. It follows that $V^m(z) > V^c(z)$ for all $z > \underline{z}(D, \chi)$, and, for the first case, $V^m(z) < V^c(z)$ for all $z < \underline{z}$. Moreover, $\underline{z}(D, \chi)$ increases in D and decreases in χ as f increases in D and decreases in χ.
(2) From equation (1.8), $\Psi^c(D) = -r(\bar{z} - D) + \sigma[u(\bar{z}) - \bar{z}]$ if $D < \bar{z}$, and $\Psi^c(D) = \sigma[u(D) - D]$ if $\bar{z} \leqslant D \leqslant y^*$. Hence, $\Psi^c(D)$ strictly increases with D for all $0 \leqslant D \leqslant y^*$. From equation (1.7), $\Psi^m = -r\bar{z} + \sigma[u(\bar{z}) - \bar{z}]$, which does not depend on D. Therefore, $\Psi^c(D) - \Psi^m$ strictly increases with D, with the maximum value $\Psi^c(y^*) - \Psi^m = \sigma[u(y^*) - y^*] - \sigma\chi - \Psi^m$ and the minimum value $\Psi^c(0) - \Psi^m = -\sigma\chi < 0$.

By the Intermediate Value Theorem, when $\Psi^c(y^*) - \Psi^m \geqslant 0$, or equivalently

$$\chi \leqslant \frac{\sigma[u(y^*) - y^*] - \Psi^m}{\sigma} \equiv \bar{\chi},$$

there exists a unique $\tilde{D} \in [0, y^*]$ that solves $\Psi^c(D) = \Psi^m$, and $\Psi^c(D) \geqslant \Psi^m$ if and only if $D \geqslant \tilde{D}$. Otherwise, $\Psi^c(D) < \Psi^m$ for any $0 \leqslant D \leqslant y^*$.
(3) Combining equation (1.3) and (1.6), we can rewrite the buyer's CM decision as

$$\max_z \{-rz + \sigma \max\{V^m(z), V^c(z)\}\}. \tag{1.21}$$

We show that problem (1.21) has the same solution as

$$\max\{\Psi^m, \Psi^c(D)\} = \max\{\max_z\{-rz + \sigma V^m(z)\}, \max_z\{-rz + \sigma V^c(z)\}\}, \tag{1.22}$$

by considering three cases.
(3.a) $D < \tilde{D}$. In this case the solution to (1.22) is such that $\Psi^m > \Psi^c(D)$ and $z = \bar{z}$. Note that $\bar{z} > \underline{z}(D, \chi)$ because

$$f(\bar{z}; D, \chi) = -r\bar{z} + \sigma[u(y^c(\bar{z}, D)) - y^c(\bar{z}, D)] - \sigma\chi - \Psi^m < \Psi^c(D) - \Psi^m < 0.$$

The solution to (1.21) is also $z = \bar{z}$: if $z > \underline{z}(D, \chi)$, then $V^m(z) > V^c(z)$ and $\max\{-rz + \sigma V^m(z)\} = \Psi^m$, which has solution $\bar{z} > \underline{z}(D, \chi)$; if $z \leqslant \underline{z}(D, \chi)$, then $V^m(z) \leqslant V^c(z)$ and $\max\{-rz + \sigma V^c(z)\} \leqslant \Psi^c(D) < \Psi^m$. Therefore $\bar{z}$ is also the unique solution of (1.21).
(3.b) $D \in [\tilde{D}, \max\{\bar{z}, \tilde{D}\})$. Then, the solution to (1.22) is such that $\Psi^m \leqslant \Psi^c(D)$ and $z = \bar{z} - D$. Note that $\bar{z} - D \leqslant \underline{z}(D, \chi)$ because

$$f(\bar{z} - D; D, \chi) = \Psi^c(D) - [-r(\bar{z} - D) + \sigma(u(\bar{z} - D) - (\bar{z} - D)] \geqslant \Psi^c(D) - \Psi^m > 0.$$

The solution to (1.21) is also $z = \bar{z} - D$: if $z \leqslant \underline{z}(D, \chi)$, then $V^c(z) \geqslant V^m(z)$, and $\max\{-rz + \sigma V^c(z)\} \leqslant \Psi^c(D)$ which has solution $z = \bar{z} - D \leqslant \underline{z}(D, \chi)$; if $z > \underline{z}(D, \chi)$, then $V^c(z) < V^m(z)$, and $\max\{-rz + \sigma V^m(z)\} = \Psi^m \leqslant \Psi^c(D)$. So the unique solution to (1.21) is also $\bar{z} - D$.
(3.c) $D \geqslant \max\{\bar{z}, \tilde{D}\}$. The solution to (1.22) is such that $\Psi^m \leqslant \Psi^c(D)$ and $z = 0$. The solution to (1.21) is also $z = 0$: if $z \leqslant \underline{z}(D, \chi)$, then $V^m(z) \leqslant V^c(z)$ and $\max\{-rz + \sigma V^c(z)\} \leqslant \Psi^c(D)$ which has solution $z = 0 \leqslant \underline{z}(D, \chi)$; if $z > \underline{z}(D, \chi)$, then $V^m(z) \geqslant V^c(z)$ and $\max\{-rz + \sigma V^m(z)\} = \Psi^m \leqslant \Psi^c(D)$. So the unique solution to equation (1.21) is the unique solution to equation (1.22).

Proof of Lemma 1.3.1

When $z = 0$, the consistency between the solution of $\max\{V^m(z), V^c(z)\}$ and $\max\{\Psi^m, \Psi^c(D)\}$ when $z = 0$ follows the same proof as in the proof of Lemma 1.3.2. As $\Psi^c(D) - \Psi^m = \sigma[u(D) - D - \chi]$ is a concave and increasing function on $D \in [0, y^*]$ with the minimum value $-\sigma\chi < 0$ and the maximum value $\sigma[u(y^*) - y^* - \chi]$, by the Intermediate Value Theorem, when $\chi \leqslant u(y^*) - y^* \equiv \hat{\chi}$, there exists a unique $\hat{D}(\chi) \in [0, y^*]$ which solves $\Psi^c(\hat{D}) - \Psi^m = 0$, with $\hat{D}'(\chi) = 1/(u'(\hat{D}) - 1) > 0$, $\hat{D}''(\chi) = -u''(\hat{D})/(u'(\hat{D}) - 1)^2 > 0$.

Proof of Lemma 1.3.3

1. When $D \in [\tilde{D}, \bar{z})$, the DM output in the coexistence equilibrium is $y = \bar{z}$, and in the pure credit equilibrium is $y = D$. As $D < \bar{z}$, $\sigma[u(\bar{z}) - \bar{z} - \chi] > \sigma[u(D) - D - \chi]$.

2. When $D \in [\hat{D}, \tilde{D})$, the DM output in the pure monetary equilibrium is $y = \bar{z}$ and in the pure credit equilibrium is $y = D$. As $D < \tilde{D}$, from Lemma 1.3.2, $\sigma[u(D) - D - \chi] = \Psi^c(D) < \Psi^m = -r\bar{z} + \sigma[u(\bar{z}) - \bar{z}] < \sigma[u(\bar{z}) - \bar{z}]$.

Proof of Lemma 1.3.4

Combine equation (1.10) and (1.11), we get equation (1.12) in Lemma 1.3.4. Rewrite equation (1.12) by a function $g(D; \epsilon, \xi)$ as
(1.23)

$$g(D; \epsilon, \xi) \equiv \Psi^c(D) - \frac{r}{\epsilon}D - \frac{r\xi}{\sigma} = \begin{cases} \Psi^m + rD - \xi\epsilon - \frac{r}{\epsilon}D - \frac{r\xi}{\sigma} & \text{if} \quad 0 \leqslant D < \bar{z}; \\ \sigma(u(D) - D) - \xi\epsilon - \frac{r}{\epsilon}D - \frac{r\xi}{\sigma} & \text{if} \quad \bar{z} \leqslant D < y^*; \\ \sigma[u(y^*) - y^*] - \xi\epsilon - \frac{r}{\epsilon}D - \frac{r\xi}{\sigma} & \text{if} \quad D \geqslant y^*, \end{cases}$$

so the incentive constraint is equivalent as $g(D; \epsilon, \xi) \geqslant 0$. Notice that $g(D; \epsilon, \xi)$ is a decreasing function of D for any (ϵ, c), with the maximum value $g(0; \epsilon, \xi) = \Psi^m - \xi\epsilon - \frac{r\xi}{\sigma}$ and the minimum value $g(\infty; \epsilon, \xi) = -\infty$. So there exists a unique $\bar{D} = \bar{D}(\epsilon; \xi)$ such that $g(\bar{D}; \epsilon, \xi) = 0$ if $\Psi^m - \xi\epsilon - \frac{r\xi}{\sigma} \geqslant 0$; otherwise let $\bar{D} < 0$. Then we have $g(D; \epsilon, \xi) \geqslant 0$ for all $D \leqslant \bar{D}$.

Proof of Proposition 1.3.1

We prove in three steps:

1. For $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, there exists $\bar{\epsilon}(\xi)$ such that $\tilde{D} \leqslant \bar{D}$ if and only if $\epsilon \in [0, \bar{\epsilon}(\xi)]$.
2. For any $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$ and any $D \leqslant \bar{\bar{D}} \equiv \bar{D}(\bar{\epsilon}(\xi); \xi)$, there is a well-defined and continuously differentiable function $\epsilon(D; \xi)$, as the inverse of $\bar{D}(\epsilon; \xi)$, such that D is implementable if and only if $\epsilon(D; \xi) \leqslant \epsilon \leqslant \bar{\epsilon}(\xi)$.
3. Properties of $\bar{D}(\epsilon; \xi)$ in Proposition 1.3.1(2.a) and (2.b).

(1) From equation (1.23), $\tilde{D} \leqslant \bar{D}$ if and only if $g(\tilde{D}; \epsilon, \xi) \geqslant g(\bar{D}; \epsilon, \xi) = 0$. So we only need to show that there exists $\bar{\epsilon}(\xi)$ such that $g(\tilde{D}; \epsilon, \xi) \geqslant 0$ if and only if $\epsilon \in [0, \bar{\epsilon}(\xi)]$ for $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$. For $\xi > \frac{\sigma}{\sigma+r}\Psi^m$, we show that $g(\tilde{D}; \epsilon, \xi) < 0$ for all $\epsilon \in [0, 1]$.

First note that $\tilde{D}$ increases in both ξ and ϵ. To see this, let $\Psi^m = \Psi^c(\tilde{D})$ as in equation (1.7) and (1.8): if $\xi\epsilon \leqslant r\bar{z}$, $\tilde{D} = \frac{\xi\epsilon}{r}$ and $\tilde{D}$ increases with ϵ and ξ. If $\xi\epsilon > r\bar{z}$, $\tilde{D}$ is determined by $\sigma[u(\tilde{D}) - \tilde{D}] = \Psi^m + \xi\epsilon$, and $\tilde{D}$ increases with ϵ and ξ because $\sigma[u(D) - D]$ increases with D. Moreover, when $\xi\epsilon > r\bar{z}$, $\frac{\tilde{D}(\epsilon;\xi)}{\epsilon}$ is also increasing in ϵ because

$$\frac{\partial \frac{\tilde{D}(\epsilon;\xi)}{\epsilon}}{\partial \epsilon} = \frac{\sigma[u(\tilde{D}) - \tilde{D}] - \Psi^m - \sigma\tilde{D}(u'(\tilde{D}) - 1)}{\epsilon^2\sigma[u'(\tilde{D}) - 1]} \geqslant \frac{\sigma[u(\bar{z}) - \bar{z}] - \Psi^m - r\bar{z}}{\epsilon^2\sigma[u'(\tilde{D}) - 1]} = 0, \tag{1.24}$$

where the inequality uses the fact that $\tilde{D} \geqslant \bar{z}$ and $\sigma[u(D) - D] - \Psi^m - \sigma D(u'(D) - 1)$ is an increasing function of D.

Plug $\tilde{D}$ into $g(D; \epsilon, \xi)$,

$$g(\tilde{D}; \epsilon, \xi) = \begin{cases} \Psi^m - \frac{\sigma+r}{\sigma}\xi & \text{if } \xi\epsilon \leqslant r\bar{z}; \\ \Psi^m - \frac{r}{\epsilon}\tilde{D}(\epsilon;\xi) - \frac{r\xi}{\sigma} & \text{if } \xi\epsilon > r\bar{z}, \end{cases} \tag{1.25}$$

where in both cases $g(\tilde{D}; \epsilon, \xi)$ decreases in ξ and decreases in ϵ. Moreover, if $\xi\epsilon > r\bar{z}$, $g(\tilde{D}; \epsilon, \xi) < g(\tilde{D}; \frac{r\bar{z}}{\xi}, \xi) = \Psi^m - \frac{\sigma+r}{\sigma}\xi$. So for any $\epsilon \in [0, 1]$, $g(\tilde{D}; \epsilon, \xi) \leqslant \Psi^m - \frac{\sigma+r}{\sigma}\xi$.

When $\xi > \frac{\sigma}{\sigma+r}\Psi^m$, it is straightforward to verify that $g(\tilde{D}; \epsilon, \xi) \leqslant \Psi^m - \frac{\sigma+r}{\sigma}\xi < 0$, so there is no implementable ϵ and $\bar{\epsilon}(\xi)$ is empty. Now we discuss $\bar{\epsilon}(\xi)$ when $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$ in two cases:

(1.a) $r\bar{z} \geqslant \frac{\sigma}{\sigma+r}\Psi^m$. Given that $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, $\xi\epsilon \leqslant \frac{\sigma}{\sigma+r}\Psi^m \leqslant r\bar{z}$, and hence $g(\tilde{D}; \epsilon, \xi) = \Psi^m - \frac{\sigma+r}{\sigma}\xi \geqslant 0$ for any $\epsilon \in [0, 1]$. Thus $\bar{\epsilon}(\xi) = 1$.

(1.b) $r\bar{z} < \frac{\sigma}{\sigma+r}\Psi^m$. We consider two subcases. If $\xi < r\bar{z}$, $\xi\epsilon < r\bar{z}$ for all $\epsilon \in [0, 1]$, and hence $g(\tilde{D}; \epsilon, \xi) = \Psi^m - \frac{\sigma+r}{\sigma}\xi > \Psi^m - \frac{\sigma+r}{\sigma}r\bar{z} > \Psi^m - \frac{\sigma+r}{\sigma}\frac{\sigma}{\sigma+r}\Psi^m = 0$. Thus, $\bar{\epsilon}(\xi) = 1$. If $r\bar{z} \leqslant \xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, notice that when $\epsilon = 1$, $g(\tilde{D}; 1, \xi)$ decreases in ξ, with the maximum value $g(\tilde{D}; 1, r\bar{z}) = \Psi^m - r\bar{z} - \frac{r}{\sigma}r\bar{z} \geqslant 0$ and the minimum value $g(\tilde{D}; 1, \frac{\sigma}{\sigma+r}\Psi^m) = \Psi^m - r\tilde{D} - \frac{r}{\sigma}\Psi^m = -r\tilde{D} + \sigma[u(\tilde{D}) - \tilde{D}] - \Psi^m \leqslant 0$. So there exists a unique threshold $\dot{\xi} \in [r\bar{z}, \frac{\sigma}{\sigma+r}\Psi^m]$, determined by $g(\tilde{D}; 1, \dot{\xi}) = 0$, such that $g(\tilde{D}; 1, \xi) \geqslant 0$ if $\xi \leqslant \dot{\xi}$, and $g(\tilde{D}; 1, \xi) < 0$ if $\xi > \dot{\xi}$. Then when $r\bar{z} \leqslant \xi \leqslant \dot{\xi}$, if $0 \leqslant \epsilon < \frac{r\bar{z}}{\xi}$, $g(\tilde{D}; \epsilon, \xi) = \Psi^m - \frac{\sigma+r}{\sigma}\xi \geqslant 0$, and if $\frac{r\bar{z}}{\xi} \leqslant \epsilon \leqslant 1$, $g(\tilde{D}; \epsilon, \xi) \geqslant g(\tilde{D}; 1, \xi) \geqslant 0$. So in the range $r\bar{z} \leqslant \xi \leqslant \dot{\xi}$, $\bar{\epsilon}(\xi) = 1$, and all $\epsilon \in [0, 1]$ is implementable. When $\dot{\xi} < \xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, if $0 \leqslant \epsilon < \frac{r\bar{z}}{\xi}$, $g(\tilde{D}; \epsilon, \xi) = \Psi^m - \frac{\sigma+r}{\sigma}\xi \geqslant 0$. If $\frac{r\bar{z}}{\xi} \leqslant \epsilon \leqslant 1$, $g(\tilde{D}; \epsilon, \xi)$ decreases in ϵ, with the maximum value $g(\tilde{D}; \frac{r\bar{z}}{\xi}, \xi) = \Psi^m - \frac{\sigma+r}{\sigma}\xi \geqslant 0$, and the minimum value $g(\tilde{D}; 1, \xi) < 0$. So there exists a unique $\bar{\epsilon}(\xi) \in [\frac{r\bar{z}}{\xi}, 1)$ which solves $g(\tilde{D}; \bar{\epsilon}, \xi) = 0$, and $g(\tilde{D}; \epsilon, \xi) \geqslant 0$ for all $\epsilon \in [0, \bar{\epsilon}(\xi)]$. So in the range $\dot{\xi} < \xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, only $\epsilon \in [0, \bar{\epsilon}(\xi)]$ is implementable.

(2) First we show that $r\tilde{D}(\epsilon; \xi) \geqslant \xi\epsilon^2$ for any $\epsilon \in [0, \bar{\epsilon}]$. We consider two cases. If $\xi\epsilon < r\bar{z}$, $\tilde{D} = \frac{\xi\epsilon}{r}$ and $r\tilde{D} - \xi\epsilon^2 = \xi\epsilon(1 - \epsilon) \geqslant 0$. If $\xi\epsilon \geqslant r\bar{z}$, $\tilde{D}$ is determined by $\sigma[u(\tilde{D}) - \tilde{D}] - \xi\epsilon = \Psi^m$.

Thus $r\tilde{D} - \xi\epsilon^2 > r\tilde{D} - \xi\epsilon = r\tilde{D} - \sigma[u(\tilde{D}) - \tilde{D}] + \Psi^m \geqslant r\bar{z} - \sigma[u(\bar{z}) - \bar{z}] + \Psi^m = 0$. It then follows that for any $\epsilon \in [0, \bar{\epsilon}]$, $\frac{\partial g(D;\epsilon,\xi)}{\partial \epsilon} = -\xi + \frac{rD}{\epsilon^2} \geqslant -\xi + \frac{r\tilde{D}}{\epsilon^2} \geqslant 0$.

Given $\epsilon = 1$, first-best is implementable if $g(y^*; 1, \xi) \geqslant 0$, which has solution $\xi < \xi^* \equiv \frac{\sigma}{\sigma+r}[\sigma(u(y^*) - y^*) - ry^*]$. Note that $\xi^* \geqslant 0$ if and only if $r \leqslant \frac{\sigma(u(y^*)-y^*)}{y^*}$; otherwise if $r > \frac{\sigma(u(y^*)-y^*)}{y^*}$, y^* is never implementable even for $\xi = 0$. Given any $r \leqslant \frac{\sigma(u(y^*)-y^*)}{y^*}$ and $\xi < \xi^*$, y^* is implementable if $g(y^*; \epsilon, \xi) \geqslant 0$ which has solution $\epsilon \geqslant \epsilon^*(\xi)$. Moreover, $\epsilon^*(\xi)$ increases in ξ. We define

$$\bar{\bar{\epsilon}}(\xi) = \begin{cases} \epsilon^*(\xi) & \text{if } \xi \leqslant \xi^*; \\ \bar{\epsilon}(\xi) & \text{if } \xi^* < \xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m. \end{cases} \tag{1.26}$$

Note that $\bar{\bar{D}}(\xi) = \bar{D}(\bar{\bar{\epsilon}}(\xi); \xi) = \bar{D}(\bar{\epsilon}(\xi); \xi)$ as the highest implementable credit limit for given ξ. Thus,

$$\bar{\bar{D}}(\xi) = \begin{cases} y^* & \text{if } \xi \leqslant \xi^*; \\ \bar{D}(\bar{\epsilon}(\xi); \xi) < y^* & \text{if } \xi^* < \xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m. \end{cases} \tag{1.27}$$

As when $\epsilon = \bar{\bar{\epsilon}}$, $\bar{\bar{D}}$ is implementable, so $g(\bar{\bar{D}}; \bar{\bar{\epsilon}}, \xi) \geqslant 0$. For any $D \leqslant \bar{\bar{D}}(\xi)$, $g(D; \epsilon, \xi)$ strictly increases in ϵ for $\epsilon \leqslant \bar{\bar{\epsilon}}$, with the minimum value $g(D; 0, \xi) = -\infty$ and the maximum value $g(D; \bar{\bar{\epsilon}}, \xi) \geqslant g(\bar{\bar{D}}; \bar{\bar{\epsilon}}, \xi) \geqslant 0$. By the Intermediate Value Theorem, there exists a unique $\epsilon(D; \xi) \in [0, \bar{\bar{\epsilon}}]$ such that $g(D; \epsilon(D; \xi), \xi) = 0$. When $\epsilon \geqslant \epsilon(D; \xi)$, $g(D; \epsilon, \xi) \geqslant 0$. As for $D \in [0, \bar{\bar{D}}]$, $g(D; \epsilon, \xi)$ is a continuously differentiable function on ϵ with positive partial derivative w.r.t ϵ, hence $\epsilon(D; \xi)$ is continuously differentiable.
(3) From equation (1.23), we define the enforcement rate when $\bar{D} = \bar{z}$ as $\underline{\epsilon}(\xi)$, which is determined by $g(\bar{z}; \underline{\epsilon}(\xi), \xi) = \Psi^m + r\bar{z} - \xi\underline{\epsilon} - \frac{r\bar{z}}{\underline{\epsilon}} - \frac{r\xi}{\sigma} = 0$ with $\underline{\epsilon}'(\xi) = \frac{\underline{\epsilon}^2(r/\sigma+\underline{\epsilon})}{r\bar{z}-\xi\underline{\epsilon}^2} > 0$, as $r\bar{z} \geqslant r\tilde{D} \geqslant \xi\epsilon$. To compute the derivatives, we use the Implicit Function Theorem and solve it by taking first-order derivative of $g(\bar{D}(\epsilon; \xi); \epsilon, \xi) = 0$ with respect to ϵ, which implies

$$\left.\frac{d\bar{D}}{d\epsilon}\right|_{0 \leqslant \epsilon < \underline{\epsilon}} = \frac{\frac{r}{\epsilon}\bar{D} - \xi\epsilon}{r(1-\epsilon)} > 0 \text{ and } \left.\frac{d\bar{D}}{d\epsilon}\right|_{\underline{\epsilon} \leqslant \epsilon < \bar{\bar{\epsilon}}} = \frac{\frac{r}{\epsilon}\bar{D} - \xi\epsilon}{r - \epsilon\sigma[u'(\bar{D}) - 1]} > 0; \tag{1.28}$$

where the inequality comes from $\frac{r}{\epsilon}\bar{D} - \xi\epsilon > \frac{r}{\epsilon}\tilde{D} - \xi\epsilon$ and $r\tilde{D} \geqslant \xi\epsilon$. Similarly, we take first-order derivative of $g(\bar{D}(\epsilon; \xi); \epsilon, \xi) = 0$ with respect to ξ to obtain

$$\left.\frac{d\bar{D}}{d\xi}\right|_{0 \leqslant \epsilon < \underline{\epsilon}} = \frac{-\frac{r}{\sigma} - \epsilon}{r(1-\epsilon)} < 0; \text{ and } \left.\frac{d\bar{D}}{d\xi}\right|_{\underline{\epsilon} \leqslant \epsilon < \bar{\bar{\epsilon}}} = \frac{-\frac{r}{\sigma} - \epsilon}{r - \epsilon\sigma[u'(\bar{D}) - 1]} < 0. \tag{1.29}$$

Finally, when $\epsilon = 0$, it is obvious that $\bar{D}(0; \xi) = 0$. When $\epsilon = \bar{\epsilon}(\xi)$, given $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$, by some algebra we can show that $\underline{\epsilon}(\xi) \leqslant \underline{\epsilon}(\frac{\sigma}{\sigma+r}\Psi^m) = \bar{\epsilon}(\xi)$, so $\bar{D}(\bar{\epsilon}(\xi); \xi) \geqslant \bar{D}(\underline{\epsilon}(\xi); \xi) = \bar{z}$.

Proof of Proposition 1.4.1

(1): When $D < \bar{z}$, $y^c = \bar{z}$ and $\mathcal{W}^c = \sigma[u(\bar{z}) - \bar{z}] - \xi\epsilon(D;\xi)$ where $\epsilon(D;\xi)$ is a strictly increasing function in D, (see Proposition 1.3.1(2.a)). Thus $\frac{d\mathcal{W}^c}{dD} = -\xi\frac{d\epsilon}{dD} < 0$.
(2): When $D \geqslant \bar{z}$, $y^c = D$. Rewrite equation (1.17) as

$$\frac{d\mathcal{W}^c}{dD} = \frac{r}{rD - \xi\epsilon^2}[\sigma D(u'(D) - 1) - \xi\epsilon(D;\xi)], \tag{1.30}$$

where $rD - \xi\epsilon^2 > 0$ (see the proof of Proposition 1.3.1(2)), and the sign of $\frac{d\mathcal{W}^c}{dD}$ depends on $\sigma D(u'(D) - 1) - \xi\epsilon(D;\xi)$. Notice that since $u(D) = \frac{D^\alpha}{\alpha}$, $\sigma D(u'(D) - 1)$ strictly decreases in D and $\epsilon(D;\xi)$ strictly increases in D. So $\sigma D(u'(D) - 1) - \xi\epsilon(D;\xi)$ decreases in $D \in [\bar{z}, \bar{\bar{D}}(\xi)]$, with maximum value $\sigma\bar{z}(u'(\bar{z}) - 1) - \xi\epsilon(\bar{z};\xi) = r\bar{z} - \xi\epsilon(\bar{z};\xi) > 0$ (see the proof of Proposition 1.3.1(2)), and the minimum value $\sigma\bar{\bar{D}}(\xi)(u'(\bar{\bar{D}}(\xi)) - 1) - \xi\epsilon(\bar{\bar{D}}(\xi);\xi)$. Now we show that $\sigma\bar{\bar{D}}(\xi)(u'(\bar{\bar{D}}(\xi)) - 1) - \xi\epsilon(\bar{\bar{D}}(\xi);\xi) < 0$ in three subcases:
(2.a) $\xi \leqslant \xi^*$. From equation (1.27), $\bar{\bar{D}}(\xi) = y^*$, so $\sigma\bar{\bar{D}}(u'(\bar{\bar{D}}) - 1) - \xi\epsilon(\bar{\bar{D}};\xi) = -\xi\epsilon(y^*;\xi) < 0$.
(2.b) $\xi^* < \xi \leqslant \dot{\xi}$. From equations (1.26) and (1.27), $\bar{\bar{D}} < y^*$ and $\epsilon(\bar{\bar{D}}, \xi) = 1$. Together with the incentive constraint $\sigma[u(\bar{\bar{D}}) - \bar{\bar{D}}] - \xi = r\bar{\bar{D}} + \frac{r\xi}{\sigma}$, we rewrite $\sigma\bar{\bar{D}}(u'(\bar{\bar{D}}) - 1) - \xi\epsilon(\bar{\bar{D}};\xi) = \frac{\sigma}{\sigma+r}[(\sigma+r)\bar{\bar{D}}u'(\bar{\bar{D}}) - \sigma u(\bar{\bar{D}})]$, which decreases in $\bar{\bar{D}} \in [\bar{z}, y^*)$, so $\sigma\bar{\bar{D}}(u'(\bar{\bar{D}}) - 1) - \xi\epsilon(\bar{\bar{D}};\xi) \leqslant \frac{\sigma}{\sigma+r}[(\sigma+r)\bar{z}u'(\bar{z}) - \sigma u(\bar{z})] < 0$.
(2.c) $\dot{\xi} < \xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$. From the proof of Proposition 1.3.1(1.b), $\xi\epsilon(\bar{\bar{D}};\xi) > r\bar{z}$, so $\sigma\bar{\bar{D}}(u'(\bar{\bar{D}}) - 1) - \xi\epsilon(\bar{\bar{D}};\xi) < \sigma\bar{\bar{D}}(u'(\bar{\bar{D}}) - 1) - r\bar{z} < \sigma\bar{z}(u'(\bar{z}) - 1) - r\bar{z} = 0$.

Therefore, by the Intermediate Value Theorem, there exists $D \in [\bar{z}, \bar{\bar{D}}(\xi)]$, below which $\frac{d\mathcal{W}^c}{dD} > 0$, and above which $\frac{d\mathcal{W}^c}{dD} < 0$.

Proof of Proposition 1.4.2

By Proposition 1.4.1(1), $\mathcal{W}^c[D, \epsilon(D;\xi)]$ strictly decreases in D for $D < \bar{z}$. Thus, we consider only $D \in [\bar{z}, \bar{\bar{D}}(\xi)]$. Moreover, for $\xi > \frac{\sigma}{\sigma+r}\Psi^m$, there is no implementable D. So we only consider $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$. In this range, $\mathcal{W}^c[D, \epsilon(D;\xi)] = \sigma[u(D) - D] - \xi\epsilon(D;\xi)$. Proposition 1.3.1(2) implies that $\mathcal{W}^c[D, \epsilon(D;\xi)]$ is continuously differentiable on (D, ξ) for $\epsilon \in [\underline{\epsilon}(\xi), \bar{\bar{\epsilon}}(\xi)]$. By the Theorem of Maximum, $\mathcal{W}^c(\xi)$ is continuous in ξ. Moreover, $\bar{\bar{D}}(\xi)$ is continuously differentiable for $\xi \in [0, \xi^*)$ and $\xi \in (\xi^*, \frac{\sigma}{\sigma+r}\Psi^m]$. Now, by the Envelope Theorem,

$$\frac{\mathcal{W}^c(\xi)}{d\xi} = -\xi\frac{d\epsilon}{d\xi} - \epsilon + \lambda\frac{d\bar{\bar{D}}}{d\xi} < 0, \tag{1.31}$$

where $\lambda > 0$, for $\xi < \xi^*$ and $\xi^* < \xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m$. This, together with the continuity of $\mathcal{W}^c$, implies that $\mathcal{W}^c(\xi)$ strictly decreases in ξ for all $\xi \in [0, \frac{\sigma}{\sigma+r}\Psi^m]$.

Now, when $\xi = 0$, $D^*(0) = y^* > \bar{z}$ and hence $\mathcal{W}^c(0) > \mathcal{W}^m$. When $\xi = \frac{\sigma}{\sigma+r}\Psi^m$, the only implementable $D = \bar{\bar{D}}(\frac{\sigma}{\sigma+r}\Psi^m) = \bar{z}$, so at optimal $D^*(\frac{\sigma}{\sigma+r}\Psi^m) = \bar{z}$ and

$$\mathcal{W}^c(\frac{\sigma}{\sigma+r}\Psi^m) = \sigma[u(\bar{z}) - \bar{z}] - \frac{\sigma}{\sigma+r}\Psi^m \bar{\epsilon}(\bar{z}; \frac{\sigma}{\sigma+r}\Psi^m) < \mathcal{W}^m.$$

Thus there exists a unique $\hat{\xi} \in [0, \frac{\sigma}{\sigma+r}\Psi^m]$ such that $\mathcal{W}^c(\hat{\xi}) = \mathcal{W}^m$.

Proof of Proposition 1.4.3

From Proposition 1.4.2, when $\xi \leqslant \hat{\xi}$, $\mathcal{W}^c(\xi) = \sigma[u(D) - D] - \xi\epsilon(D;\xi)$ and is twice continuously differentiable. The ranges of the choice variable $D \in [\bar{z}, \bar{\bar{D}}(\xi)]$ are both weakly decreasing in ξ. Moreover, there exists $\tilde{\xi}$ such that $\frac{\partial^2 \mathcal{W}^c(\xi)}{\partial D \partial \xi} < 0$ for $\xi < \tilde{\xi}$, because $\frac{\partial^2 \mathcal{W}^c(\xi)}{\partial D \partial \xi}$ is continuous on ξ and when $\xi = 0$:

$$\frac{\partial^2 \mathcal{W}^c(\xi)}{\partial D \partial \xi}|_{\xi=0} = (-\frac{\partial \epsilon}{\partial D} - \xi \frac{\partial^2 \epsilon}{\partial D \partial \xi})|_{\xi=0} = -\frac{\partial \epsilon}{\partial D}|_{\xi=0} = -\frac{r\epsilon - \epsilon^2 \sigma[u'(D) - 1]}{rD} < 0. \tag{1.32}$$

Then by supermodularity, the optimal credit limit $D^*(\xi)$ decreases in ξ when $\xi \leqslant \tilde{\xi}$.

Given functional form $u(y) = \frac{y^\alpha}{\alpha}$ $(0 \leqslant \alpha \leqslant 1)$ and $\xi \leqslant \hat{\xi}$, we discuss the optimal credit limit $D^*(\xi)$ in two cases:

1. $r < \sigma\frac{1-\alpha}{\alpha}$. From Proposition 1.4.1(2), social welfare first increases in D and then decreases in D, so for any $\xi \in [0, \hat{\xi}]$, the optimal credit limit is uniquely determined by FOC, $\sigma D^*[u'(D^*) - 1] = \xi\epsilon(D^*, \xi)$, with $\frac{dD^*}{d\xi} = \frac{\epsilon + c\frac{d\epsilon}{d\xi}}{\sigma[u'(D^*) - 1 + D^* u''(D^*)] - \xi\frac{d\epsilon}{dD^*}} < 0$.

2. $r \geqslant \sigma\frac{1-\alpha}{\alpha}$. From Proposition 1.3.1, for any $\xi \leqslant \hat{\xi}$, $\bar{\epsilon}(\xi) = 1$ because $r\bar{z} \geqslant \frac{\sigma}{\sigma+r}\Psi^m$. Then by the similar method used in the proof of Proposition 1.4.1(2.b), social welfare in a pure credit equilibrium now increases in D as $\frac{d\mathcal{W}^c[D, \epsilon(D;\xi)]}{dD} > 0$ for all $D \in [\bar{z}, \bar{\bar{D}}(\xi)]$. So the optimal credit limit is $D^*(\xi) = \bar{\bar{D}}(\xi)$, which decreases with ξ as $\frac{d\bar{\bar{D}}}{d\xi} = \frac{d\bar{D}(1;\xi)}{d\xi} < 0$ as in equation (1.29).

Proof of Proposition 1.4.4

Rewrite the incentive constraint as in equation (1.23): $g(D; \epsilon, \xi) = \Psi^c(D) - \frac{r}{\epsilon}D - \frac{r}{\sigma}\xi v(\epsilon)$, which is a decreasing function on D with minimum value negative. We prove that $\xi \leqslant \min\{\frac{1}{v'(1)}\frac{\sigma}{\sigma+r}\Psi^m, r\bar{z}\}$ is a sufficient condition for implementability by showing that $g(\tilde{D}; \epsilon, \xi) \geqslant 0$ for any $\epsilon \in [0, 1]$. Note that given $\xi \leqslant r\bar{z}$ and the assumption $v(1) = 1$, $\xi v(\epsilon) \leqslant r\bar{z}$ for any ϵ. So $\tilde{D} = \frac{\xi v(\epsilon)}{r}$ and $g(\tilde{D}; \epsilon, \xi) = \Psi^m - \frac{\sigma+r}{\sigma}\frac{\xi v(\epsilon)}{\epsilon} > \Psi^m(1 - \frac{v'(\epsilon)}{v'(1)}) \geqslant 0$.

The highest implementable credit limit $\bar{D}$ increases with ϵ as: when $\bar{D} < \bar{z}$, $\frac{d\bar{D}}{d\epsilon} = \frac{\Psi^c(\bar{D})-\xi v'(\epsilon)(\epsilon+\frac{r}{\sigma})}{r(1-\epsilon)} \geqslant \frac{\Psi^c(\tilde{D})-\xi v'(\epsilon)(\epsilon+\frac{r}{\sigma})}{r(1-\epsilon)} = \frac{\Psi^m-\xi v'(\epsilon)(\epsilon+\frac{r}{\sigma})}{r(1-\epsilon)} \geqslant \frac{\Psi^m(1-\frac{v'(\epsilon)}{v'(1)})}{r(1-\epsilon)} \geqslant 0$; and when $D \geqslant \bar{z}$, $\frac{d\bar{D}}{d\epsilon} = \frac{\Psi^c(\bar{D})-\xi v'(\epsilon)(\epsilon+\frac{r}{\sigma})}{r-\sigma\epsilon(u'(\bar{D})-1)} \geqslant \frac{\Psi^m(1-\frac{v'(\epsilon)}{v'(1)})}{r-\sigma\epsilon(u'(\bar{D})-1)} \geqslant 0$.

Proof of Proposition 1.4.5

The proof follows the same path as under linear cost. Optimal social welfare strictly decreases with ξ is proved by the Envelope Theorem as in equation (1.31), and with convex cost, $\frac{d\mathcal{W}^c(\xi)}{d\xi} = -\epsilon v'(\epsilon)\frac{d\epsilon}{d\xi} + \lambda\frac{\bar{\bar{D}}}{d\xi} - v(\epsilon) < 0$ with $\lambda > 0$. The optimal credit limit decreases with ξ when ξ is sufficiently small is proved by supermodularity as in equation (1.32), and with convext cost:

$$\frac{\partial^2\mathcal{W}^c(\xi)}{\partial D\partial\xi}|_{\xi=0} = (-v'(\epsilon)\frac{\partial\epsilon}{\partial D} - \xi v'(\epsilon)\frac{\partial^2\epsilon}{\partial D\partial\xi})|_{\xi=0} = -v'(\epsilon)\frac{\partial\epsilon}{\partial D}|_{\xi=0} < 0.$$

Proof of Proposition 1.4.6

Given $i \geqslant r$, the equilibrium real balances is $\bar{z}(i) \leqslant \bar{z}$ which solves $\sigma[u'(\bar{z}(i))-1] = i$, and the continuation value of using money only is $\Psi^m(i) = -i\bar{z}(i) + \sigma[u(\bar{z}(i)) - \bar{z}(i)] \leqslant \Psi^m$. The continuation value of using credit when $D < \bar{z}(i)$ is $\Psi^c(D;i) = \Psi^m(i) + iD - \xi\epsilon$; otherwise it is the same as described in Lemma 1.3.4.

The incentive constraint is now: $g(D;\epsilon,\xi,i) = \Psi^c(D;i) - \frac{r}{\epsilon}D - \frac{r}{\sigma}\xi \geqslant 0$, which is either a strictly decreasing function on D, or a first increasing then decreasing function on D, depending on the value of ϵ and i. We will show that in both cases, given $\xi \leqslant \min\{\frac{\sigma}{\sigma+r}\Psi^m(i), i\bar{z}(i)\}$, $g(\tilde{D}(\epsilon;\xi,i);\epsilon,\xi,i) \geqslant 0$ for any $\epsilon \in [0,1]$, therefore, there exists $\bar{D} > \tilde{D}$ such that $g(\bar{D};\epsilon,\xi,i) = 0$. To prove it, note that if $\xi \leqslant i\bar{z}(i)$, $\tilde{D}(\epsilon;\xi,i) = \frac{\xi\epsilon}{i} \leqslant \bar{z}(i)$. So $g(\tilde{D}(\epsilon;\xi,i);\epsilon,\xi,i) = \Psi^m(i) - \frac{r}{i}\xi - \frac{r}{\sigma}\xi \geqslant \Psi^m(i) - \xi - \frac{r}{\sigma}\xi \geqslant 0$ as $\xi \leqslant \frac{\sigma}{\sigma+r}\Psi^m(i)$.

To show that the welfare implication is the same as it is without inflation, we need to prove that the highest implementable credit limit $\bar{\bar{D}}(\xi) = \bar{D}(1;\xi)$ satisfies $\bar{D}(1;\xi) \geqslant \bar{z}(i)$ when $\xi \leqslant \min\{\frac{\sigma}{\sigma+r}\Psi^m(i), i\bar{z}(i)\}$, which is shown by $g(\bar{z}(i);1,\xi,i) = -r\bar{z}(i) + \sigma[u(\bar{z}(i)) - \bar{z}(i)] - \frac{\sigma+r}{\sigma}\xi \geqslant \Psi^m(i) - \frac{\sigma+r}{\sigma}\xi \geqslant 0$.

Chapter 2

Optimal Tax Policy in The Presence of Informality

2.1 Introduction

The patterns of inflation rates and tax revenues in developing countries are very different from those in developed countries. As shown in Table 2.1, in low income countries, seigniorage revenue accounts for around 10.53% of government tax revenue, much larger than their counterparts in high income countries at around 1.65%. Similarly, the corresponding inflation rates in developing countries (6.71%-7.5%) are also significantly higher than those in developed countries (1.95%-4.78%). In this paper, we propose a monetary model with both formal and informal sectors to answer the question: What is the optimal joint structure of inflation and taxation, taking into account the need for public goods and the different prevalence of informality in different countries?

We answer this question by taking the frictions in the informal sector seriously. We study how these underlying frictions not only affect governments' abilities to collect revenue but also how they impact trade and hence welfare from the private sector. In particular, we consider a stylized model where the government needs to collect taxes from the economy to finance public expenditure. However, such taxation, such as income tax or VAT, requires monitoring and record-keeping, which can be done only in the formal sector.[1] Alternatively, government can finance its expenditures by increasing the currency

[1] See Gordon and Li (2009) for detailed sources of government revenue from income tax, corporate income tax, consumption and production tax and border taxes. The author find that income tax accounts for 54.3% of tax revenue in high income countries and 31.2% in developing countries. Border taxes are 8.6% of tax revenue in developing countries and 0.7% in developed countries. We do not distinguish the type of tax on the formal market and call the total amount of tax that the government collected from

Table 2.1: Sources of government revenue, inflation and size of informal sector (2002-2017)

Income level	Tax revenue (% of GDP)	Seigniorage revenue (% of tax revenue)	Inflation rate	Informal sector
Low income	10.66%	10.53%	7.50%	40.94%
Lower middle income	16.28%	6.19%	6.71%	38.31%
Upper middle income	16.06%	4.84%	4.78%	36.52%
High income	19.95%	1.65%	1.95%	21.25%

Notes: author calculation. Income level is according to World Bank country classifications (2021-2022). Tax revenue and inflation rate data are from *World Development Indicators*, the World Bank group, 2022. The informal sector data are from the World Bank Informal Economy Database (2021), and the size of the informal sector is estimated by the informal sector output as a percentage of GDP, using the Multiple Indicators Multiple Causes (MIMIC) model-based estimates method. Seignorage revenue is calculated by the change of currency in circulation as a percentage of GDP, the currency in circulation data is from IMF International Financial Statistics (IFS) 2022.

in circulation, which would then indirectly tax the informal sector as money is used as means-of-payments. However, such inflation tax is distortionary as it affects consumers' incentives to hold cash and hence decreases trade surpluses from the informal sector. A benevolent government, when choosing a combination of fiscal policy and monetary policy to finance public goods, would take into account these trade-offs between public good provision and trade surpluses in the private sector. The optimal monetary policy and tax scheme depends, therefore, on the underlying structure of the economy. The optimal inflation rate is then shown to increase with the size of informal sector. The optimal policy is deflationary when the size of informal sector is small but is inflationary when the informal sector is large.

We use the Lagos-Wright (2005) (henceforth L-W) framework and propose a model with both formal market and informal market. The main distinctions between the two markets are the underlying frictions that prevail in the sector. In the formal market agents can access the record-keeping system to use credit but, because of such record-keeping, the government can levy taxes directly. In the informal market such monitoring is absent and agents can only use money as means-of-payment. At the same time, this lack of monitoring implies that agents can avoid taxation in these trades. The L-W framework provides a tractable framework that allows for the endogenous demand of means-of-payments. Thus higher inflation implies a higher opportunity cost of carrying cash and decreases trade surpluses in the informal markets. In the formal sector, agents can use both money and credit, but credit is preferred as it does not entail such an

formal trade “credit tax”.

opportunity cost.

In developing countries, there exist large-scale informal sectors. Table 2.1 shows that the size of the informal sector, measured by the percentage of informal output to GDP, ranges from 40.94% in low income countries to 21.25% in high income countries. There are many stylized differences between formal and informal activities; for example, employers in the informal market are low-skilled, production in the informal sector is labour-intensive and with low productivity; furthermore, the informal sector involves many illegal activities. But in this paper, we focus on the different tax schemes and payment methods between formal and informal activities, and our interest is mainly in the fact that the large "nontaxable" sector distorts the tax scheme and influences the efficiency of tax policy. In our paper, the informal sector refers to activities that are hidden from government regulation, and where traders do not have access to finance and use money as the only means-of-payment.

The model proposed here has some attractive features that fit with the the findings from La Porta and Shleifer (2014) regarding the informal sectors in developing countries. First, the informal sector is less productive but it produces different goods from those of the formal sector. Accordingly, our model features two markets for two types of goods, and consumers enjoy utilities from both markets. However, the trade surplus from the informal sector is smaller. Second, while the informal sector has very little access to credit, financial access for the formal sector is far from perfect.[2] Accordingly, we consider different debt limit in the formal sector to capture the different formal sector's financial access level. Finally, although tax schemes and government regulations influence the trade surplus in both the formal and informal market, regulation is not the main reason for informality, at least in the short term.[3] Accordingly, the size of the informal market is taken as given and we analyze the optimal fiscal policy and monetary policy.

Our main finding is that the optimal nominal inflation rate increases with the size of the informal sector. Optimal monetary policy can be either deflationary, constant, or inflationary, as determined by informal sector size. When the size of the informal sector is small, the seigniorage income has a small tax base. Thus it is optimal to conduct a

[2]See Table 2 "Obstacles to Doing Business" in La Porta and Shleifer (2014). 43.8% of informal enterprises and 18.5% of formal enterprises find the most important obstacle is "access to financing". Only 0.1% of informal enterprises and 5.3% of formal enterprises report the most important obstacle is "tax administration", likewise 1.8% and 3.3% for "labour regulations", and 3.3% and 0.8% for "legal system".

[3]See Sri Lanka. De Mel, McKenzie, and Woodruff (2013) and the discussion in La Porta and Shleifer (2014). Field experiments that test the effect of government actions to induce informal firms to register and formalize show that removing the entry cost and regulation fee for informal firms barely incentivizes them to be formal.

deflationary policy and use the tax revenue from the formal sector to increase the value of money and encourage informal trade. When the size of the informal sector is large, the optimal monetary policy is inflationary. This is because although a high inflation rate increases the opportunity cost of using money and hurts the informal trade surplus, the high seignorage revenue that could be collected can finance more public goods, which is welfare improving. Moreover, we find that better credit conditions can alleviate this effect. A higher credit limit directly increases trade surplus in formal markets, and allows the government to collect more credit tax and rely less on seigniorage, so optimal inflation rate decreases with credit limit.

We contribute to the literature in three ways. First, our work provides a tractable model and clear equilibrium analysis to explain the underlying mechanisms by which informality influences optimal taxation and monetary policy. The following analysis could easily be solved with paper and pencil, so it also leaves significant room for further extension. Second, we adopt a monetary framework and describe the means-of-payment available in formal and informal markets explicitly. By doing so, the model is made suitable for studying the impact of inflation policy on people's participation in the credit market, and also the impact of credit conditions on the optimal inflation policy. We show that the overall effect depends on the fundamentals, but in general better credit helps to decrease inflation. The third contribution here is that we consider the endogenous government spending decisions. We assume that the optimal tax scheme and inflation policy should be designed to reallocate resources not only between the formal and the informal sectors but also between the public and the private sectors. We show that the main positive relationship between the size of informal sector and inflation policy still holds, but that optimal policy now depends on the government's preference over the private sector and public sector.

The motivation for endogenous government spending comes from the fact that tax revenue and government spending increase with GDP. As shown in Table 2.1, in low-income countries, the total tax revenue accounts for on average 10.66% of GDP, while in high-income countries the ratio is 19.95%. The different public expenditure levels can be interpreted exogenously. For example, governments have differing military needs and are obliged to finance a certain amount of public goods. The only problem is how to finance expenditure efficiently from formal and informal sectors. We consider the exogenous government spending in section 2.4.1. Then in section 2.4.2 we interpret the different tax revenue as the optimal policy that maximizes social welfare. We assume that the provision of public goods, such as infrastructure, education, and health services, has welfare implications, and that governments take this effect into consideration. Therefore, optimal tax structure and monetary policy are designed to maximize not only the formal and the informal trade outputs, but also the optimal provision of public goods.

In both considerations the optimal inflation rate is shown to increase with the size of informal sector, but through different mechanisms. In the case of exogenous spending, large informal sectors restrict the gain from formal market trade surpluses and impact the government's fiscal capacity. Therefore, to collect enough revenue, governments must rely on more seigniorage and a higher inflation rate. In the case of endogenous spending case, the positive relationship is shown to come from a balance between the benefit from using seigniorage revenue to increase public good supply and the cost of a lower informal trade surplus. We show that if the government has a relatively high preference for public goods over the informal market performance, optimal inflation can be expected to increase with an increase in informal sector size.

Related Literature

It has been traditional to discuss the effect of inflationary finance and its welfare implications. The main idea is that "an increase of the inflation rate performs like a tax to restrain consumption demand and thus releases resources for capital formation or public use", as discussed in Phelps (1973). But this argument has been challenged. For example, Kimbrough (1986), Chari et al. (1996), and Correia and Teles (1996) showed that the Friedman rule continues to be optimal even though distorting taxes must be levied for revenue purposes. One underlying assumption in the discussions listed here is that agents have preferences for both money-goods and credit-goods, and hence they hold money to trade, which creates a tax base for seigniorage. At the same time they assume all consumptions are fully taxable. However, in many developing countries, agents only operate in the shadow economy in order to avoid tax. Given the large non-taxable economy, seigniorage revenue and inflation could then be the only channel of taxation.

There are many recent works that calibrate optimal monetary policy and seigniorage revenue in the developing countries with a large informal economy. Koreshkova (2006) proposes a model with formal and informal sectors and emphasizes the difference between the two sectors on productivity and tax policy. It is assumed that both sectors have access to money and credit, subject to the exogenous cost functions of using credit and of a cash-in-advance constraint. Gordon and Li (2009) propose an overlapping-generations model and assume firms can choose to be formal or informal. They focus on the free entry condition and conclude that the optimal policy should make the marginal profits from being formal and informal equal. They assume that by accessing the financial system, firm revenue increases by a fraction, but is subject to taxation. By being informal, firms face inflation costs and an exogenous cost of theft. The main results of the present paper are consistent with their findings, which is that the optimal inflation rate increases with the size of informal sectors. But we adopt a monetary model, do a full equilibrium analysis, and describe the use of cash and credit explicitly.

In this paper, a New Monetarist method (Lagos and Wright, 2005, or "L-W" model) is used to study optimal monetary policy. The LW model prescribes the searching process between traders where an exchange media is necessary to overcome market frictions such as double-coincidence of wants and lack of commitment. Given the micro foundation of money, the L-W framework is useful for constructing an analysis of the use of money and credit and it is widely used in monetary policy papers. There are some papers that study the informal sector and tax avoidance within the LW framework. Gomis-Porqueras et al. (2014) use the L-W framework to discuss the optimal size of the informal sector. They focus on the long-run effect of regulation and inflation on agents' preference for being both formal and informal, and use the model to calibrate the size of the shadow economy.

A more closely related paper is Aruoba (2021), where the L-W framework and Ramsey optimal-policy approach are used to explain the optimal inflation rate and income tax rate for financing public expenditure in developing countries. We ask a similar research question but the modeling strategies used by us are different. Aruoba (2021) assumes that the formal market is perfectly competitive and frictionless. All sellers operate in the formal market, subject to income tax, and choose the extent of informal activity, which is liquidity-constrained but capable of avoiding tax. The credit conditions in our paper are not always perfect. We assume the credit limit can very be low, capturing the fact that the gain from access to financial intermediaries is smaller in poorer countries. In our model, sellers need to balance the benefit of accessing credit and being taxed, with the option that they can use cash to trade. Therefore, the tax that government can collect from the formal sector depends both on the credit level and the monetary policy. By doing so, we are able to capture the idea that is indicated in Gordon and Li (2009), "high inflation not only generates additional revenue, but also can induce firms to make use of the financial sector", and hence, we find that, in some cases, high inflation would increase the government fiscal capacity from the formal sector.

The model we put forward is also related to the literature that models the coexistence of money and credit, as in Lotz and Zhang (2016), Araujo and Hu (2018), and Jiang and

Shao (2020). But our main focus is on public finance and tax avoidance. There are other papers that incorporate two DM rounds in the L-W framework. Telyukova and Wright (2008) study the 2 DM rounds to address the credit card puzzle. Araujo and Hu (2018) study the optimal intervention with 2 DM rounds. In Araujo and Hu (2018), it is shown that it is always welfare improving for government to tax the credit trade and conduct deflationary policy, or to set inflation high and use the seignorage revenue to subsidize credit. Our model borrows their framework but with a different role of government. In our paper, government needs tax revenue and seigniorage revenue from both money and credit trades to finance spending.

We need to emphasize that there are many other factors that influence optimal public financing in developing countries. We are not considering these in our theoretical work as our intention is to keep the model tractable, and also because empirical evidence shows that the positive impact of informality on optimal inflation is persistent. Political instability can be an important reason for high inflation, see Cukierman et al (1992) and Aisen and Veiga (2006). But Baklouti and Boujelbene (2019) have found that in countries with high political instability, government spending is financed more by seigniorage than tax, and a large-sized informal sector accelerates this effect. Another factor is central bank independence. Mazhar and Méon (2017) have pointed out that in countries with a well developed central banking system, the government may not be able to control inflation or seigniorage to finance spending. But they find that the positive impact of the informal sector on the inflation rate still exists and that such an effect is stronger in developing countries which have weak central bank independence.

In the next section, we introduce the environment. In Section 2.3, the discussion looks at a baseline model with no taxation and a constant money supply. Then Section 2.4 discusses the optimal credit tax and inflation rate.

2.2 Environment

The model follows the tradition of Lagos and Wright (2005). Time is discrete and infinite. The economy is populated by two types of infinitely-lived agents, a unit measure of buyers and a unit measure of sellers. Agents are subject to limited commitment friction. Sellers are different in production technology and access to financial services. There are $1-\alpha$ "formal sellers" who have access to a record-keeping technology and accept credit transactions. There are α informal sellers who do not have access to the credit system and only accept cash.

The record-keeping and enforcement technology can observe, monitor, record, and retain internet data, and it is further assumed that the system has the power to enforce people to repay debt up to a threshold D. The credit limit measures the financial development level, and in advanced economies the credit limit is high; but in poor countries credit limit can be very low. D can also be interpreted as the contract enforcement level, influenced by the quality of judicial processes, government coercion power, cost and efficiency in the court system, and many other factors. Such credit limit is exogenous given, and we assume that the record-keeping technology keeps history data for just one period. Therefore, as people are subject to limited commitment friction, any debt above the credit limit D cannot be promised to repay.

While formal sellers have access to the credit system and accept delayed settlement, whether they also accept cash if they do not access the system is under discussion. It would be possible to model the formal sectors as the supermarkets, restaurants, or retailer shops that accept both cash and credit card. But they could also be modeled as online retailers or manufacturers where trades are subject to spatial and temporal separation. Therefore, cash transaction may be difficult to implement, and even though firms choose not to use credit, the transactions also rely on the record-keeping system and hence are subject to taxation. Allowing agents to access different means-of-payment in the formal sector influences the outside option of tax avoidance, and hence influences the government's tax capacity. We discuss the assumption that formal trade cannot be facilitated by cash, and then discuss the other assumption in the Appendix.

Each day is divided into three stages: a decentralized formal market (DM1), a decentralized informal market (DM2), and a centralized market (CM). Buyers participate in each market sequentially, and sellers are only active in their own markets and in the centralized market. In the first stage, buyers randomly meet formal sellers in the decentralized formal market (DM1) and trade "formal good" y_f. Then the meeting breaks and buyers enter the second stage. In the second stage, buyers and informal sellers randomly meet bilaterally in a decentralized informal market (DM2) and trade "informal good" y_i, and then the match breaks. In the last stage, all buyers and sellers meet in a centralized market(CM) and trade "general good" x. Agents pay taxes, repay debt, and balance money holdings in a competitive market for money. They also receive public goods and services G. All goods are non-storable in each market.

Buyers consume formal goods and informal goods with utility $u_f(y_f)$ and $u_i(y_i)$ but do not have production technology in the decentralized markets. Sellers have the production technology with cost $c_f(y_f)$ and $c_i(y_i)$ but do not want to consume in the decentralized markets. We assume that buyers' utility functions satisfy $u'(\cdot) > 0$, $u''(\cdot) < 0$, $u'(0) = +\infty$ and $u'(\infty) = 0$, and assume that sellers' cost functions are linear: $c_f(y_f) = \rho_f y_f$ and $c_i(y_i) = \rho_i y_i$. We define the first-best trade outcomes as y_j^* $(j = f, i)$ that maximize the trade surplus $u_j(y_j^*) - \rho_j y_j^*$. All agents consume and produce CM goods with linear utility and cost x ($x < 0$ means production), and buyers get utility from the public goods and services G with utility $U(G)$.[4] Agents discount between each day with discount factor β, and denote the discount rate as $r = 1/\beta - 1$. The instantaneous utility functions of buyers and sellers are given by:

$$U^b(y_f, y_i, x) = u_f(y_f) + u_i(y_i) + x + U(G) \text{ and } U^j(y_j, x) = -\rho_j y_j + x \text{ for } j = i, f.$$

In the first two decentralized markets, the matching process is described as follows: In DM1, measure $1 - \alpha$ formal sellers randomly match with measure 1 of buyers and the number of successful matchings is $\min\{1, 1 - \alpha\} = 1 - \alpha$. So with probability $1 - \alpha$, a buyer meets a formal seller and trades formal good y_f. The corresponding probability for formal sellers to meet a buyer is 1. Then in DM2, measure 1 of matched buyers and measure α informal sellers meet in an informal market with the number of informal trade in DM2 $\min\{1, \alpha\} = \alpha$. With probability α a buyer successfully matches an informal seller, and with probability 1 the seller meets the buyer. Note that the seller's probability of matching does not depend on the size of the markets or the sequence of the markets.

In every successful meeting in DM1 and DM2, buyers need to decide the quantity of good y_j they would like to consume and also the payment method, money b_i^m and credit b_f^c (in terms of CM good) they use. Money transfer is instantaneous and debt transfer needs to be repaid in the CM. The trade objects (y_f, b_f^c) and (y_i, b_i^m) are determined by proportional bargaining, and the buyer's bargaining power is θ.

In the subsequent CM, all buyers and sellers meet in a central market to repay debts, pay taxes, get public goods and rebalance their liquidity holdings. The government has the technology to transfer CM goods into public goods G one-to-one, and public spending is financed by direct tax from formal trade and indirect inflation tax by printing money. It is assumed here that the tax authority has perfect information about the formal trade history (y_f, b_f^c) and has perfect enforcement power to tax buyers in the CM based on history. In the baseline model, we assume there is no taxation levied on credit trade and

[4] Because of the linear utility function assumption in CM, it could be shown that agents would not produce general goods for their own consumption. The only reason they want to consume and produce general goods is to balance money holdings. It can also be shown that in the CM, sellers do not bring any liquidity to the next period.

that money growth is constant. Then we discuss the model where governemnt can collect fixed credit tax χ for those who access credit.

The total amount of fiat money in circulation in period t is M_t with money growth rate γ, so $M_{t+1} = \gamma M_t$. Let ϕ_t to denote the price for money in terms of CM goods and define the real balance in each period as $Z_t = \phi_t M_t$. We focus on steady-states where real balances are constant: $\phi_t M_t = \phi_{t+1} M_{t+1}$.

Given allocation (y_f, y_i, G), the social welfare is:

$$\mathcal{W}(y_f, y_i, G) = (1-\alpha)[u_f(y_f) - \rho_f y_f] + \alpha[u_i(y_i) - \rho_i y_i] + v(G), \tag{2.1}$$

where $v(G) \equiv -G + U(G)$. Define the first-best government spending level as G^* that solves $U'(G^*) = 1$. We suppose that G^* is high enough and is not achieved in the equilibrium. In section 2.4.1, we assume government takes $G < G^*$ as exogenously given. In section 2.4.2 we endogenize optimal G. Note that the utility function and production function should ensure that social welfare decreases with size of informal sector in the baseline model, so we assume that in equilibrium, $u_f(y_f) - \rho_f y_f \geq u_i(y_i) - \rho_i y_i$. In the analysis, it is shown that the assumption always holds when the credit limit is high such that buyers are willing to use credit in the formal market, and when the informal sector is unproductive, so the cost of production ρ_i is high.

We begin from a baseline model with constant money supply and no taxation in section 2.2.3. Then we discuss the optimal tax policy in Section 2.2.4.

2.3 No Tax and Constant Money Supply

Assume that the money supply is constant over time and that there is no direct tax. Each buyer will take debt limit D as given and decide on an offer $(y_f, b_f^c,)$ if he meets the formal seller in DM1, and on offer (y_i, b_i^m) if he meets the informal seller in DM2. We use (y_i^0, y_f^0) to describe the equilibrium object (superscripts 0 denote the equilibrium outcome in the baseline model).

In CM, because we assume lineal utility and cost function, it could be shown that sellers and buyers' value functions are lineal in their real balance holding z and debt b_f^c. To be specific, buyers with real balance z and debt b_f^c have value $W^b(z, b_f^c) = z - b_f^c + U(G) + \max_{z' \geq 0} \left\{-z' + \beta V_f^b(z')\right\}$. This can be expressed in a more convenient way: $W^b(z, b_f^c) = z - b_f^c + W^b$, where W^b is a constant value.

In DM2, buyers meet sellers in the informal market and decide the trade object (y_i, d_i^m) through proportional bargaining. Combining with the linearity of W^b, buyers with real balance z and formal debt b_f^c have continuation value:

$$\begin{aligned} V_i^b(z, b_f^c) &= \alpha \max_{y_i, b_i^m} \{u_i(y_i) - \phi b_i^m\} + W^b(z, b_f^c), \\ \text{s.t.} \quad u_i(y_i) - \phi b_i^m &= \frac{\theta}{1-\theta}(-\rho_i y_i + \phi b_i^m) \text{ and } \phi b_i^m \le z. \end{aligned} \tag{2.2}$$

Denote the real balance that the buyers pay the sellers to consume y_i as $z_i(y_i) \equiv (1-\theta)u_i(y_i) + \theta\rho_i y_i = \phi b_i^m$. Informal trade surplus for buyers is $\theta[u_i(y_i) - \rho_i y_i]$ and for seller is $(1-\theta)[u_i(y_i) - \rho_i y_i]$. So buyers' DM2 problem can be rewritten as:

$$\begin{aligned} V_i^b(z, b_f^c) &= \alpha \max_{y_i} \theta[u_i(y_i) - \rho_i y_i] + W^b(z, b_f^c), \\ \text{s.t.} \quad z_i(y_i) &= (1-\theta)u_i(y_i) + \theta\rho_i y_i \le z. \end{aligned} \tag{2.3}$$

Equation (2.3) indicates that if buyers enter DM2 with enough real balance $z \ge (1-\theta)u_i(y_i^*) + \theta\rho_i y_i^*$, then first-best outcome is reached. Otherwise, buyers will use all the money to purchase informal market goods. The equation (2.3) also implies that the buyer's credit condition does not influence his informal trade outcome, hence we have $V_i^b(z, b_f^c) = -b_f^c + V_i^b(z, 0)$.

The decision in DM1, when buyers meet formal sellers, is similar and the trade objects are (y_f, b_f^c), including the amount of formal good y_f and debt b_f^c. Given debt limit D, the buyer's question in DM1 is:

$$\begin{aligned} V_f^b(z) &= (1-\alpha) \max_{y_f, b_c^f} \left\{u_f(y_f) + V_i^b(z, b_f^c)\right\} + \alpha V_i^b(z, 0), \\ \text{s.t.} \quad u_f(y_f) - b_f^c &= \frac{\theta}{1-\theta}(-\rho_f y_f + b_f^c), \text{ and } b_f^c \le D. \end{aligned} \tag{2.4}$$

Again, the bargaining problem implies that the buyer uses all the credit limit to purchase DM1 goods and the problem (2.4) can be rewritten as:

$$\begin{aligned} V_f^b(z_f, z_i) =& (1-\alpha) \max_{y_f} \theta[u_f(y_f) - \rho_f y_f] + V_i^b(z, 0), \\ & \text{s.t.} (1-\theta)u_f(y_f) + \theta\rho_f y_f \le D. \end{aligned} \tag{2.5}$$

If the debt limit is high $D \ge (1-\theta)u_f(y_f^*) + \theta\rho_f y_f^*$, then the first-best outcome in formal trade y_f^* can be achieved; otherwise, buyers use credit D to purchase formal goods $y_f < y_f^*$.

Combining equation (2.3) and (2.5), buyers' real balance holding problem in CM is:

$$\begin{aligned} & \max_{0 \le y_f \le y_f^*, 0 \le y_i \le y_i^*} \left\{-r z_i(y_i) + (1-\alpha)\theta[u_f(y_f) - \rho_f y_f] + \alpha\theta[u_i(y_i) - \rho_i y_i]\right\}, \\ & \text{s.t. } (1-\theta)u_f(y_f) + \theta\rho_f y_f \le D, \text{ and } z_i(y_i) = (1-\theta)u_i(y_i) + \theta\rho_i y_i, \end{aligned} \tag{2.6}$$

where $r = \frac{1-\beta}{\beta}$ is the real interest rate. Equation (2.6) indicates that buyers need to balance the opportunity cost of holding money and the benefits from holding money which enable them to trade with sellers in the informal decentralized market. The optimal DM trade outcomes y_f, y_i are given by:

$$y_i = \bar{y_i} \quad \text{and} \quad \begin{cases} y_f = y_f^* & \text{if } D > (1-\theta)u_f(y_f^*) + \theta\rho_f y_f^*, \\ (1-\theta)u_f(y_f) + \theta\rho_f y_f = D & \text{otherwise.} \end{cases} \tag{2.7}$$

where $\bar{y_i}$ solves $rz_i'(\bar{y_i}) = \alpha(u_i'(\bar{y_i}) - \rho_i)$ and denote the equilibrium real balance holding. We define $\bar{y_f}$ that solves $r[(1-\theta)u_f'(\bar{y_f}) + \theta\rho_f] = (1-\alpha)[u_f'(\bar{y_f}) - \rho_f]$ to denote the formal trade outcome in a pure monetary equilibrium and from now on we focus on equilibrium such that the credit limit $D \geq (1-\theta)u_f(\bar{y_f}) + \theta\rho_f\bar{y_f}$.

In the baseline model with no direct tax and where money supply is constant, the government collects no tax to finance any public good $G = 0$ and the social welfare is given by:

$$\mathcal{W}(y_f^0, y_i^0) = (1-\alpha)[u_f(y_f^0) - \rho_f y_f^0] + \alpha[u_i(y_i^0) - \rho_i y_i^0]. \tag{2.8}$$

2.4 Fixed Credit Tax and Inflation Tax

Now we consider the case when the government collects fixed credit tax from formal credit trades and inflation tax from money holding, then use both tax income and seigniorage revenue to provide public goods. It will be shown that the optimal credit tax and inflation rate depend on the size of the informal sector in the economy, under both exogenous government spending and endogenous government spending assumption.

To be specific, buyers need to pay credit tax χ (in CM good) if they trade in the formal meeting. Assume the government sets a money growth rate $\gamma \equiv M_{t+1}/M_t$. If $\gamma > 1$, the government conducts inflationary policy and "punishes" holding money. If $\gamma < 1$, the government conducts a deflationary policy and increases the value of money. In both cases, government budget constraint is:

$$(1-\alpha)\chi + (\gamma - 1)\phi M \geq G, \tag{2.9}$$

where the left-hand side is the credit tax income $(1-\alpha)\chi$ from $1-\alpha$ formal meetings and seigniorage income $(\gamma - 1)\phi M = (1-\gamma)z_i(y_i)$ from real balance holding $z_i(y_i)$. The right-hand side is the government spending on public goods G.

Note that in CM, the buyer's value function is still linear in money holding and debt. DM1 problem is the same as in the baseline model because the lump-sum credit tax χ is not distortionary and does not influence the trade outcome in the formal meetings, as long as the credit tax is not too high. In DM2, the informal trade problem is the same as in (2.3). Therefore the question and solution for real balance holding in CM are given by:

$$\max_{0 \leq y_i \leq y_i^*} \left\{ -iz_i(y_i) + \alpha\theta(u_i(y_i) - \rho_i y_i) \right\}, \tag{2.10}$$

and

$$\alpha\theta[u'(\bar{y}_i) - 1]\rho_i = iz_i'(\bar{y}_i), \tag{2.11}$$

where i is the nominal interest rate defined by Fisher equation $(1+i) = (1+r)\gamma$. Note that the opportunity cost of bringing money is i and the equilibrium $\bar{y}_i$ decreases with the nominal interest rate and increases with the informal sector size α.

2.4.1 Exogenous government spending

Suppose that the government takes public goods spending G as given and designs the fiscal policy and monetary policy to finance it. Here we adopt the assumption that the formal sector is subject to time and spatial separation friction and hence cash payment is not feasible. In order to trade, buyers and sellers have to leave records in the system and the tax authority can tax the formal trade based on these records. But if the government sets the credit tax higher than the buyer's total trade surplus from formal trade, the buyer will avoid tax by simply skipping DM1. Therefore, the participation constraint for buyers is:

$$\chi(i) \leq \theta[u_f(y_f) - \rho_f y_f], \tag{2.12}$$

and the government problem is to choose credit tax χ and money growth rate γ to maximize social welfare and to finance exogenous G:

$$\max_{\gamma,\chi} \left\{ (1-\alpha)[u_f(y_f(D)) - \rho_f y_f(D)] + \alpha[u_i(\bar{y}_i) - \rho_i \bar{y}_i] + v(G) \right\}, \tag{2.13}$$

s.t.

$$\chi \leq \theta[u_f(y_f(D)) - \rho_f y_f(D)],$$
$$G \leq (1-\alpha)\chi + (\gamma - 1)z_i(\bar{y}_i).$$

Note that social welfare is independent with the lump-sum credit tax χ but strictly increases with real balance holding $\bar{y}_i$. The government problem can be transformed to choose the minimum inflation such that the budget constraint is binding. In the next theorem we discuss the optimal nominal interest i for given credit limit D, size of informal sector α and exogenous government spending G. We let $\theta = 1$ and assume utility functions take forms $u_j(y_j) = y_j^\eta/\eta$ $(0 < \eta < 1)$ for both $j = i, f$.

Theorem 2.4.1. *Suppose that $\theta = 1$ and $u(y) = y^\eta/\eta$, for given D and α, there exists $\bar{G}$ above which government spending can not be financed. When $G \leq \bar{G}$, there exists unique i^* that maximizes social welfare. Moreover, optimal i^* increases with α and decreases with D, when ρ_i is sufficiently high.*

Theorem 2.4.1 discusses the optimal inflationary policy when the government need to finance an exogenous given public spending. Note that for given D and α, the highest tax revenue and seigniorage revenue determines the upper bound $\bar{G}$ that can be financed. When G is lower than the threshold, Theorem 2.4.1 shows that the optimal inflation rate increases with the size of the informal sector, as long as the informal market is not too productive. The reason is that when the informal sector becomes larger, the total tax revenue and seigniorage revenue decreases.

To see this, note that with a large informal sector, buyers have a higher probability of meeting an informal seller, therefore they bring more liquidity, which provides a larger tax base for seigniorage. But the large informal sector decreases the probability of meeting a formal seller, and hence the total trade surplus from the formal market decreases, which constrains the government capacity to collect tax revenue from the formal market. The overall effect of a large-sized informal market on the total revenue depends on the benefit from seigniorage and loss from credit tax. We show that when ρ_i is sufficiently high, meaning that formal trade is more productive than informal trade, the latter effect outweighs the former effect and, therefore, a larger informal sector leads to lower total tax revenue.

Therefore, as higher α harms the tax base and decreases the total revenue, the optimal nominal interest rate i^* has to increase with α to compensate for the lack of revenue. Note that the optimal i^* is always lower than the nominal interest rate to maximize seigniorage revenues, hence, by increasing i, more seigniorage can be collected.

Theorem 2.4.1 also shows that the optimal nominal interest rate i^* decreases with credit limit D. The reason for this is that better credit conditions can alleviate the tax burden on seignorage revenue so lower inflation is needed. For a given size of informal sector, if the credit limit is high, the trade surplus from formal trade is high, so the government can collect more tax revenue from the formal market to finance spending. Therefore, the government can then rely less on seignorage revenue and set the optimal i^* lower.

When we assume that the outside option for buyers to evade tax in the formal market is to use cash, the impact of informal sector size on the inflation rate is not that straightforward. In the Appendix, we use a numerical example to discuss this assumption, and we show that in some cases, optimal nominal interest rate may decrease with the size of the informal sector. Here we restrict discussion to some underlying trade-offs.

When buyers can avoid tax by using money in the formal meeting, the inflation rate influences buyers' participation constraint. Hence, high inflation will have two effects on revenue. First, inflation influences the seigniorage revenue through the typical Laffer curve effect. Second, high inflation decreases the value of using money, so buyers are more likely to use credit in the formal trade, which increases the credit tax that governments can collect. With the two effects of inflation on revenue, in the Appendix we show that the total tax revenue may not always decrease with the size of the informal market and hence the optimal inflation rate may also decrease with informality.

2.4.2 Endogenous government spending

Now we move on to the endogenous government spending assumption. In this section we show that a positive relationship between the size of informal sector and the optimal inflation rate can also exist, but, different from in Section 2.4.1, the main trade-off now is between the provision of public goods and the trade outcome in the private sector.

We are still focusing on the case $D \geq (1-\theta)u_f(\bar{y_f}) + \theta\rho_f\bar{y_f}$ such that using credit leads to higher trade surplus than if buyers are allowed to use cash and credit is not available in DM1. And we also assume that cash transaction is not implementable in DM1. The social planner problem is to choose the tax policy, monetary policy and the government expenditure (γ, χ, G) to maximize social welfare subject to the equilibrium outcome $y_i = \bar{y_i}$, $y_f = y_f(D)$, budget constraint (2.9) and buyer's participation constraint (2.12):

$$(2.14) \qquad \max_{\gamma,\chi,G} \left\{(1-\alpha)[u_f(y_f(D)) - \rho_f y_f(D)] + \alpha[u_i(\bar{y_i}) - \rho_i\bar{y_i}] + v(G)\right\},$$

s.t.

$$\chi \leq \theta[u_f((y_f(D)) - \rho_f y_f(D)],$$
$$G \leq (1-\alpha)\chi + (\gamma - 1)z_i(\bar{y_i}).$$

Note that we assume the first-best public goods provision G^* is high enough and is not achieved in equilibrium so the welfare function strictly increases with G. At optimal, the participation constraint and budget balance constraint bind. So we rewrite the government problem by choosing i:

$$\max_{i\geq 0}\left\{\alpha(u_i(\bar{y}_i)-\rho_i\bar{y}_i)+v\left[\theta(1-\alpha)[u_f(y_f(D))-\rho_f y_f(D)]+\frac{i-r}{1+r}z_i(\bar{y}_i)\right]\right\}. \tag{2.15}$$

Equation (2.15) indicates that high inflation on the one hand directly decreases the real balance holding and informal trade surplus, but, on the other hand, it increases the inflation tax rate, which may generate high seigniorage revenue. The revenue can be used to finance more government spending and hence improve social welfare. To illustrate the trade-off clearly, from now on we assume that $\theta = 1$, utility functions take form $u_j(y) = y_j^\eta/\eta$ $(0 < \eta < 1)$ for both sectors $j = i, f$ and the utility from public goods is linear $v(G) = G$. The next Lemma discusses the optimal monetary policy under the specific functional forms.

Lemma 2.4.1. *Suppose that $\theta = 1$, $u(y) = y^\eta/\eta$ and $v(G) = G$, for given credit limit $D \geq (1-\theta)u_f(\bar{y_f}) + \theta\rho_f\bar{y}_f$, social welfare $\mathcal{W}$ first increases then decreases with i.*

Lemma 2.4.1 indicates a non-monotonic relationship between the inflation rate and social welfare and we describe it in Figure 2.1. The trade-off is between the informal sector trade outcome and public good provision. We find that there exists a range of low nominal interest rates i such that the social welfare increases with inflation. This is because when the nominal interest rate is small enough, the informal trade outcome approaches first-best, so the marginal loss of trade surplus from a small increase of i is low. Instead, a small increase in i leads to high seigniorage revenue to finance public goods, and the marginal effect is large. So welfare increases with higher inflation.

When the nominal interest rate i is higher than a threshold, Lemma 2.4.1 shows that welfare decreases with i. Because when i is too large, the gain from seigniorage revenue cannot offset the cost of informal trade surplus. The non-monotonicity is not only due to the Laffer Curve effect which implies a trade-off between the inflation tax rate i and the inflation tax base $\rho_i\bar{y}_i$, but also considers the trade-off between private sector performance and public sector output. The next theorem discuss the optimal nominal interest rate i^*.

Theorem 2.4.2. *Suppose that $\theta = 1$, $u(y) = y^\eta/\eta$ and $v(G) = G$, there exists unique i^* that maximizes social welfare, and i^* increases with α. Moreover, when $r \leq \bar{r}$, there exists $\bar{\alpha}$ such that $i^* \leq r$ when $\alpha \leq \bar{\alpha}$, and $i^* > r$ otherwise. When $r > \bar{r}$, $i^* < 0.5$ for any $\alpha \leq 0.5$.*

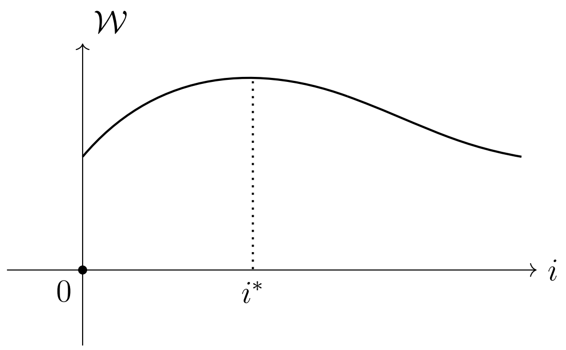

Figure 2.1: Social welfare and nominal interest rate with endogenous government spending

Theorem 2.4.2 shows that the optimal nominal interest rate i strictly increases with the size of the informal sector α. When the economy has a small informal sector, $\alpha \leq \bar{\alpha}$, that uses cash as the only means-of-payment, the total money holding in the economy is small and thus the total seigniorage revenue is low. Therefore, it is optimal to conduct a deflationary policy and set a low nominal interest rate, $i \leq r$, to increase the value of money and encourage informal trades, because the marginal gain from private sector output is larger than the seigniorage revenue loss. This relationship captures the fact that in developed countries with a small informal sector the optimal policy is a low inflation rate and low (or negative) seigniorage revenue. When the size of the informal sector is large, $\alpha > \bar{\alpha}$, the optimal monetary policy is inflationary. As the total money holding in the economy is large, high inflation increases the seigniorage revenue to finance more public goods and the positive effect is stronger than the negative impact on the trade surplus. The relationship between the optimal nominal interest rate and the size of the informal sector is described in Figure 2.2.

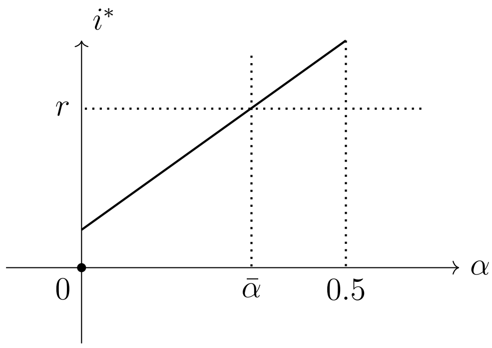

Figure 2.2: Optimal nominal interest rate and size of informal sector with endogenous government spending

Note that in Lemma 2.4.1 and Theorem 2.4.2, we let $v(G)$ be linear, indicating that the marginal utility gain from more public goods provision is constant at 1. But if we let $v(G)$ be a concave utility function, higher G will lead to lower marginal utility from consuming public goods, and the overall effects will be much more difficult to solve. However, as long as $U(\cdot)$ is not too concave such that the first-best G^* is high enough to be achieved in the equilibrium, the non-monotonic relationship between inflation and social welfare still exists and the optimal inflation increases with the size of the informal sector. We use a numerical example in Figure 2.3 to show this. The utility function take forms $u_j(y_j) = \sqrt{y_j}/0.5$ for $j = i, f$, $v(G) = \sqrt{G}$, $r = 0.1$, $\rho_f = \rho_i = 1$ and $D = 1$. When there is no informal sector, welfare is 2 and is independent of inflation policy. When the economy has a small informal sector $\alpha = 10\%$, the optimal nominal interest rate is $0.058 < 0.1$; and when the economy has larger informal sector $\alpha = 40\%$, the optimal nominal interest rate is $0.135 > 0.1$.

But note that when such marginal utility is too low or G^* is too low, the positive effect of high inflation on the seigniorage revenue may lead to a low utility gain. If the gain is too low to offset the decrease of informal trade surplus, the non-monotonic relation between i and $\mathcal{W}$ may turn into a strictly negative relationship.

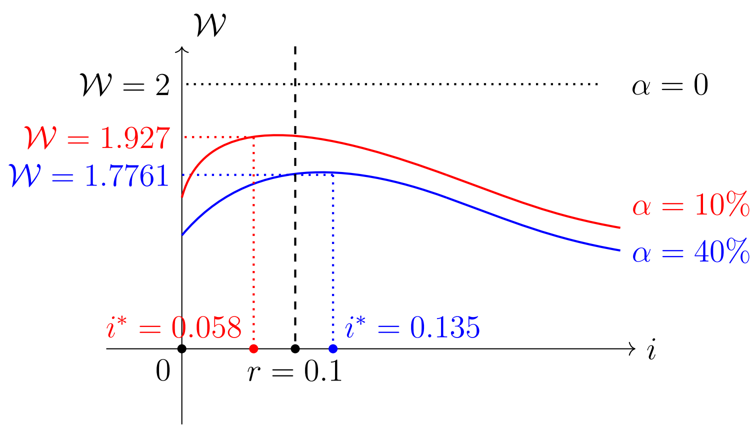

Figure 2.3: Social welfare and nominal interest rate with fixed credit tax with $v(G) = \sqrt{G}$

Another important trade-off that is missing in this section is the effect of credit conditions on the optimal monetary policy. The assumption that $v'(G) = 1$ implies that the taxation from the formal sector does not influence the marginal utility gain from increased public goods, so it does not influence the optimal monetary policy.

The assumption that buyers cannot use cash in the formal market owing to physical restriction also simplifies the analysis. This is because the outside option for not using credit in the formal meeting is no trade. The optimal credit tax, determined by the participation constraint, is independent of inflation. We discuss the case where formal trade accepts cash and a pure cash trade would not be recorded and taxed in the Appendix. We show that it provides another mechanism by means of which inflation policy influences welfare. As buyers and formal sellers can always successfully avoid tax by using money only, the credit tax cannot be too high to ensure voluntary participation. The value of using money decreases with high inflation, so by setting a high i, the government can collect not only more seigniorage revenue but also more credit tax. In the Appendix, we use a numerical example to show that social welfare might strictly increase with inflation under this assumption.

To sum up, the model provides some explanations for the high inflation and large informal sector phenomenon in developing countries. We show high inflation in some situations is the optimal monetary policy from a public finance perspective. We also show that the results hold with exogenous government spending, or with endogenous government spending.

2.5 Concluding Remarks

This paper points out the importance of the public finance motive behind inflation in the presence of a large informal sector in many developing countries. We use a monetary model to characterise the different payment methods that formal sector and informal sector can use and the different government regulation.

Along the same lines as in Chpater 1, we take the underlying lack of commitment friction seriously, and therefore means-of-payment are essential to facilitate trade. We highlight that there exists a record-keeping technology in the economy but the technology is imperfect. In Chapter 1, the imperfection is characterised by a probability that default behaviour cannot be successfully recognised and recorded. In this Chapter, we assume that only part of all trades, called formal trades, can be recorded, and there exists another type of trades, called informal trades, where no monitoring or record-keeping technology is feasible. In both Chapter, accessing the record-keeping technology are costly and buyers need to pay credit fee to use credit. But in Chapter 1 we focus more on the effect of costly record-keeping on social welfare and the optimal level of credit provision, and in this Chapter, we introduce public goods and discuss more on the optimal taxation and monetary policy.

Our main discussion is about the positive relationship between the size of the informal sector and the inflation rate. We show that the optimal monetary policy can be either deflationary, constant, or inflationary, depending on the size of the informal sector. We provide two theoretical explanation to account for the results, one from the trade-off between formal and informal sectors, and one from the trade-off between private and public sectors. Though the model simplifies the tax structure in the formal market, the details of financial services and credit market in developing countries, and the difference between formal economy and shadow economy, the results provide some implications that can inform an understanding of the coexistence of a large informal sector, high inflation, and the high seigniorage revenue phenomenon in developing countries.

2.6 Appendix A: if Buyers can Avoid Credit Tax by Using Money

In this part we discuss the optimal monetary policy if buyers are allowed to use cash to avoid tax in the formal trade. In this situation, in order to keep tractability, we assume that only those buyers who have matched a formal seller in DM1 can enter the informal market DM2, and have a chance to meet an informal seller. For those who have not matched with a formal seller, they stay unmatched in the following informal market in this period. By doing so, the buyer's equilibrium liquidity holding before DM2 is degenerate. [5]

When the credit limit is high, the buyer's problem is the same as in Section 4. But to avoid taxation, the outside option is different. As buyers could bring liquidity $\rho_f \bar{y_f}$ into the formal meeting and leave no record in the record-keeping system, but are subject to inflation tax. Assume $\theta = 1$ for simplicity, $\bar{y_f}$ solves $i = (1-\alpha)[u'_f(\bar{y_f}) - \rho_f]$. Suppose that the credit limit is high, $D \geq \rho_f \bar{y_f}$ hence credit equilibrium exists, buyers' participation constraint is:

$$\chi \leq \frac{(1-\alpha)\left(u_f\left(\frac{D}{\rho_f}\right) - D\right) - (-i\rho_f \bar{y_f} + (1-\alpha)[u_f(\bar{y_f}) - \rho_f \bar{y_f}])}{1-\alpha}, \tag{2.16}$$

where $-i\rho_f \bar{y_f} + (1-\alpha)[u_f(\bar{y_f}) - \rho_f \bar{y_f}]$ is the buyer's continuation value in a pure monetary equilibrium.

Government collects credit tax and seigniorage revenue to finance public good subject to the budget constraint (2.9) and participation constraint (2.16), and at optimal we have:

$$G = (1-\alpha)\left(u_f\left(\frac{D}{\rho_f}\right) - D\right) - (-i\rho_f \bar{y_f} + (1-\alpha)[u_f(\bar{y_f}) - \rho_f \bar{y_f}]) + \frac{i-r}{1+r}\rho_i \bar{y_i}. \tag{2.17}$$

Note that the high inflation rate now has two effects on the revenue. It not only has a non-monotonic Laffer Curve effect on the seigniorage revenue but also increases the credit tax levied on formal trade.

The effect of α on G is also mixed. Higher α decreases the value from formal trade and hence decreases the total credit tax. But as buyers can easily avoid tax by using money, such an effect can be relatively small, especially when the credit limit is low.

[5] As in Araujo and Hu (2018), another way to keep tractability is to keep tractability is to assume that the buyer's probability of meeting in DM1 is 1. Therefore, before the second market, all buyers are in the same situation with the same liquidity holding. However, this assumption means that the social optimal level of informality is $\alpha^* = 1$, which is not proper for our work.

Therefore, together with the positive effect of α on seigniorage, the overall effect can be positive. When total revenue increases with the size of the informal sector, the optimal inflation rate may decrease with α if government finances are exogenous G. In Figure 2.4 we use a numerical example to show that the total revenue increases with α and the optimal inflation rate decreases with α. The parameter values are $\eta = 0.5$, $G = 0.01$, $D = 0.65$, $r = 0.1$, $\rho_f = 1$ and $\rho_i = 3$.

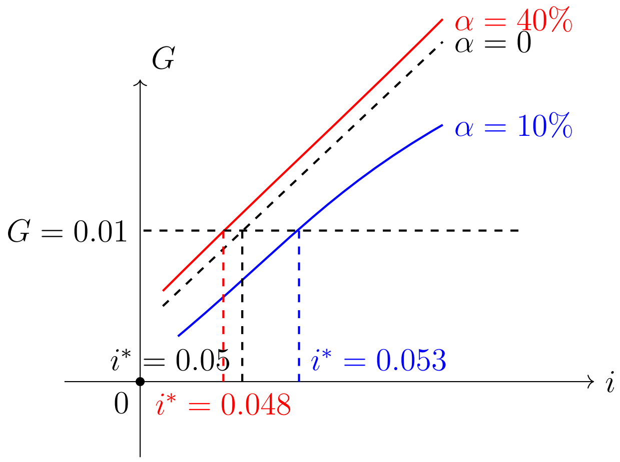

Figure 2.4: Government revenue G and nominal interest rate i

Suppose that G is endogenous. Social planner's problem is choosing the monetary policy γ and public good G to maximize social welfare:

(2.18)
$$\max_{\gamma,G}\left\{(1-\alpha)\left(u_f\left(\frac{D}{\rho_f}\right)-D\right)+\alpha[u_i(\bar{y}_i)-\rho_i\bar{y}_i]+v\Big\{(1-\alpha)\left(u_f\left(\frac{D}{\rho_f}\right)-D\right)\right.$$
$$\left.-(-i\rho_f\bar{y}_f+(1-\alpha)[u_f(\bar{y}_f)-\rho_f\bar{y}_f])+\frac{i-r}{1+r}\rho_i\bar{y}_i)\Big\}\right\},$$

and we let $v(G) = G$, the problem can be rewritten by choosing i:

$$\max_{i\geq 0}\left\{\alpha(u_i(\bar{y}_i)-\rho_i\bar{y}_i)-(-i\rho_f\bar{y}_f+(1-\alpha)[u_f(\bar{y}_f)-\rho_f\bar{y}_f])+\frac{i-r}{1+r}\rho_i\bar{y}_i\right\}. \tag{2.19}$$

Now the increases in the nominal interest rate have three effects on social welfare. It directly decreases the informal trade surplus, and indirectly influences the taxation and public good supply. The overall effect depends on the magnitude of the three effects, and in Figure 2.5 we provide a numerical example to illustrate the relationship between inflation and social welfare. The utility function take forms $u_j(y_j) = \sqrt{y_j}/0.5$ for $j = i, f$, $D = 0.65$, $r = 0.1$, $\rho_f = 1$ and $\rho_i = 3$. We show that the overall effect of inflation on social welfare can be positive when the size of the informal sector is small.

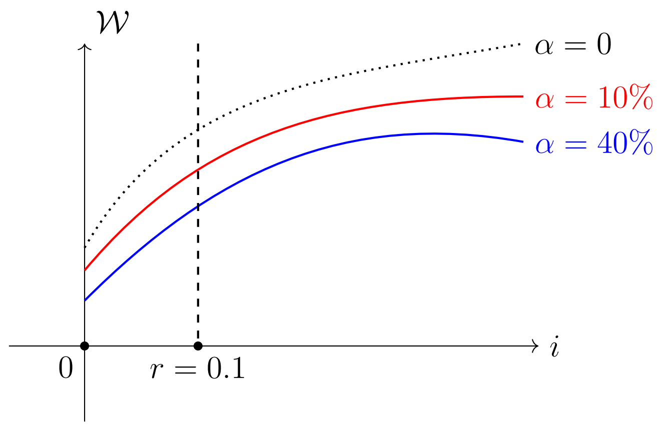

Figure 2.5: Social welfare and nominal interest rate when cash is accepted in DM1

Proof of Theorem 2.4.1

We denote the total revenue as $R(i; D, \alpha) \equiv (1-\alpha)[u(D/\rho_f) - D] + \frac{i-r}{1+r}\rho_i \bar{y}_i$. So we have

$$\begin{aligned} \frac{\partial R}{\partial i} &= \frac{\rho_i \bar{y}_i}{1+r} + \frac{i-r}{1+r}\rho_i \frac{\partial \bar{y}_i}{\partial i} \\ &= \frac{\rho_i \bar{y}_i}{1+r}\left[1 - \frac{i-r}{1-\eta}\frac{1}{\alpha+i}\right], \end{aligned} \tag{2.20}$$

where we use specific functional form $u(y) = y^\eta/\eta$ so $\bar{y}_i$ solves $i = \alpha[u'(\bar{y}_i) - \rho_i]$ with $\frac{\partial \bar{y}_i}{\partial i} = \frac{\rho_i}{\alpha u''(\bar{y}_i)} = \frac{-\bar{y}_i}{(1-\eta)(\alpha+i)}$. It is easy to see that $\frac{\partial R}{\partial i} \geq 0$ when $i \leq \frac{(1-\eta)\alpha + r}{\eta}$ and $\frac{\partial R}{\partial i} < 0$ otherwise. So total revenue first increases with i and then decreases with i, with the highest value $R(\frac{(1-\eta)\alpha+r}{\eta}; D, \alpha) \equiv \bar{G}$.

For any $G \leq \bar{G}$, there exist $i^* = i^*(D, \alpha)$ such that $R(i; D, \alpha) \geq G$ holds, and when $R(0; D, \alpha) > G$, $i^* = 0$. Note that i^* decreases with D as $\frac{\partial R}{\partial D} = (1-\alpha)[u'_f(D/\rho_f)/\rho_f - 1] > 0$. A sufficent condition for i^* increases with α is $\rho_i \geq 1/(1+r)$ as:

$$\begin{aligned} \frac{\partial R}{\partial \alpha} &= -[u_f(D/\rho_f) - D] + \frac{i-r}{1+r}\frac{i\bar{y}_i}{\alpha(1-\eta)(\alpha+i)} \\ &< -(u_i(\bar{y}_i) - \rho_i \bar{y}_i) + \frac{i-r}{1+r}\frac{i\bar{y}_i}{\alpha(1-\eta)(\alpha+i)} \\ &< -\frac{i}{\alpha}\bar{y}_i(\rho_i - \frac{1}{1+r}). \end{aligned} \tag{2.21}$$

The first inequality uses the assumption that welfare decreases with α, so $u_f(D/\rho_f) - D > u_i(\bar{y}_i) - \rho_i \bar{y}_i$. The second inequality uses the buyer's continuation value of bringing money to consume in DM2 is positive $-i\rho_i\bar{y}_i + \alpha[u_i(\bar{y}_i) - \rho_i\bar{y}_i] \geq 0$.

Proof of Lemma 2.4.1

By assuming $v(G) = G$ and assuming buyers make take-it-or-leave-it offer with $\theta = 1$, social planner's problem becomes:

$$\max_{i \geq 0} \left\{ \alpha[u_i(\bar{y}_i) - \rho_i \bar{y}_i] + \frac{i-r}{1+r} \rho_i \bar{y}_i \right\}, \tag{2.22}$$

where $\bar{y}_i$ solves $i = \alpha[u'(\bar{y}_i) - \rho_i]$ with $\frac{\partial \bar{y}_i}{\partial i} = \frac{\rho_i}{\alpha u''(\bar{y}_i)} < 0$. Take first order derivative with respect to i:

$$\frac{d\mathcal{W}}{di} = \frac{\rho_i}{\alpha u''(\bar{y}_i)} \left(i\rho_i + \frac{\alpha u''(\bar{y}_i)}{\rho_i} \frac{\rho_i \bar{y}_i}{1+r} + \frac{i-r}{1+r} \rho_i \right) \equiv \frac{\rho_i}{\alpha u''(\bar{y}_i)} f(i;\alpha), \tag{2.23}$$

and the sign of $\partial \mathcal{W}/\partial i$ depends on the sign of $f(i;\alpha)$. Plug in specific function form $u(y) = \frac{y^\eta}{\eta}, (0 < \eta < 1)$ we have $\bar{y}_i = (\frac{\alpha}{\alpha\rho_i + i\rho_i})^{1/1-\eta}$ and:

$$\begin{aligned} f(i;\alpha) &= i\rho_i + \frac{\alpha}{1+r}(\eta - 1) y_i^{\eta-1} + \frac{i-r}{1+r} \rho_i \\ &= \frac{\rho_i}{1+r} \left(i(1+r) + i - t - (1-\eta)(\alpha + i) \right), \end{aligned} \tag{2.24}$$

with $\frac{df(i;\alpha)}{di} = 1 + r + \eta > 0$ and the minimum value $f(0;\alpha) = \frac{\rho_i}{1+r}(-r - (1-\eta)\alpha) < 0$ and the maximum value $f(\infty;\alpha) = \infty$. So there exists unique $i^* = i^*(\alpha)$ that solve $f(i^*;\alpha) = 0$ such that when $i \leq i^*$, $d\mathcal{W}/di > 0$ and when $i > i^*$, $d\mathcal{W}/di < 0$.

Proof of Theorem 2.4.2

The optimal nominal interest rate solves $d\mathcal{W}/di = 0$ and with specific functional form, the solution is:

$$i^* = \frac{(1-\eta)\alpha + r}{1 + r + \eta},$$

and i^* increases with α. Let $\bar{\alpha}$ solves $i^* = r$ so $\bar{\alpha} = \frac{r(r+\eta)}{(1-\eta)}$. Note that $\bar{\alpha}$ increases with r with $\bar{\alpha} \leq 0.5$ if $r \leq \bar{r} = \frac{-\eta + \sqrt{\eta^2 + 2r^2(1-\eta)}}{2}$. So when $\alpha \leq \bar{\alpha}$, $i^* \leq r$ and when $\alpha > \bar{\alpha}$, $i^* > r$.

Chapter 3

Optimal Banking Regulation and Monetary Policy

3.1 Introduction

The serious interruption of the real economy from the global financial crisis of 2008 has given rise to a renewed interest in understanding the role of financial intermediaries and how to regulate them. At the same time, central banks' interest-rate policies are now implemented through interest on excess reserves in most advanced economies in environments with large reserve balances, while many developing economies struggle with high inflation. These new monetary policies also highlight their fiscal nature, as interest on reserves may be regarded as a form of subsidies to banks and to deposit holders. Here we propose a tractable model that focuses on the liquidity-provision role of banks in the form of deposits as means of payments. We use this framework to study the optimal monetary policy together with optimal banking regulations, in a flexible setting that allows for various fiscal capacities.

Our model of financial intermediaries features endogenous liquidity provision by introducing *banks* into a standard monetary model à la Lagos and Wright (2005) to maintain tractability. In our economy, banks are the only agents with the necessary expertise to manage and monitor loans to earn returns. However, banks' asset holdings require external financing, and banks may obtain this financing by issuing deposits which can serve as a means of payment for households. The usual frictions (lack of commitment and monitoring) in the Lagos-Wright environment render a means of payment essential, and this gives rise to an endogenous demand for bank liabilities.

We consider two main frictions in the banking sector. First, it is costly for banks to manage and monitor loans to entrepreneurs. There is an economy of scale in the sector by way of a fixed cost of bank operations and an increasing marginal cost of managing assets, which determine the potential profits to banks. Second, banks cannot fully commit to honour their future obligations; instead, they can only credibly pledge their assets that the court can seize upon bankruptcy. This friction constrains the amount of liquidity banks can provide and may prevent the first-best level of consumption for households to be achieved.

While most of the literature assumes that the central bank has unlimited access to lump-sum taxation to finance its policies, here we assume that the central bank has only limited fiscal resources, or, alternatively, it is required to hand in certain level of revenue to the fiscal authority. We have in mind the different fiscal capacities faced by different governments: for advanced economies who have better taxation institutions and face a relatively low interest rate that effectively allow for seigniorage revenue from the rest of the world, the fiscal requirement is low and the central bank may be subject to subsidies for its policies; in contrast, for developing economies that rely partially on domestic seigniorage for financing public expenditures, the fiscal requirement is high and the central bank is subject to revenue requirements.

Given this background environment, we consider three main policy instruments: reserve requirement, interest on excess reserves, and nominal asset creation rate. The first instrument is similar to the liquidity requirements in Basel III. The second instrument is how the central bank sets the interest-rate policy. The third is the traditional monetary policy and is closely related to inflation rate, according to the quantity theory of “money”. Given the fiscal requirement, however, the central bank cannot choose the three policies independently. Instead, for a given fiscal requirement upon the central bank, a positive interest on excess reserves would require real resources, which could be financed either by reserve creation or by fiscal transfers. However, in our pure “deposit” economy, the only channel through which seigniorage revenue can be generated is reserve requirements, under which banks are required to hold nominal assets.

In our setup, households demand deposits issued by banks to meet their liquidity needs for trades, which endogenously determines liquidity needs in the economy. The supply of deposits, however, is restricted by the amount of pledgeable assets that banks hold due to limited commitment. If there are abundant pledgeable assets, banks compete to increase the (real) interest rate on deposits to the point that households’ demand is satiated. Otherwise, liquidity is tight and deposit holdings are inefficiently low that prevent the first-best level of trades. Without intervention, a tight liquidity would also push up asset prices.

Both monetary policy and banking regulations affect banks' liquidity provision, however. Since reserves are fully pledgeable, these measures may increase liquidity provision by inducing banks to hold more reserves. Mandatory reserve requirements do so without the need of financing but affect bank profits. The proportional reserve requirement affects bank profits through the spread between the real deposit rate and banks' cost of equity, and is limited by that spread. The fixed reserve requirement would decrease bank profits directly by acting as a tax, and is limited by banks' profitability. In contrast, a positive (real) interest on excess reserves would induce banks to hold more excess reserves, but that would require fiscal resources to finance the interest payment.

Our main results characterize optimal monetary policy and banking regulations, depending on the fiscal requirement faced by the central bank. If that requirement is relatively low and if there are abundant pledgeable assets from the private sector, then it is optimal to have no regulations at all other than a potential fixed reserve requirement to generate seigniorage revenue, if needed. However, if the liquidity provision from the private sector is tight, then it is optimal to have both the fixed and proportional reserve requirements, together with a positive IOER. Assuming that the central bank maintains a constant inflation target, the optimal IOER would increase with productivity of the private sector and hence is pro-cyclical. These characteristics would mimic the patterns of monetary policy in the advanced economies.

In contrast, if the fiscal requirement for the central bank is high, the optimal policy never uses a the IOER, and banks do not hold excess reserves in equilibrium. Moreover, the amount of seigniorage revenue the central bank can generate is not monotonic in the productivity of the private sector. If the private sector is very productive and hence provides abundant pledgeable assets, the central bank can only generate seigniorage revenue by imposing a fixed reserve requirement with constant creation of more reserves. However, if the liquidity provision from the private sector is tight, then both the fixed and the proportional reserve requirements are feasible and allow the central bank to generate more seigniorage revenue. In both cases, the optimal inflation rate would be determined by the fiscal requirement. These characteristics seem consistent with many developing countries that feature persistent high inflation rates.

Related Literature

Our framework inherits features from two strands of the literature, and here we focus on the key differences and the novelty of our model. The first includes studies on financial intermediation and credit market frictions which build on, e.g., Kiyotaki and Moore (1997, 2001), and the second includes those that explicitly model the transaction role of means of payment using models that are based on, e.g., Lagos and Wright (2005) and Rocheteau and Wright (2005). Regarding the first strand, our pledgeability constraint is similar to that in Gertler and Kiyotaki (2010). There are two key differences, however. First, we explicitly model deposits as a means of payment and focus on banks' liquidity role for depositors. In particular, our model features an essential reserve requirement where reserves are fiat objects. Second, we focus on banks' limited commitment and optimal regulations to deal with both. Our approach to modeling bank assets as (one-period) Lucas trees effectively assumes that all agency issue between the banks and the end borrowers is captured by the management and monitoring cost, an approach shared by some recent papers such as Begenau and Landvoigt (2021).

Regarding the second strand, in Williamson (2012, 2016) both fiat money and deposits are used as means of payment, but he assumes that in some transactions deposits cannot be used and hence, money is valued even though it is dominated in the return. Our model integrates banking and fiat money (reserves) via reserve requirements, and demonstrates the essential role of both institutions for welfare. In addition, Williamson (2016) considers that bank deposits are backed by different fractions of long- and short-maturity assets and equity, subject to pledgeability constraints in the style of Kiyotaki and Moore (1997, 2001). Gu et al. (2013) consider an environment where bank deposits enable intertemporal exchange, but banks have limited commitment and deposits are unsecured. Monnet and Sanches (2015) study the incentive problem when banks' deposit issuance is not observable. In both papers, like ours, charter value (or future bank profit) is the crucial instrument proposed for dealing with limited commitment.[1] In contrast to papers in this strand, our model generates novel implications of monetary policy to optimal bank regulations and bank profits.

3.2 Environment

The environment is borrowed from Lagos and Wright (2005). Time is discrete and has an infinite horizon, $t \in \mathbb{N}_0$. The economy is populated by three sets of agents. The first set consists of a unit measure of *households*, the second consists of measure m of *banks*, and the third consists of a unit measure of *entrepreneurs*. Each date has two stages: the first stage has random pairwise meetings between households in a decentralized market (called the DM), and the second has a centralized market (called the CM) where all agents

[1] Keeley (1990) provides some evidence that charter value restricts banks' risk-taking behavior.

meet. In each DM, a household with a successful meeting can either be a *consumer* or a *producer*, and the probability that a household has a successful meeting and becomes a consumer is $\sigma \leq 1/2$, and the probability that a household has a successful meeting and becomes a producer is also σ. There is a single perishable good produced in each stage, with the CM good taken as the numéraire. Households' labels as consumers and producers depend on their roles in the DM, where only producers are able to produce and only consumers wish to consume. All agents except for banks can produce and consume in the CM.

A household's preference is represented by the following utility function

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t\left[u(q_t)-c(q_t)+x_t-h_t\right],$$

where $\beta \equiv (1+r)^{-1} \in (0,1)$ is the discount factor, q_t is DM consumption or production (depending on the household's role), x_t is CM consumption, and h_t is the supply of hours in the period-t CM. The first-stage utility functions, $u(q)$ and $-c(q)$, are increasing and concave, twice differentiable, and $u(0) = c(0) = 0$. The surplus function, $u(q) - c(q)$, is strictly concave, with $q^* = \arg\max\left[u(q) - c(q)\right]$. Moreover, $u'(0) = c'(\infty) = \infty$ and $c'(0) = u'(\infty) = 0$. All households have access to a linear technology to produce the CM good from their own labour, $x = h$. In the DM there are limitations in commitment, enforcement and record keeping, and households are anonymous. Thus, means of payment are necessary to facilitate DM trades.

There is only one real asset, projects from entrepreneurs. Each entrepreneur is endowed with $\bar{A}$ projects, which materialize within a single period, and each unit has a gross return τ (in terms of the CM good) that pays off in the next CM. The entrepreneur needs at least $\underline{\phi}$ units of CM goods for each unit of the project. Banks have no capacity to produce in any market, but they have the required expertise to perform costly monitoring and management to receive the return from these projects, and, to do so, a bank has to become active and purchase these projects from a competitive market. For simplicity we assume that even the entrepreneurs themselves cannot obtain the returns without banks, and that they simply sell their projects to banks.[2] Since each unit of project requires at least $\underline{\phi}$ units of CM good, this sets the minimum for the project price, denoted by ϕ (in terms of CM goods).

[2]As in Gertler and Kiyotaki (2010), one can also think of the bank's claim on these projects as equity. We could have assumed that the projects require input from CM goods that can only come from external financing and then considered formal loan contracts between banks and entrepreneurs. However, this will complicate the exposition without adding any insights.

On the liability side, each bank can issue deposit certificates to finance its asset purchase in the open market. We assume that these certificates are perfectly divisible and cannot be counterfeited. Such liabilities are payable to the bearer. Thus, households may use such certificates to finance their consumptions in the DM. There is a public record of banks' liabilities and asset holdings, but there is no record keeping of households' deposit holdings and their transaction records.[3]

There are two frictions associated with this financial intermediation. The first friction is the cost associated with managing and monitoring the projects.[4] Only active banks can hold assets and issue deposits; to become active, a bank has to pay a fixed cost of γ each period. There is also a cost of asset management: for an active banker to hold a projects, he needs to pay $\psi(a)$ to monitor and manage them. We assume that $\psi(0) = 0$, $\psi(a)$ is twice differentiable, strictly increasing, and strictly convex. Second, banks have limited liability and cannot commit to future actions, except for pledging the returns from their project holdings to repay their deposits.[5] Banks maximize their life-time profits with discount factor β.

Finally, we discuss social welfare in our economy. An allocation consists of the quantity of goods traded in each successful DM meeting, denoted by q. Given an allocation q, the total welfare is given by

$$\mathcal{W}(q,m) = \underbrace{\sigma[u(q) - c(q)]}_{(a)} - \underbrace{\left[m\psi(\bar{A}/m) + m\gamma\right]}_{(b)}, \tag{3.1}$$

in which term (a) evaluates efficiency in DM production, q, and term (b) evaluates efficiency in asset management that depends on m.[6] Thus, the first-best allocation, q^*, maximizes term (a) , according to the FOC's $u'(q^*) - c'(q^*) = 0$. Note that since we take m as given, it is not a choice here. However, as a benchmark, note that term (b) is maximized at m^*, according to the FOC $\Psi(\bar{A}/m^*) = \gamma$, where

$$\Psi(a) = \psi'(a)a - \psi(a). \tag{3.2}$$

[3]A historical example of this deposit claim is banknotes, and a modern counterpart is stored-value cards issued by banks. We can also introduce record keeping of households' deposits holdings and there will be room for credit card issuance, but that would complicate the analysis without affecting our main results.

[4]This assumption is consistent with the delegated monitoring model of Diamond (1984) and Williamson (1986). However, given our focus on liquidity provision by banks, we do not model the agency problem between banks and entrepreneurs in detail.

[5]Assuming partial pledgeability of the bank's project holdings does not affect the main results.

[6]We could have included the returns from the projects, $\bar{A}\tau$, in total welfare, but it would not affect the entry decisions of banks and allocations.

Here we define the fundamental prices of the assets. If the households can "hire" the banks to manage the assets without any friction, the banks would maximize

$$\Pi(a;\phi) = -\phi a - \gamma - \psi(a) + \beta\tau a.$$

The optimal asset holding then solves $-\phi+\beta\tau = \psi'(a)$. By substituting this FOC back into $\Pi(a;\phi)$, we obtain banks' profits under optimal asset holding of a projects as $\Psi(a)-\gamma$. For the given measure of banks, m, market clearing requires $\phi(m) = \beta\tau - \psi'(\bar{A}/m)$, and each active bank's profit is given by $\Psi(\bar{A}/m) - \gamma$. The equilibrium asset price is then

$$\phi^*(m) = \frac{\tau}{1+r} - \psi'\left(\frac{\bar{A}}{m}\right), \tag{3.3}$$

which can be regarded as the *fundamental value* of projects. To avoid uninteresting cases, we assume that

$$0 < \underline{\phi} < \phi^*(m). \tag{3.4}$$

The assumption (3.4) ensures that the projects are feasible when traded at the fundamental price.

Monetary Policy and Banking Regulations

In this section we lay out the monetary policy and banking regulations in our model. There is only one nominal asset, bank reserves. We assume that the central bank has no control over the fiscal policy but is required to make a transfer to the fiscal authority of an amount T (in terms of the CM good per household) each period. Note that T can be negative, which means that the central bank receives a transfer. The bank reserves are book-entry accounts at the central bank. The central bank pays a nominal interest rate on excess reserve holdings, a rate denoted by ι_g.

We consider only stationary equilibrium and hence the real value of reserves is constant over time. We use M_t to denote the nominal amount of reserves at the end of period-t CM and φ_t to denote the price of reserves in terms of CM good in period-t CM, and hence $z_t \equiv \varphi_t M_t$ is the real balances at the end of period-t CM. In terms of banking regulations, there are two reserve requirements, fixed and proportional. The first requires banks to hold a fixed amount of reserves and the second require banks to hold at least reserves that are η proportion to their deposit issuance in terms of the real value. The bank reserve is a nominal liability of the regulator, and banks can fully pledge their reserve holdings. The central bank also controls the growth rate of nominal assets, denoted by π.

We can then write the regulator's budget constraint. The regulator has to finance the interest on excess reserves by seignorage revenue, and the transfer can be either revenue (when negative) or additional expense (when positive). Thus, the budget balancedness requirement for period-$(t+1)$ CM is given by

$$\iota_g M_t^e \varphi_{t+1} + T \leq \pi(M_t^e + M_t^f + M_t^p)\varphi_{t+1}, \tag{3.5}$$

where M_t^e is the amount of excess reserves, M_t^f and M_t^p are the amounts of required fixed and proportional reserves, respectively, at the end of period-t CM. Note that we evaluate them at the price in period-$(t+1)$ CM as the interest on excess reserves is paid then. Expressed in real terms, we have

$$\frac{\iota_g}{1+\iota_g} Z_e + T = \pi \left\{ \frac{1}{1+\iota_g} Z_e + Z_f + Z_p \right\}, \tag{3.6}$$

where $Z_e \equiv (1+\iota_g)M_t^e\varphi_{t+1}$, $Z_f \equiv M_t^f\varphi_{t+1}$, and $Z_p \equiv M_t^p\varphi_{t+1}$, the corresponding real values for excess reserves, required fixed reserves and required proportional reserves, respectively.

In sum, taking T as given, the regulator has control over (ι_g, π) as monetary policy and (η, z_f) as banking regulation, where z_f is the fixed reserve requirement expressed in real terms for each bank. Besides the budget constraint (3.6), the regulator also has to respect to equilibrium conditions, which we turn next.

3.3 Equilibrium Allocations

In this section we consider equilibrium bank contracts with households under a given set of monetary policy and regulations. We first describe the time line and the general characteristics of the bank contracts.

The course of events. In the period-t CM, the course of events is as follows:

1. banks settle obligations to holders of deposit certificates issued in period $t-1$;
2. banks issue deposit contracts, promising a gross return R_t (in exchange for CM good);
3. banks buy projects in the competitive market at price ϕ_t (in terms of CM good).

We use d to denote the total amount of deposits that the bank promises to give out in the next CM (and hence it will receive d/R_t in the current period-t CM). Note that there are two different spot markets in the CM—one for assets and the other for deposits. Because only banks can monitor and manage entrepreneurial projects, with no loss of generality we assume that households do not participate in the asset market.

In the DM, upon a successful meeting with a producer, the consumer makes a take-it-or-leave-it offer, (q, p), where q is the DM production from the producer and p is the amount of deposit transfer (in terms of the coming CM goods).

We restrict our attention to stationary equilibria, where goods exchanged in the DM and the real value of projects are constant over time; i.e., $q_t = q_{t+1}$ and $\phi_t = \phi_{t+1}$, for all t, and in what follows we drop the time subscript t. Moreover, Finally, stationarity also means that aggregate real balance of the serves is constant over time, and hence $\varphi_{t+1} = \varphi_t/(1+\pi)$. Note also that the real balances are measured in terms of their coming CM real value.

Here we determine the equilibrium allocations for a given set of policy parameters, $(\iota_g, \pi, \eta, z_f)$. Recall that banks can fully pledge their holdings of projects and reserves. This then gives rise to the following pledgeability constraint,

$$d \leq \tau a + z, \tag{3.7}$$

where a is the bank's asset holding, z the total reserve holding. Note that here we measure the reserves in real terms at the following CM.

To formulate reserve requirement, we use z_f to denote a bank's fixed reserve holding, z_p the required proportional reserve holding and z_e excess reserve holding with $z = z_f + z_p + z_e$. The reserve requirement is then

$$z_p \geq \max\{0, \eta(d - z_e - z_f)\}. \tag{3.8}$$

Note that (3.8) allows deposits held against excess reserves to be exempted from the reserve requirement.

Given R and ϕ, a bank's profit from holding a projects, issuing d deposits and holding required reserves z_p and z_f, and excess reserves z_e at period-t CM is given by

(3.9)

$$\begin{aligned}\Pi(a, d, z_e, z_p, z_f) &= \frac{d}{R} - \phi a - (1+\pi)(z_p + z_f) - \frac{(1+\pi)z_e}{1+\iota_g} - \gamma - \psi(a) + \beta\{\tau a + z - d\} \\ &= \beta\left\{s_d d + [\tau - (1+r)\phi]a - (1+r)[\psi(a) + \gamma] - s_m(z_p + z_f) - s_g z_e\right\},\end{aligned}$$

where

$$s_d \equiv \frac{1+r}{R} - 1, s_m = (1+\pi)(1+r) - 1, \ \ s_g = \frac{1+r}{R_g} - 1 \text{ and } R_g = \frac{1+\iota_g}{1+\pi}.$$

We have three spreads, s_d for (real) deposit rate, s_m for money, and s_g for IOER, all against the discount rate, r.

A bank chooses a, d, and z_e to maximize (3.9) subject to (3.7) and (3.8). Note that z_f is required whenever $d > 0$, and, given the choice of a and d, z_p is pinned down by (3.8) at equality whenever $s_m > 0$. Moreover, whenever $s_d > 0$, the constraint (3.7) is binding and it is optimal to issue deposits. Plug in these binding constraints, we obtain the profit

(3.10)

$$\Pi(a, d, z_e) = \beta \left\{ \frac{s_d - \eta s_m}{1-\eta}\tau a + [\tau - (1+r)\phi]a - (1+r)[\psi(a) + \gamma] + (s_d - s_m)z_f + (s_d - s_g)z_e \right\}.$$

Thus, it is optimal to have $a > 0$ (and hence $d > 0$) only if

$$\eta \le \frac{s_d}{s_m}, \tag{3.11}$$

which states that the profit margin of issuing deposits, s_d, must be greater than the marginal cost of holding reserves for issuing deposits, ηs_m. We use $A(\phi, s_d)$ to denote the asset demand that maximizes (3.10) for $s_d \ge \eta s_m$, which is determined by the FOC

$$\frac{s_d - \eta s_m}{1-\eta}\tau + [\tau - (1+r)\phi] = (1+r)\psi'(a). \tag{3.12}$$

The corresponding required proportional reserves, denoted by $z_p(\phi, s_d)$, is pinned down by $A(\phi, s_d)$ and (3.8). If $s_d < \eta s_m$, it is optimal not to issue any deposits and hold any reserves, i.e., $z_p(\phi, s_d) = 0$. If $s_d \ge \eta s_m$, both (3.7) and (3.8) are binding, and

$$z_p(\phi, s_d) = \frac{\eta}{1-\eta}\tau A(\phi, s_d). \tag{3.13}$$

Now we consider the demand for excess reserves, denoted by $z_e(s_d, s_g)$. From (3.10), if $s_d < s_g$, then $z_e(s_d, s_g) = 0$. If $s_d > s_g$, then $z_e(s_d, s_g)$ is infinite and this cannot be an equilibrium. Finally, if $s_d = s_g$, any level of excess reserve holding is optimal, and we should think of $z_e(s_d, s_g)$ as a correspondence. Thus, banks hold excess reserves in equilibrium only if $R = R_g$, that is, $\iota_g = \iota_d$, the nominal rate of deposits.

Now we turn to households' behavior. We use $V(d)$ to denote a household's continuation value upon entering period-t DM with deposit d, and use $W(d)$ to denote a household's continuation value upon entering period-t CM with deposit d, facing bank deposit returns R. Note that d is the *promised value* of the deposit in the coming CM.

Assuming that banks repay their deposit obligations (which will be the case in equilibrium), standard Lagos-Wright (2005) arguments show that $W(d) = d + W(0)$; that is, W is linear in d. Now we consider the household's DM problem. The consumer with deposit d facing a producer with deposit d' solves the following problem:

$$\max_{(p,q)} u(q) + W(d-p), \quad \text{subject to} \quad p \leq d, \ -c(q) + W(d'+p) \geq W(d'), \tag{3.14}$$

in which p is the transfer of deposits and q is the DM production for the consumer. Given the linearity of W, it is straightforward to see the solution to (3.14) is given by

$$\begin{aligned} q(d) &= c^{-1}(d) \quad \text{and} \quad p(d) = d, \ \text{if } d < c(q^*); \\ q(d) &= q^* \quad \text{and} \quad p(d) = c(q^*), \ \text{otherwise.} \end{aligned} \tag{3.15}$$

Given R, a household's deposit choice in the CM is given by

$$\max_{d \geq 0} -\frac{d}{R} + \beta\{\sigma\{u[q(d)] + W[d - p(d)]\} + (1-\sigma)W(d)\},$$

where $[q(d), p(d)]$ is given by (3.15). By linearity of $W(\cdot)$, we can simplify the problem into

$$\max_{d \geq 0} -\frac{d}{R} + \beta \left\{\sigma[u[q(d)] - c[q(d)]] + d\right\}, \tag{3.16}$$

with the FOC

$$s_d = \frac{\sigma\{u'[q(d)] - c'[q(d)]\}}{c'[q(d)]}. \tag{3.17}$$

Let $D(s_d)$, the deposit demand per household, be the solution to (3.17). Note that for any $s_d > 0$, $D(s_d)$ is uniquely determined; when $s_d = 0$, $D(s_d)$ is not pinned down but $D(s_d) \geq c(q^*)$. Without loss of generality we set $D(0) = c(q^*)$, the minimum value of $D(s_d)$ when $s_d = 0$. Then, $D(s_d)$ is continuous and strictly decreasing in s_d.

Equilibrium requires market clearing conditions for deposits and assets and bank participation constraint:

$$D(s_d) \leq \tau \bar{A} + m z_p(\phi, s_d) + m z_f + m z_e(\phi, s_d), \tag{3.18}$$

$$m A(\phi, s_d) = \bar{A}, \tag{3.19}$$

$$\Psi\left(\frac{\bar{A}}{m}\right) - \gamma - \frac{s_m - s_d}{1+r} z_f \geq 0, \tag{3.20}$$

with equality in (3.18) whenever $s_d > 0$. The right side of (3.18), $\tau\bar{A} + mz_p(\phi, s_d) + mz_e(\phi, s_d)$, is derived from binding (3.7) and the market-clearing condition (3.19). In (3.18), we have inequality because (3.7) does not have to bind when $s_d = 0$. In (3.20), the left-side is the optimal bank profit, derived from (3.10) by plugging in the optimal asset and reserve holdings, plus the market clearing condition for asset holding, (3.19), and it has to be nonnegative for banks to be active. For given size of banks m, the fixed reserve requirement z_f is an entry cost for banks to participate and it cannot be too high. When the banks size are subject to free entry $m = m^*$, the only entry cost has to be $z_f = 0$.

Now we characterize the equilibrium allocations, taking the primitives and the policies as given. We are particularly interested in how equilibrium allocation varies with the asset return, τ, and the policy $(\pi, \iota_g, \eta, z_f, m)$. Note that here the transfer T is taken as passive without taking the budget constraint (3.5) into account, which will be brought back when considering optimal policy.

Proposition 3.3.1. *Let π, ι_g, z_f and m be given. Assume that m is not too low so that $\phi \geq \underline{\phi}$ is not binding and assume z_f is now too high such that (3.20) is not binding. When banks are active, there is a unique equilibrium with equilibrium ϕ determined by*

$$\phi = \frac{\left(1 + \frac{s_d - s_m\eta}{1-\eta}\right)\tau - (1+r)\psi'\left(\frac{\bar{A}}{m}\right)}{1+r}. \tag{3.21}$$

Let

$$\tau^* = \max\left\{\frac{D(0) - mz_f}{\bar{A}}, 0\right\}.$$

1. *Suppose that $\tau \geq \tau^*$. Banks are active only if $\eta = 0$. In that case, equilibrium $q = q^*$, $s_d = 0$ and $z_e = 0$.*

2. *Suppose that $\tau < \tau^*$. There exists a threshold $\bar{\eta} \in (0, \frac{s_g}{s_m})$ such that banks are active only if $\eta \leq \bar{\eta}$. We have two subcases.*

    a) *Suppose that*

    $$\tau \geq \hat{\tau} \equiv \max\left\{\frac{\frac{s_m - s_g}{s_m}\left[D(s_g) - mz_f\right]}{\bar{A}}, 0\right\}.$$

    For any $\eta \in [0, \bar{\eta}]$, there exists $\underline{\eta}$ such that

    - *for all $\eta > \underline{\eta}$, equilibrium $s_d < s_g$ and s_d strictly decreases with η, $z_e = 0$;*
    - *for all $\eta \leq \underline{\eta}$, equilibrium $s_d = s_g$, $z_e > 0$.*

    b) *Suppose that $\tau < \hat{\tau}$. Then $\bar{\eta} = s_g/s_m$, and for all $\eta \in [0, \bar{\eta}]$, equilibrium $s_d = s_g$ with $z_e > 0$.*

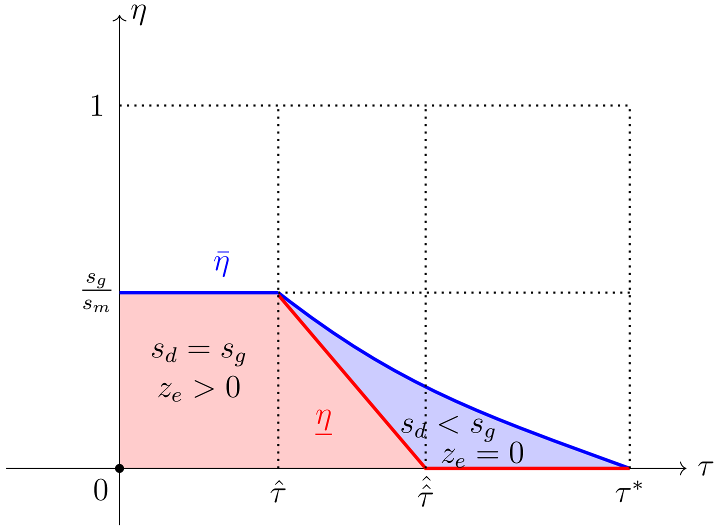

Figure 3.1: Different equilibrium given (π, ι_g, z_f, m) and $mz_f \leq D(s_g)$

Proposition 3.3.1 gives a full characterization of the monetary equilibrium for a given policy, $(\tau, \iota_g, \eta, z_f, m)$. The equilibrium spread of the deposit rate depends on the policy through the overall liquidity supply on the right-hand side of (3.18). According to Proposition 3.3.1 (1), when τ is sufficiently large ($\tau \geq \tau^*$), the equilibrium DM trade is at the first best level and depositors enjoy real interest rate equal to r. The intuition is that in this range there is plenty of liquidity around so that banks bid up interest on deposits to the point that the opportunity cost of holding deposits is zero. For this range of τ's, *any* reserve requirement can hurt the economy: it will make banking unprofitable and suppress the liquidity. Intuitively, proportional reserve requirement means that any deposit issuance requires the bank to pay the cost of holding reserves, but when $s_d = 0$ or $r_d = r$, banks make no profit from deposit issuance and hence any positive η will drive banks out of business. Note also that in this range the interest on reserves is not effective: banks already supply abundant liquidity that no excess reserves are needed. The only possible reserve holding is due to a strict charter system such that banks earn positive profit to pay for the fixed reserve requirement.

For the lower range of τ's, Proposition 3.3.1 (2) gives an upper bound on η, $\bar{\eta}$, above which banks are inactive. We have seen that for $\eta > s_d/s_m$, holding deposits is not optimal, and indeed $\bar{\eta} = s_d/s_m$ for τ sufficiently low. However, as depicted in Figure 3.1, this threshold $\bar{\eta}$ is lower than s_g/s_m for $\tau > \hat{\tau}$ and strictly decreases with τ. Intuitively, a higher τ means more liquidity supply and hence a higher interest on deposits. When η is too high, the proportional reserve requirement implies the associated proportional tax to issue deposits is too high for banks to be profitable. For η below $\bar{\eta}$, however, banks are always active, and equilibrium interest on deposits is below r, or, equivalently, $s_d > 0$.

Proposition 3.3.1 (2a) show that for $\tau > \hat{\tau}$, there is another threshold $\underline{\eta} < \bar{\eta}$ above which s_d is strictly in between 0 and s_g, and banks do not hold excess reserves. Intuitively, in this range, banks have intermediate liquidity supply supplemented by the proportional reserves required so that the interest on deposits is higher than interest on reserves and hence no banks hold excess reserves in equilibrium. Accordingly, for η below $\underline{\eta}$, the amount of liquidity provided by pledgeable assets and the corresponding required proportional reserves plus the required fixed reserves does not meet the need from the depositors at the interest rate on reserves, and excess reserves are held by the banks to meet that gap. Finally, for τ below $\hat{\tau}$, this always happens for η below $\bar{\eta}$ and hence equilibrium interest on deposits is determined by interest on reserves and banks hold excess reserves.

3.4 Optimal Banking Regulations

Now we move on to discussing the optimal design of the monetary policy, reserve requirement, pledgeability constraint and the charter system from the central bank's perspective. The central bank needs to pay the interest on excess reserve to banks and also pay a fixed amount of transfer T in terms of CM good to the government ($T < 0$ means central bank receive fiscal transfer from the central bank). The total expense is financed by inflation tax and the tax base is bank's proportional and fixed reserve holding. The central bank budget constraint is:

$$m\frac{r - s_g}{1+r}z_e + T \le \frac{s_m - r}{1+r}\frac{\eta}{1-\eta}\tau\bar{A} + \frac{s_m - r}{1+r}mz_f. \tag{3.22}$$

Note that for given T and m, the optimal social welfare (3.1) is achieved when buyers are subject the lowest cost of using deposit s_d, or equivalently, as indicated by Proposition 3.3.1, the total liquidity supply in the deposit market is the highest. Now we assume that T and m is given, and the central bank's optimal policy can be expressed by the following lemma:

Lemma 3.4.1. *Let m and T be given. For a given s_d, the optimal policy $(\eta, s_m, s_g, z_e, z_f)$ that solves*

$$\mathcal{S}(s_d; m, T) = \max_{\eta, s_m, s_g, z_e, z_f} \frac{\tau\bar{A}}{1-\eta} + mz_e + mz_f,$$

subject to $s_g \ge s_d$, (3.11), (3.20) and (3.22), is given by $\eta = s_d/s_m$, $s_g = s_d$, $z_f = \frac{1+r}{s_m - s_d}\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]$. $s_m \ge r$ and $z_e \ge 0$ are jointly determined by (3.22).

The the optimal value is given by

$$\mathcal{S}(s_d; m, T) = \frac{r}{r-s_d}\tau\bar{A} - \frac{1+r}{r-s_d}T + \frac{1+r}{r-s_d}m\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]. \tag{3.23}$$

Lemma 3.4.1 implies that the optimal liquidity provision does not depend on the inflation rate, but only on the transfer from the fiscal authority, T and the charter system m. The reason is high inflation rate directly increases the seigniorage revenue from bank's proportional reserve and fixed reserve holding, but it decreases the tax base. Overall, these two effects cancel out. The decrease of proportional reserve holding is due to the reserve requirement η decreases as inflation increases. Higher inflation means that for one unit of deposit contract, the required proportional reserve now is subject to higher marginal cost. To ensure bank are active, the optimal reserve requirement η decreases. The same trade-off also applies on fixed reserve requirement. Higher inflation directly increases bank cost of holding fixed reserve and bank earns less profit, so the optimal fixed reserve requirement decreases to ensure bank participates. Overall, higher inflation has no effects on liquidity provision.

But Lemma 3.4.1 implies that the monetary policy s_m and the excess reserve holding z_e should be designed to make sure the fiscal transfer T can be financed. We will discuss the optimal monetary policy in detail later. Before it, we first discuss the equilibrium existence condition.

For any given m and T, under the optimal policy in Lemma 3.4.1, the equilibrium s_d is determined by

$$D(s_d) \leq \frac{r}{r-s_d}\tau\bar{A} - \frac{1+r}{r-s_d}T + m\frac{(1+r)\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]}{r-s_d}, \tag{3.24}$$

with equality if $s_d > 0$. The equilibrium condition (3.24) uniquely determines the optimal spread s_d, denoted by $s_d(T, m)$. The next theorem describes the range of T for an equilibrium to exist.

Theorem 3.4.1. *Let $m \leq m^*$ be given. With the optimal policy, an equilibrium exists if there exists s_m, T and $z_e \geq 0$ such that (3.22) and (3.24) are satisfies. Specifically, there exist $\tilde{T}$ and $\underset{\sim}{T}$ such that*

(1) For $T \leq \underset{\sim}{T}$, equilibrium $s_d(T, m) = 0$ and $q = q^*$.

(2) For $\underset{\sim}{T} < T \leq \tilde{T}$, equilibrium $s_d(T, m) > 0$ and $q < q^*$.

(3) For $T > \tilde{T}$, no equilibrium exists.

Theorem 3.4.1 implies that when T is too high, there is no equilibrium because T is too costly to be financed or the liquidity supply is too low. When T is lower, equilibrium exists, and when T is low enough or negative, the equilibrium is first-best. The equilibrium for different T and τ is described in Figure 3.2.

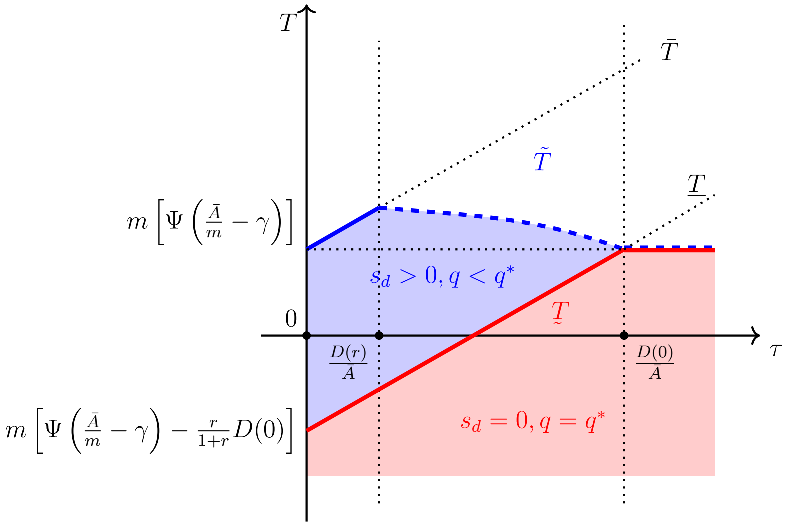

Figure 3.2: Equilibrium (s_d, q) for given $m \leq m^*$

When $\tau \geq \frac{D(0)}{\bar{A}}$, for any $T < \tilde{T} = m[\Psi(\bar{A}/m) - \gamma]$, the only equilibrium is first-best. Though the optimal proportional reserve requirement is $\eta = 0$, the central bank could still use the seignorage revenue taxed on the fixed reserve requirement to finance some $T > 0$. The total tax income is constrained by bank size m, and if central bank let the charter system $m = m^*$, then banks make no profit and pay no reserve in equilibrium, so there is no room for any fiscal payment.

When $\frac{D(r)}{\bar{A}} \leq \tau < \frac{D(0)}{\bar{A}}$, depending on T the economy can still be first best and pay a low fiscal payment or receive fiscal transfer. If T is too high, the central banks set positive reserve requirement and rely on the proportional reserve to finance T. But note that the highest $\tilde{T}$ is constrained by the budget constraint, and is lower than $\bar{T}$, meaning that the liquidity supply is plentiful to sustain a higher fiscal policy, but there is no enough inflation tax income to finance it.

When $\tau < \frac{D(r)}{\bar{A}}$, the upper bound of $T = \bar{T}$ such that the for any fiscal payment that is not too high to ensure positive liquidity supply, there always monetary policy to finance it.

To determine the optimal monetary policy, (π, ι_g), note that there is indeterminacy as (3.24) does not depend on π. However, we can either determine the minimum π and ι_g, or, alternatively, determine ι_g for a given inflation target $\bar{\pi}$.

Theorem 3.4.2. *(Minimum Inflation) Let $m \leq m^*$ be given. The minimum monetary policy, (π, ι_g) is characterized as follows.*

(1) Suppose that $T \leq 0$. Then minimum $\pi(T, m) = 0$ and minimum $\iota_g = \frac{1+r}{1+s_d(T,m)} - 1$ with

$$z_e = -\frac{1+r}{r - s_d(T,m)} \frac{T}{m} > 0. \tag{3.25}$$

(2) Suppose that $T > 0$, then minimum $z_e = 0$ with $\iota_g < \frac{(1+r)(1+\pi)}{1+s_d(T,m)} - 1$, and

(a) Suppose that $0 < T \leq \underset{\sim}{T}$, the minimum $\pi(T, m) = \frac{T}{(1+r)\left(m\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right] - T\right)} > 0$.

(b) Suppose that $\underset{\sim}{T} < T < \tilde{T}$,the minimum $\pi(T, m) = \frac{T}{D(s_d) - \tau \bar{A}} > 0$.

Theorem 3.4.2 discusses the minimum monetary policy to implement the optimal policy. Note that when $T \leq 0$, central bank receives positive fiscal transfer from the government to finance interest on excess reserve and does not rely on seigniorage revenue, so the minimum money growth rate is zero and the amount of excess reserve only depends on the size of fiscal transfer.

When $T > 0$, in order to finance fiscal payment, central bank set a positive money growth rate and use seignorage revenue to finance T. The tax base of seigniorage depends on the total reserve that banks save in the central bank. When asset is plentiful and the fiscal payment is not too high ($0 < T \leq \underset{\sim}{T}$), the optimal deposit provision is sufficient to achieve the first-best level and equilibrium $s_d = 0$, and hence, any proportional reserve requirement hurts the economy. Therefore, in this case, banks only hold fixed reserve and the total seigniorage base is $m\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]$. When the fiscal payment requirement is high $\underset{\sim}{T} < T < \tilde{T}$, at optimal $s_d > 0$, which allow central bank sets positive proportional reserve requirement and the total reserve includes both proportional reserve and fixed reserve. But in both case, the minimum monetary policy is inflationary, and the minimum excess reserve is 0.

Now we discuss the equilibrium outcome for a given inflation target $\bar{\pi} > \pi(T, m)$.

Theorem 3.4.3. *(Inflation Target) Let $m \leq m^*$ and $T < \tilde{T}$ be given. If central bank sets inflation target $\bar{\pi} > \pi(T, m)$, the optimal $\iota_g = \frac{(1+\bar{\pi})(1+r)}{1+s_d(T,m)} - 1$ with $\frac{\partial \iota_g}{\partial \tau} > 0$. The excess reserve holding is*

$$z_e = \frac{s_m - r}{s_m - s_d} \frac{s_d}{r - s_d} \frac{\tau \bar{A}}{m} + \frac{s_m - r}{s_m - s_d} \frac{1+r}{r - s_d} \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right] - \frac{1+r}{r - s_d} \frac{T}{m}. \tag{3.26}$$

where $s_m = (1 + \bar{\pi})(1 + r) - 1$ and $s_d = s_d(T, m)$.

Theorem 3.4.3 indicates that as long as the fiscal policy if not too tight, the central bank can always use inflation and seigniorage revenue to finance interest on reserve. The nominal interest rate on excess reserve increases with inflation rate, and also increases with the asset return. When the productivity of the private sector increases, banks have more pledgeable asset and increases the deposit supply. The equilibrium deposit spread decreases and the nominal interest on excess reserve is high.

3.5 Concluding Remarks

We have taken the liquidity role of banks seriously and derived the optimal banking regulations under different fiscal requirement. Our results demonstrate that when banks are subject to limited commitment friction and the productive assets are scarce, banking regulations, such as reserve requirement and interest on excess reserve policy, are optimal for welfare. Under those regulations, banks are required or encouraged to hold more pledgeable assets, and thus the total liquidity supply increases.

Compared to most of the literature, we have shown that considerations of liquidity provision can change some conventional wisdom about banking regulation. First, our model features three sets of policy instruments, a reserve requirement, interest on excess reserves, and nominal asset creation rate. We have shown that all three are essential to implement constrained efficient allocations. Second, our model emphasizes the welfare role of banks' liquidity provision and its implications for the optimal policy under different fiscal requirement. We show that for more advanced economy with fiscal subsidise to the central bank, low inflation but positive interest on excess reserve are optimal. In this case, a higher inflation rate will leads to more excess reserve holding, but does not improve the liquidity provision and welfare. When the central bank are subject to fiscal payment requirement, as in some developing countries, our work implies that a high inflation rate is optimal to generate seigniorage and banks hold no excess reserve.

Our results are robust to a few assumptions that we made. First, we can introduce heterogeneous assets, both in terms of pledgeability and in terms of maturities, and our main conclusions would remain the same. In such an extension, one can apply our methodology to study optimal banking regulations and how they affect the optimal security design of bank portfolios. Second, the take-it-or-leave-it offer in the DM can be replaced by Kalai bargaining or even the optimal mechanism as in Hu et al. (2009), without affecting any of the results qualitatively. Third, while households are anonymous in our model to simplify the analysis, we can allow banks to keep track of households' deposit holding and spending. This would not change our main results on banking regulation, but adding the ability to track the households' trade histories would make our framework a natural model for credit card issuance and welfare analysis. Indeed, in such a world, it would be optimal to allow households with extra liquidity needs to issue unsecured credit, with future exclusion from banking services as a threat to induce repayment.

3.6 Appendix: Proofs of Lemmas and Propositions

Proof of Proposition 3.3.1

Define a lower bound of m as $\underline{m}$ that solves $\frac{1}{1+r}[\tau-(1+r)\psi'(\bar{A}/\underline{m})] = \underline{\phi}$. Note that $m \geq \underline{m}$ is a sufficient condition for $\phi \geq \underline{\phi}$ because in equilibrium, banks are active only if $s_d \geq s_m\eta$, so the equilibrium asset price $\phi = \frac{\left(1+\frac{s_d - s_m\eta}{1-\eta}\right)\tau - (1+r)\psi'\left(\frac{\bar{A}}{m}\right)}{1+r} \geq \frac{1}{1+r}[\tau - (1+r)\psi'(\bar{A}/m)]$. As it is an increasing function on m, so for any $m > \underline{m}(\tau)$, the equilibrium price $\phi > \underline{\phi}$.

The banks participates if and only if the fixed reserve requirement is not too high and bank makes nonnegative profit $z_f \leq \frac{1+r}{s_m - s_d}[\Psi(\bar{A}/m) - \gamma]$. As the s_d is either constant or strictly decreasing in z_f, the RHS is an decreasing function on z_f, so there exists a unique upper bound $\bar{z_f}$ such that for $z_f \leq \bar{z_f}$, banks earn nonnegative profit and participate in the deposit market.

The threshold $\bar{\eta}$ is determined by $D(\bar{\eta}s_m) = \frac{\tau\bar{A}}{1-\bar{\eta}} + mz_f$ and decreases with τ as $\frac{\partial\bar{\eta}}{\partial\tau} = \frac{\bar{A}}{1-\bar{\eta}}\frac{1}{D'(\bar{\eta}s_m)s_m - \tau\bar{A}/(1-\bar{\eta})^2 - m'(s_m\bar{\eta})s_m} < 0$ with $\bar{\eta} = 0$ when $\tau = \tau^*$.

The threshold $\underline{\eta}$ is determined by $D(s_g) = \frac{\tau\bar{A}}{1-\underline{\eta}} + mz_f$ if $D(s_g) - m(s_g)z_f \geq 0$ and $\underline{\eta} = 0$ otherwise. When $\underline{\eta} \geq 0$, it strictly decreases and when $\tau = \hat{\tau}$, $\underline{\eta} = s_g/s_m$ and when $\underline{\eta} = 0$, we define $\tau = \hat{\hat{\tau}} = D(s_g) - mz_f$.

Proof of Lemma 3.4.1

We solves the optimal $(\eta, s_m, s_g, z_e, z_f)$ step by step. First, note that the for given s_m, η and z_f, the RHS of budget constraint (3.22) is given and at optimal there is no waste of seigniorage revenue, so the budget constraint is binding. We plug in mz_e from (3.22) into $\mathcal{S}(s_d; m, T)$ and get:

$$\begin{aligned}\mathcal{S}(s_d; m, T) &= \frac{\tau}{1-\eta}\bar{A} + mz_e + mz_f \\ &= \frac{1}{1-\eta}\tau\bar{A} + \left(\frac{s_m - r}{r - s_d}\frac{\eta}{1-\eta}\tau\bar{A} + \frac{s_m - r}{r - s_d}mz_f - \frac{1+r}{r-s_d}T\right) + mz_f.\end{aligned} \tag{3.27}$$

Then note that the liquidity supply increases with η and also increases with z_f, so at optimal we have $\eta = \frac{s_d}{s_m}$ and $z_f = \frac{1+r}{s_m - s_d}\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]$ and plug into the liquidity provision we have

$$\mathcal{S}(s_d; m, T) = \frac{r}{r - s_d}\tau\bar{A} - \frac{1+r}{r-s_d}T + m\frac{1+r}{r-s_d}\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right].$$

The monetary policy doesn't not influence the supply of liquidity, as long as seignorage revenue could finance T, or:

$$T \leq \frac{s_m - r}{1+r}\frac{s_d}{s_m - s_d}\tau\bar{A} + m\frac{s_m - r}{s_m - s_d}\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right].$$

The RHS is an increasing function on s_m with RHS=0 when $s_m = r$, and RHS=$\frac{s_d}{1+r}\tau\bar{A} + m\left[\Psi(\bar{A}/m) - \gamma\right]$ when $s_m = \infty$. So for any $T < \frac{s_d}{1+r}\tau\bar{A} + m\left[\Psi(\bar{A}/m) - \gamma\right]$ there exists a unique lower bound $\underline{s_m} = \underline{s_m}(s_d; T)$ such that for any $s_m \geqslant \underline{s_m}$, the seignorage revenue could finance T. When $T < 0$, we let $\underline{s_m} = r$.

The corresponding optimal excess reserve holding z_e for given other policies just solves the budget constraint (3.22) at equality. For any T, if $s_m = \underline{s_m}$, $z_e = 0$, and when $s_m > \underline{s_m}$, $z_e = \frac{s_m - r}{r - s_g}\frac{\eta}{1-\eta}\frac{\tau\bar{A}}{m} + \frac{s_m - r}{r - s_g}z_f - \frac{1+r}{r-s_g}\frac{T}{m} > 0$.

Proof of Theorem 3.4.1

Define $\bar{T} \equiv \frac{r}{1+r}\tau\bar{A} + m\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]$. When $T \geq \bar{T}$, the market clearing condition (3.24) never holds. When $T < \bar{T}$, the left-hand side of (3.24) is a decreases function of s_d from $D(0)$ to $D(r)$, and the right-hand side is an increasing function of s_d from $\tau\bar{A} - \frac{1+r}{r}T + m\frac{(1+r)}{r}\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]$ to ∞. Then define $\underline{T} \equiv \frac{r}{1+r}\left[\tau\bar{A} - D(0)\right] + m\left[\Psi(\bar{A}/m) - \gamma\right]$ such that when $T \leq \underline{T}$, (3.24) holds for any s_d, and when $T > \underline{T}$, (3.24) binds.

Given the two important thresholds, now we first discuss the condition such that first-best equilibrium exists, and then discuss the equilibrium where $s_d(T, m) > 0$.

If in equilibrium $s_d(T, m) = 0$, Lemma 3.4.1 implies that the optimal $\eta = 0$ $z_p = 0$ and $z_f = \frac{1+r}{s_m}\left[\Psi(\bar{A}/m) - \gamma\right]$. Such equilibrium exists if there exists T and s_m such that (3.22) and (3.24) satisfy $\mathcal{S}(0; T, m) \geq D(0)$ and $T \leq \frac{s_m - r}{s_m}m\left[\Psi(\bar{A}/m) - \gamma\right]$. Note that the seignorage income $\frac{s_m - r}{s_m}m\left[\Psi(\bar{A}/m) - \gamma\right]$ is an increasing function of s_m with the minimum value 0 when $s_m = r$ and the maximum value approaches $m\left[\Psi(\bar{A}/m) - \gamma\right]$ when $s_m \to \infty$. So we rewrite the two conditions as:

$$\text{(3.28)} \qquad \begin{cases} T \leq \frac{r}{1+r}\left[\tau\bar{A} - D(0)\right] + m\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right] = \underline{T}, \\ T < m\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right], \end{cases}$$

which has solution denoted by $T < \underset{\sim}{T} = \min\left\{\underline{T}, m\left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma\right]\right\}$.

Then we consider the equilibrium condition if $s_d(T, m) > 0$. In this case Lemma 3.4.1 implies that the optimal $\eta = s_d/s_m$ $z_p = \frac{s_d}{s_m - s_d}\tau\bar{A}$ and $z_f = \frac{1+r}{s_m - s_d}\left[\Psi(\bar{A}/m) - \gamma\right]$. In the equilirbium s_m and T should satisfies $0 < \mathcal{S}(0; T, m) < D(0)$ and the budget constraint

(3.24) which is

$$T \leq \frac{s_m - r}{1+r} \frac{s_d}{s_m - s_d} \tau \bar{A} + m \frac{s_m - r}{s_m - s_d} \left[\Psi(\bar{A}/m) - \gamma \right].$$

The right-hand side seignorage revenue is an increasing function of s_m with value 0 when $s_m = r$ and the maximum value approaches $m \left[\Psi(\bar{A}/m) - \gamma \right] + \frac{s_d}{1+r} \tau \bar{A}$ when $s_m \to \infty$. Again we rewrite the two conditions as:

(3.29)

$$\begin{cases} \underline{T} = \frac{r}{1+r}(\tau \bar{A} - D(0)) + m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] < T < \frac{r}{1+r} \tau \bar{A} + m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] = \bar{T} \\ T < m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] + \frac{s_d}{1+r} \tau \bar{A}, \end{cases}$$

where condition (3.22) holds for any $T \in (\underline{T}, \bar{T})$, and we will show that there exists unique $\bar{\bar{T}} \in (\underline{T}, \bar{T})$ such that (3.24) holds if and only if $T \leq \bar{\bar{T}}$. We prove it by rearranging the second equation with $D(s_d) = \mathcal{S}(s_d; T, m)$:

$$m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] + \frac{s_d}{1+r} \tau \bar{A} - T = \frac{r - s_d}{1+r} (D(s_d) - \tau \bar{A}), \tag{3.30}$$

and the sign depends on the sign of $D(s_d) - \tau \bar{A}$ which is a decreasing function of T. When $T \to \underline{T}$, $s_d \to 0$ so $D(s_d) - \tau \bar{A} \to D(0) - \tau \bar{A}$; when $T \to \bar{T}$, $s_d \to r$ so $D(s_d) - \tau \bar{A} \to D(r) - \tau \bar{A}$.

Note that when $\tau \geq D(0)/\bar{A}$, for any $T \in (\underline{T}, \bar{T})$ (3.30) is always negative, so there does not exists any T such that budget constraint $T < m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] + \frac{s_d}{1+r} \tau \bar{A}$ holds. We let $\tilde{T} = \underline{T}$ when $\tau \geq D(0)/\bar{A}$.

When $\tau \leq D(r)/\bar{A}$, for any $T \in (\underline{T}, \bar{T})$ (3.30) is always positive, so budget constraint $T < m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] + \frac{s_d}{1+r} \tau \bar{A}$ always holds. So let $\tilde{T} = \bar{T}$ when $\tau \leq D(r)/\bar{A}$.

When $D(r)/\bar{A} < \tau < D(0)/\bar{A}$, there exists a unique $\bar{\bar{T}} \in (\underline{T}, \bar{T})$ such that $D(s_d) - \tau \bar{A} = 0$, and when $T < \bar{\bar{T}}$, $m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] + \frac{s_d}{1+r} \tau \bar{A} > T$, and when $T > \bar{\bar{T}}$, $m \left[\Psi\left(\frac{\bar{A}}{m}\right) - \gamma \right] + \frac{s_d}{1+r} \tau \bar{A} < T$. So we use $\tilde{T} \equiv \min \left\{ \bar{T}, \bar{\bar{T}} \right\}$ to denote the solution of (3.29), which is when $T < \tilde{T}$, there exists s_m such that $s_d(T, m) > 0$.

Proof of Theorem 3.4.2

We discuss that for given T, the minimum inflation rate such that T can be financed. When $T \leq 0$, there is no need to use inflation to finance it, so $\underline{s_m} = r$ and $\pi(T, m) = 0$. The corresponding reserve holding z_e at $\pi = 0$ solves (3.22) at equality: $z_e = -\frac{1+r}{r - s_d(T,m)} \frac{T}{m} > 0$.

When $0 < T < \tilde{T}$, we claim that there exists a unique lowest $\underline{s_m} > r$ for each given T and the minimum excess reserve $z_e = 0$. When $0 < T \leq \utilde{T}$, the $\underline{s_m}$ solves:

$$T = \frac{\underline{s_m} - r}{\underline{s_m}} \left[\Psi \left(\frac{\bar{A}}{m} \right) - \gamma \right] \text{ and } \pi(T, m) = \frac{T}{(1+r)\left(m \left[\Psi \left(\frac{\bar{A}}{m} \right) - \gamma \right] - T \right)} > 0$$

When $\tilde{T} < T < \tilde{T}$, $\underline{s_m}$ is determined by:

$$T = \frac{\underline{s_m} - r}{1+r} \frac{s_d}{\underline{s_m} - s_d} \tau \bar{A} + m \frac{\underline{s_m} - r}{1+r} \frac{1+r}{\underline{s_m} - s_d} \left[\Psi \left(\frac{\bar{A}}{m} \right) - \gamma \right],$$

and substitute the last part from (3.24), $m(1+r) \left[\Psi \left(\frac{\bar{A}}{m} \right) - \gamma \right] = (r - s_d) D(s_d) - r\tau\bar{A} + (1+r)T$. We rearrange and get

$$\underline{s_m} = r + \frac{(1+r)T}{D(s_d) - \tau\bar{A}} > r \text{ and } \pi(T, m) = \frac{T}{D(s_d) - \tau\bar{A}} > 0,$$

and note that $\tilde{T} \leq \bar{\bar{T}}$ ensures that $D(s_d) \geq \tau\bar{A}$.

Proof of Theorem 3.4.3

Note that for given $T < \tilde{T}$, from Theorem 3.4.2 we know that there exists finite lowest $\pi(T, m)$ such that T can be financed by seigniorage. Now as $\pi > \pi(T, m)$, excess reserve z_e solves the budget constraint (3.22) at equality.

Bibliography

[1] Aisen, A. and Veiga, F. J. (2006). "Does political instability lead to higher inflation? A panel data analysis, " *Journal of Money, Credit and Banking*, 1379-1389.

[2] Andolfatto, D.(2013). "Incentive-feasible deflation, " *Journal of Monetary Economics*, 60, 383-390.

[3] Araujo, L. and Hu, T. W. (2018). "Optimal monetary interventions in credit markets, " *Journal of Economic Theory*, 178, 455-487.

[4] Aruoba, S. B. (2021). "Institutions, tax evasion, and optimal policy, " *Journal of Monetary Economics*, 118, 212-229.

[5] Baklouti, N. and Boujelbene, Y. (2019). "The economic growth–inflation–shadow economy trilogy: Developed versus developing countries, " *International Economic Journal*, 33, 679-695.

[6] Begenau, J. and Landvoigt, T. (2021)."Financial regulation in a quantitative model of the modern banking system," working paper.

[7] Bethune, Z., Hu, T. W. and Rocheteau, G. (2018). "Indeterminacy in credit economies, " *Journal of Economic Theory*, 175, 556-584.

[8] Bethune, Z., Rocheteau, G. and Rupert, P. (2015). "Aggregate unemployment and household unsecured debt, " *Review of Economic Dynamics*, 18, 77-100.

[9] Chari, V. V., Christiano, L. J. and Kehoe, P. J. (1996). "Optimality of the Friedman rule in economies with distorting taxes, " *Journal of Monetary Economics*, 37, 203-223.

[10] Chiu, J., Dong, M. and Shao, E. (2018). "On the welfare effects of credit arrangements, "*International Economic Review,* 59, 1621-1651.

[11] Correia, I. and Teles, P. (1996)."Is the Friedman rule optimal when money is an intermediate good? " *Journal of Monetary Economics,*, 38, 223-244.

[12] Cukierman, A., Edwards, S. and Tabellini, G. (1992). "Seigniorage and political instability, " *American Economic Review*, 82, 537-555.

[13] Diamond, Douglas W. (1984) "Financial intermediation and delegated monitoring," *Review of Economic Studies* 51, 393-414.

[14] Dong, M. and Huangfu, S. (2021). "Money and costly credit, " *Journal of Money, Credit and Banking*, 53, 1449-1478.

[15] Gertler, M. and Kiyotaki, N. (2010) "Financial intermediation and credit policy in business cycle analysis, " *Handbook of Monetary Economics* Volume 3, 547-599.

[16] Gomis-Porqueras, P., Peralta-Alva, A. and Waller, C. (2014). "The shadow economy as an equilibrium outcome, " *Journal of Economic Dynamics and Control*, 41, 1-19.

[17] Gordon, R. and Li, W. (2009). "Tax structures in developing countries: Many puzzles and a possible explanation, " *Journal of Public Economics*, 93, 855-866.

[18] Gu, C., Mattesini, F. and Wright, R. (2016). "Money and credit redux," *Econometrica*, 84, 1-32.

[19] Hu, T. W., Kennan, J. and Wallace, N. (2009). "Coalition-proof trade and the friedman rule in the Lagos-Wright model," *Journal of Political Economy* 117, 116-137.

[20] Jiang, J. H. and Shao, E. (2020). "The cash paradox," *Review of Economic Dynamics,*, 36, 177-197.

[21] Kehoe, T. J. and Levine, D. K. (1993). "Debt-constrained asset markets," *The Review of Economic Studies*, 60, 865-888.

[22] Kimbrough, K. P. (1986). "The optimum quantity of money rule in the theory of public finance, " *Journal of Monetary Economics*, 18, 277-284.

[23] Kiyotaki, N. and Moore, J. (1997)."Credit cycle," *Journal of Political Economy* 105, 211-248.

[24] Kiyotaki, N. and Moore, J. (2001) "Liquidity, business cycles and monetary policy," manuscript.

[25] Koch, C., Richardson, G. and Van Horn, P. (2016). "Bank leverage and regulatory regimes: Evidence from the great depression and great recession," *American Economic Review: Papers and Proceedings* 106, 538-542.

[26] Kocherlakota, N. R. (1998). "Money is memory," *Journal of Economic Theory*, 81, 232-251.

[27] Kocherlakota, N. R. and Wallace, N.(1998). "Incomplete record-keeping and optimal payment arrangements, " *Journal of Economic Theory*, 81, 272-289.

[28] Koreshkova, T. A. (2006). "A quantitative analysis of inflation as a tax on the underground economy," *Journal of Monetary Economics*, 53, 773-796.

[29] Lotz, S. and Zhang, C. (2016). "Money and credit as means of payment: A new monetarist approach, " *Journal of Economic Theory,* 164, 68-100.

[30] Lagos, R. and Wright, R. (2005). "A unified framework for monetary theory and policy analysis," *Journal of Political Economy*, 113, 463-484.

[31] Mazhar, U. and Méon, P. G. (2017). "Taxing the unobservable: The impact of the shadow economy on inflation and taxation," *World Development*, 90, 89-103.

[32] Monnet, C. and Sanches, D. R. (2015). "Private Money and banking regulation," *Journal of Money, Credit and Banking* 47, 1031-1062.

[33] Phelps, E. S. (1973). "Inflation in the theory of public finance," *The Swedish Journal of Economics*, 67-82.

[34] Rocheteau, G. and Wright, R. (2005)."Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium," *Econometrica* 73, 175-202.

[35] Sanches, D. and Williamson, S. (2010). "Money and credit with limited commitment and theft, " *Journal of Economic theory*, 145, 1525-1549.

[36] Telyukova, I. A. and Wright, R. (2008). "A model of money and credit, with application to the credit card debt puzzle," *The Review of Economic Studies*, 75, 629-647.

[37] Wang, L., Wright, R. and Liu, L. Q. (2020). “Sticky prices and costly credit, ” *International Economic Review*, 61, 37-70.

[38] Williamson, S. D. (1986). “Costly monitoring, financial intermediation, and equilibrium credit rationing,” *Journal of Monetary Economics* 18, 159-79.

[39] Williamson, S. D. (2012). “Liquidity, Monetary policy, and the financial crisis: A new monetarist approach,” *American Economic Review* 102, 2570-2605.

[40] Williamson, S. D. (2016). “Scarce collateral, the term premium, and quantitative easing,” *Journal of Economic Theory* 164, 136-165.