FEDERAL RESERVE BANK of ST. LOUIS

RESEARCH DIVISION

Working Paper Series

Stabilization Policy at the Zero Lower Bound

Paola Boel
and
Christopher J. Waller

Working Paper 2015-027B
https://doi.org/10.20955/wp.2015.027

September 2015

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Stabilization Policy at the Zero Lower Bound*

Paola Boel — Sveriges Riksbank

Christopher Waller† — FRB St. Louis

October 29, 2018

Abstract

We construct a monetary economy in which agents face aggregate liquidity shocks and heterogeneous idiosyncratic preference shocks. We show that, in this environment, not all agents are satiated at the zero lower bound even when the Friedman rule is the best interest-rate policy the central bank can implement. As a consequence, there is scope for central bank stabilization policy, which takes the form of repo arrangements in response to aggregate demand shocks. We find such a policy temporarily relaxes the liquidity constraint of impatient agents without harming the patient ones, thus improving welfare even at the zero lower bound. Due to a pecuniary externality, the policy may also have beneficial general-equilibrium effects for the patient agents even if they are unconstrained in their holdings of real balances.

Keywords: Money, Heterogeneity, Stabilization Policy, Zero Lower Bound

JEL codes: E40, E50

---

*We thank participants in workshops and seminars at the Chicago Fed, St. Louis Fed, Riksbank, Central Bank of Costa Rica, UC Riverside, Deakin University, University of Sydney, Macquarie University, Arizona State University, SED Meetings, the African Search and Matching Workshop and the Vienna Macro Workshop. We especially thank Aleks Berentsen, Daria Finocchiaro, Pedro Gomis-Porqueras, Per Krusell, Jesper Lindé and Cyril Monnet. This paper was previously circulated under the title "On the Theoretical Efficacy of Quantitative Easing at the Zero Lower Bound." The views expressed are those of the individual authors and do not necessarily reflect official positions of Sveriges Riksbank, the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

†Corresponding author. Research Division, Federal Reserve Bank of St. Louis, St. Louis, MO 63166-0442, USA. Email: cwaller@stls.frb.org.

1 Introduction

After a major shock such as the 2007–08 financial crisis, central banks around the world drove the policy rate to zero. Upon doing so, many argued that the central bank is 'out of ammunition' to deal with subsequent, but less severe, shocks to the economy. In short, the central bank has no ability to conduct stabilization policy once the zero lower bound (ZLB) is hit. The basic argument of the ineffectiveness of monetary policy goes back to Keynes and the idea of a liquidity trap. Once the ZLB is hit, the opportunity cost of holding money is zero. So agents can be patient in disposing of any excess money balances. Consequently, any further injections of money will be held as idle balances. As a result, fiscal policy must play a more important role to stabilize subsequent shocks or the central bank must rely on more unconventional policies such as quantitative easing and forward guidance.[1]

An underlying assumption of the liquidity trap argument is that agents are homogeneous and they all face the same constraints. Most importantly, agents have the same degree of patience in terms of their willingness to hold liquid assets. If agents vary in their degree of patience, then the opportunity cost of holding liquid assets is not the same across individuals in society and impatient agents will make different portfolio decisions compared to patient agents. This pattern of portfolio differences is borne out by microeconomic evidence; Kaplan, Violante and Wedner (2014) show that up to 1/3 of U.S. households hold almost no liquid assets and face high borrowing costs. In such a world, at the ZLB, some agents may be unconstrained in their liquid asset holdings while others are still constrained. Furthermore, shocks to the economy can worsen these constraints for some agents in society.

Does this heterogeneity in liquid asset holdings give the central bank an opportunity to engage in stabilization policy in a beneficial way? The objective of this paper is address this question. To do so, we construct a New Monetarist model to explore the stabilization response of the central bank to aggregate shocks at the ZLB. In NM models, the key friction is limited commitment, which induces agents to hold liquid assets in order to engage in spot trade. A key change to the standard model is to introduce idiosyncratic shocks to agents discount factors. Consequently, some agents will be more patient than others in terms of their opportunity costs of holding liquid assets across time. More patient agents will hold larger quantities of liquid assets than impatient agents, which will affect their consumption in some trades.

[1] Christiano, Eichenbaum and Rebelo (2011) study the impact of fiscal policy at the ZLB while Eggertsson and Woodford (2003) examine the use of unconventional policies at the ZLB.

Furthermore, differences in impatience means that patient agents are willing to accept a lower interest rate on assets relative to patient agents. This has a dramatic impact on monetary policy. The central bank can only drive the interest rate to zero for the most patient agents in the economy—for all the others, they are still constrained by their money holdings even at the ZLB. The central bank would like to drive the nominal interest rate below zero to benefit the impatient agents but is constrained by the patient agents portfolio decisions. It is in this sense that the ZLB is a true constraint on monetary policy.

The severity of the liquidity constraints on impatient agents will vary with aggregate shocks to the economy. But due to the ZLB, the central bank cannot use traditional interest rate policy to respond to the shocks. We first examine how a passive monetary policy at the ZLB affects the real allocation of consumption. Doing so mimics a central bank that is "out of ammunition" and does not respond to shocks while operating at the ZLB. We show that for small "aggregate demand" shocks, all agents hold enough liquid assets to consume the first best. But for large demand shocks, the impatient agents are constrained by their liquid asset holdings, so they consume less than the first best. In addition, due to a general equilibrium price effect, the patient agents actually consume *more* than the first best. These inefficiencies in consumption open the door for monetary policy to affect consumption and improve welfare.

We show how the central bank can use liquidity injections, in conjunction with forward guidance, to respond to aggregate shocks. Forward guidance takes the form of committing to undo the injections at a later date. As a result, the liquidity injections are *temporary* (i.e., repurchase agreements) much like open market operations in the federal funds market. If the injections were permanent, then the only effect would be a jump in nominal prices and there would be no real effects on the economy.

We show that this policy involves no liquidity injection in the lowest aggregate demand state and then monotonically increasing the size of the injections with the size of the shock. In this sense, monetary policy is *pro-cyclical*. These injections effectively redistribute consumption from low aggregate demand states to high aggregate demand states thereby improving welfare. In addition, by exploiting the general equilibrium price effect, the central bank redistributes consumption across agents—patient agents consume less in high demand states and impatient agents consume more—in a way that improves welfare.

The stabilization policy works differently in our environment compared to a standard New Keynesian (NK) model because the frictions are not the same. In a NK model, the relevant frictions are nominal rigidities. When there is a demand shock (e.g. a positive marginal utility shock to consumption), then demand for goods increases. With sticky prices, absent monetary policy action, agents end up consuming more than the efficient quantity since prices do not rise. Thus, the role of monetary policy is to restrain demand and pull consumption back closer to the efficient quantity. It does this by increasing the interest rate. With a negative demand shock the reverse is true—agents decrease consumption too much so the monetary authority lowers the interest rate to stimulate demand. Thus, in a NK model, monetary policy tries to restrain consumption in high demand states and encourage consumption in low demand states. In this sense, monetary policy is *counter-cyclical*.

In our model, the relevant friction is a liquidity or borrowing constraint. When the constraint binds, consumption is below the efficient quantity. Now if a positive demand shock occurs, the efficient quantity increases but unless the liquidity constraint is relaxed actual consumption does not change. Hence, actual consumption is even further away from the efficient quantity. Thus, optimal stabilization policy attempts to relax the liquidity constraint in order to increase consumption and bring it closer to the efficient quantity. So, when liquidity or borrowing constraints are the relevant friction, monetary policy tries to increase consumption in high demand states and discourage it in low demand states. In this case, monetary policy is *pro-cyclical*. This discussion illustrates one of the contributions of our paper—optimal stabilization policy works differently depending on the frictions that affect the economy.

2 Brief Literature Review

In recent years, economists have become concerned with the effects of monetary policy in heterogeneous agent models. Several papers examine this topic in environments with incomplete markets and sticky prices or wages. In this area, researchers have begun to separate the direct effects of monetary policy, working through the nominal interest rate channel and intertemporal substitution, from indirect general equilibrium effects.[2] However, with a few exceptions, most of this research is conducted assuming the central bank is away from the ZLB since it focuses on unexpected changes in interest rates or a Taylor rule for describing the evolution of the nominal policy rate.[3]

In the New Monetarist literature, heterogeneity across agents has been modeled in a variety of

[2] A non-exhaustive list includes Auclert (2017), Guerrieri and Lorenzoni (2017), Bayer et al. (2018), McKay and Reis (2016), Gornemann, Kuester and Nakajima (2016), Kaplan, Moll and Violante (2018) and Werning (2015).

[3] Notable exceptions are Guerrieri and Lorenzoni (2017) and McKay, Nakamura and Steinsson (2016).

ways, but our work is most closely related to Boel and Camera (2006). In their model, agents have different discount factors and these differences are permanent, which makes the definition of welfare difficult. By introducing stochastic discount factor shocks at particular points in time, instead, our model allows us to have a well-defined welfare function that mimics that for a representative agent.

Little work has been done on stabilization policy, instead, particularly at the ZLB. In this respect, our paper is closely related to Berentsen and Waller (2011) who study optimal stabilization policy away from the ZLB. In that model, the ZLB corresponds to the Friedman rule which yields the first-best allocation. So, at the ZLB there is no role for stabilization policy. In our paper, instead, there is still a role for stabilization policy at the ZLB. Furthermore, the Friedman rule may not be optimal due to a pecuniary externality.

In the New Monetarist literature, Rocheteau, Wright and Xiao (2018) study how open market operations affect the real economy. However, they analyze changes in the steady state stock of money and b:onds, but do not consider the use of open market operations to stabilize shocks to the economy. Gu, Han and Wright (2016), instead, study a case where the central bank surprises the economy with an increase in the monetary base but then promises to unwind it at a future date. They show that output and welfare increase with such a policy. This exercise has the flavor of our results but differs from them in two important ways. First, they study an 'MIT shock' where the announced policy is a zero probability event ex ante, i.e., it is not a response to a shock to the economy. In our paper, agents understand the stochastic process behind the demand shocks and the related central bank's policy response. Second, Gu, Han and Wright (2016) assume that the interest rate is away from the ZLB—otherwise the policy action would have no effect on the economy. We do assume the ZLB binds, yet policy can still increase output and welfare thanks to appropriate temporary injections of liquidity.

There is of course a vast literature on optimal monetary policy at the zero lower bound.[4] For example, a large number of papers in both the New Keynesian and New Monetarist literature have studied quantitative easing (QE) at the ZLB. We do not intend to survey all of them, but we do want to highlight that a common element of all of these studies is to study the effects of QE in response to severe and long lasting economic shocks which are intended to capture an economic crisis such as the 2007-2008 financial one. Most of the policy actions studied in these papers are

[4] Adam and Billi (2006), Adam and Billi (2008) and Jung et al. (2005) are among the first to have studied this topic. Many proposals have been put forward for monetary policy in a liquidity trap: raising the central bank's inflation target (e.g. Billi (2011)); quantitative easing (e.g. Woodford (2012), Gertler and Karadi (2013), Williamson (2016)); forward guidance (e.g. Eggertsson and Woodford (2003), Woodford (2012), Campbell et al. (2016)); negative interest rates (e.g. Eggertsson et al. (2017)); money injections (e.g. Gali (2017)).

permanent or long-term policy responses to the crisis shock. Unlike our paper, very few investigate the role of monetary policy for stabilizing subsequent and smaller economic shocks while still at the ZLB.

In addition to QE, other solutions have been studied for stimulating the economy at the ZLB such as raising the inflation target of the central bank, negative interest rates and forward guidance. Nearly all of these proposals have been studied in representative-agent New Keynesian models and have focused on changing demand via the intertemporal subsitution channel. Raising the inflation target and/or imposing negative nominal rates are intended to lower the real interest rate to discourage saving and increase spending. Forward guidance works through a wealth effect—by keeping real interest rates "low for long", future output and income will be very high, so agents want to bring some of that income forward to today. Hence, consumption today increases.

None of these policies are relevant in our model because of the frictions that are present. Since money is not super-neutral in our framework, raising steady-state inflation distorts consumption in a welfare-reducing way. Negative interest rates are not possible because we study the use of cash, not bank reserves. Forward guidance is required in our model but in a different way. It takes the form of anchoring inflation expectations so that temporary monetary injections do not have inflationary effects on the economy. It does not work through a wealth channel as in Eggertsson and Woodford (2003).

One paper that we do want to discuss in more detail is Auerbach and Obstfeld (2005). They use a cash-in-advance model with and without sticky prices to investigate if the central bank can enact policies to increase output in a liquidity trap. They argue that if the economy is in a liquidity trap at time t but there is some probability that the ZLB will not hold in the future, then a *permanent* injection of money at time t will increase real wages and consumption. In our paper, instead, injections of money must be *temporary* to have positive real effects—permanent injections simply lead to an immediate increase in the price level with no real effects. So, it is clear that the frictions and trading environment considered matter for the conduct of monetary policy.

3 The model

The model builds on Lagos and Wright (2005), Boel and Camera (2006) and Berentsen and Waller (2011). Time is discrete, the horizon is infinite and there is a large population of infinitely-lived agents who consume perishable goods and discount only across periods. In each period agents may visit two sequential rounds of trade—we refer to the first as DM and the second as CM.

Rounds of trade differ in terms of economic activities and preferences. In the DM, agents face an idiosyncratic trading risk such that they either consume, produce, or are idle. An agent consumes with probability α_b, produces with probability α_s and is idle with probability $1-\alpha_b-\alpha_s$. We refer to consumers as buyers and producers as sellers. Buyers get utility $\varepsilon u(q)$ from $q > 0$ consumption, where ε is a preference parameter, $u'(q) > 0$, $u''(q) \leq 0$, $u'(0) = +\infty$ and $u'(\infty) = 0$. Furthermore, we impose that the elasticity of utility $e(q) = qu'(q)/u(q)$ is bounded. Producers incur a utility cost $c(y)$ from supplying $y \geq 0$ labor to produce y goods, with $c'(y) > 0$, $c''(y) \geq 0$ and $c'(0) = 0$. Everyone can consume and produce in the CM. As in Lagos and Wright (2005), agents have quasilinear preferences $U(x) - n$, where the first term is utility from x consumption, and the second is disutility from n labor to produce n goods. We assume $U'(x) > 0$, $U''(x) \leq 0$, $U'(0) = +\infty$ and $U'(\infty) = 0$. Also, let q^* be the solution to $\varepsilon u'(q) = c'[(\alpha_b/\alpha_s)q]$ and x^* be the solution to $U'(x) = 1$.

3.1 Shocks

The economy is subject to both aggregate and idiosyncratic demand shocks, but agents are heterogeneous only with respect to the latter. Specifically, at the beginning of each CM, agents draw an idiosyncratic time-preference shock $\beta_z \in \{\beta_L, \beta_H\}$ determining their interperiod discount factor. This implies at the beginning of each period an agent can be either patient (type H) with probability ρ or impatient (type L) with probability $1-\rho$. We consider the case $0 < \beta_L < \beta < \beta_H < 1$ with no serial correlation in the draws and β being the average discount factor. Note that time-preference shocks introduce ex-post heterogeneity across households, but the long-run distribution of time preferences is invariant.

We also assume ε is stochastic like in Berentsen and Waller (2011), which allows us to study the optimal response of a central bank to aggregate demand shocks. The random variable ε has support $\Omega = [\underline{\varepsilon}, \bar{\varepsilon}]$, with $0 < \underline{\varepsilon} < \bar{\varepsilon} < \infty$. Shocks are serially uncorrelated and $f(\varepsilon)$ denotes the density function of ε. As shown below, output in the CM is constant so any volatility in total output per period is driven by ε shocks in the DM.

3.2 Information frictions, money and credit

The preference structure we selected generates a single-coincidence problem in the DM since buyers do not have a good desired by sellers. Moreover, two additional frictions characterize the DM. First, agents are anonymous as in Kocherlakota (1998), since trading histories of agents in the goods markets are private information. This in turn rules out trade credit between individual buyers and sellers. Second, there is no public communication of individual trading outcomes, which in turn eliminates the use of social punishments to support gift-giving equilibria. The combination of these two frictions together with the single coincidence problem implies that sellers require immediate compensation from buyers. So, buyers must use money to acquire goods in the DM.

Money is not essential for trade in the CM instead, and indeed agents can finance their consumption by getting credit, working or using money balances acquired earlier. To model credit, we assume agents are allowed to borrow and lend through selling and buying nominal bonds, subject to an exogenous credit constraint A. Specifically, agents lend $-p_{at}a_{t+1}$ (or borrow $p_{at}a_{t+1}$), where p_{at} is the price of a bond that delivers one unit of money in $t+1$, and receive back a_t. We assume that any funds borrowed or lent in the CM are repaid in the following CM. One can show that, even with quasi-linearity of preferences in the CM, there are gains from multi-period contracts due to time-preference shocks. Of course, default is a serious issue in all models with credit. However, to focus on optimal stabilization, we simplify the analysis by assuming a mechanism exists that ensures repayment of loans in the CM.

3.3 Policy tools

We assume a government exists that is in charge of monetary policy and is the only supplier of fiat money, of which an initial stock $M_0 > 0$ exists. Monetary policy has both a long-run and a short-run component. The long-run policy focuses on the trend inflation rate, whereas the short-run one is concerned with the output stabilization response to aggregate shocks. We denote the gross growth rate of money supply by $\pi = M_t/M_{t-1}$, where M_t denotes the money stock in the CM in period t. The central bank implements its long-term inflation goal by providing deterministic lump-sum injections of money $\tau = (\pi - 1)M_{t-1}$, which are given to private agents at the beginning of the CM. If $\pi > 1$, agents receive lump-sum transfers of money, whereas for $\pi < 1$ the central bank must be able to extract money via lump-sum taxes from the economy.

The central bank implements its short-term stabilization policy through state-contingent changes in the stock of money. We let $\tau_1(\varepsilon) = T_1(\varepsilon)M_{t-1}$ and $\tau_2(\varepsilon) = T_2(\varepsilon)M_{t-1}$ denote state-contingent cash injections received by private agents in the DM and CM respectively. We assume injections in the DM are undone in the CM, so that $\tau_1(\varepsilon) + \tau_2(\varepsilon) = 0$. Changes in $\tau_1(\varepsilon)$ thus affect the money stock in the DM without affecting the long-term inflation rate in the CM. This means that the long-term inflation rate is still deterministic since $\tau = (\pi - 1)M_{t-1}$ is not state dependent. Note that the state-contingent injections of cash can be viewed as a type of repurchase agreement—the central bank sells money in the DM under the agreement that it is being repurchased in the CM.

4 Efficient allocation

We start by discussing the allocation selected by a benevolent planner subject to the same physical and informational constraints faced by the agents. We will refer to this allocation as constrained-efficient. As is the case with any model with heterogeneous agents, defining the planner problem is not a trivial issue. The environment's frictions imply the planner can observe neither types nor identities in the DM and therefore has no ability to transfer resources across agents over time in that market. Furthermore, at the start of the DM, all agents are identical ex ante since the previous period β shock is no longer relevant and the DM shocks have not been realized. Thus, if we look at welfare from this point in time, we effectively have a representative-agent problem.

Therefore, the planning problem in the DM corresponds to a sequence of static maximization problems subject to the technological constraints. This implies in the DM the planner must solve:

$$\begin{aligned} &\underset{q,y}{Max} \int_{\Omega} \{\alpha_b \varepsilon u[q(\varepsilon)] - \alpha_s c[y(\varepsilon)]\} f(\varepsilon) d\varepsilon \\ &\text{s.t. } \alpha_b q(\varepsilon) = \alpha_s y(\varepsilon) \end{aligned} \tag{1}$$

In the CM, once the β shocks are realized, the agents are heterogeneous with regards to intertemporal choices. We also do not have the informational frictions in this market that exist in the DM. Consequently, the planner can transfer resources across agents over time and therefore chooses consumption and labor sequences $\{x_{z0}, x_{z1}, ..\}$ and $\{n_{z0}, n_{z1}, ..\}$ for $z = H, L$ that maximize a weighted sum of individual utility functions subject to feasibility and non-negativity constraints:

$$\begin{aligned} &Max \sum_{z=H,L} \sigma_z \left[U(x_{z0}) - n_{z0} + \sum_{t=1}^{\infty} \beta_z \beta^{t-1} (U(x_{zt}) - n_{zt}) \right] \\ &\text{s.t. } \rho x_{Ht} + (1-\rho) x_{Lt} = \rho n_{Ht} + (1-\rho) n_{Lt} \quad \text{for } t = 0, 1, 2, ... \\ &\text{s.t. } n_{zt} \geq 0 \quad \text{for } z = H, L \text{ and } t = 0, 1, 2, ... \end{aligned} \tag{2}$$

Here σ_H and σ_L are positive utility weights. A solution to this problem is characterized by:

$$U'(x_{z0}) = 1 - \mu_t^z/\sigma_z \qquad \text{for } z = H, L \text{ and } t = 0 \tag{3}$$

$$U'(x_{zt}) = 1 - \mu_t^z/\sigma_z\beta_z\beta^{t-1} \qquad \text{for } z = H, L \text{ and } t \geq 1 \tag{4}$$

where μ_t^z denotes the Kuhn-Tucker multiplier associated with the non-negativity constraint on n_{zt}. Note that the difference between equation (3) and (4) implies a different allocation when $t = 0$ from when $t \geq 1$. In short, once $t = 1$ is reached, the planner would prefer to reoptimize and give each agent the allocation solving (3) rather than (4) evaluated at $t = 1$. This implies the social planner problem is not time consistent. Consequently, satisfying (3) and (4) requires that the planner be able to commit to fulfill future promises of consumption and labor in CM exchange. If the planner cannot commit to this, then the only consistent solution to this problem is $\mu_t^z = 0$ in all periods. This implies that a discretionary planner allocation has $U'(x_{zt}) = 1$ and $n_{zt} > 0$ for $z = H, L$ and $t \geq 0$—the discretionary planner wants both types to work and consume a constant and equal amount in every period.

Since our focus will be on monetary policy stabilization, we adopt the allocation corresponding to the discretionary planner as our benchmark for welfare. We do so for several reasons. First, monetary policy will only be able to offset shocks that hit the DM since the absence of frictions in the CM eliminates any role for stabilization policy there. Second, at the beginning of the DM all agents are ex ante the same. So, viewing welfare from this point in time is equivalent to having a representative-agent problem. Finally, there are no ex-post welfare gains from transferring labor across agents based on the β shocks because of quasi-linear utility—shifting labor from one agent to fulfill earlier promises is zero sum ex post.

In sum, in the constrained-efficient allocation we focus on an allocation such that marginal consumption utility equals marginal production disutility in each market and in every period. Such allocation is stationary and defined by $\varepsilon u'[q(\varepsilon)] = c'[(\alpha_b/\alpha_s)q(\varepsilon)]$ for all ε in the DM and $U'(x) = 1$ in the CM. The constrained-efficient consumption is therefore defined by $q_H = q_L = q^*$ and $x_H = x_L = x^*$, thus implying equal consumption for H and L agents in both DM and CM.

5 Stationary monetary allocations

In what follows, we want to determine if the constrained-efficient allocation can be decentralized in a monetary economy with competitive markets.Thus, we focus on stationary monetary outcomes such that end-of-period real money and bonds balances are time invariant.

We simplify notation omitting t subscripts and use a prime superscript and a -1 subscript to denote variables corresponding to the next and previous period respectively. We let p_1 and p_2 denote the nominal price of goods in the DM and the CM respectively of an arbitrary period t. We also let β_j and β_z denote the discount factors drawn in period $t-1$ and t respectively. In addition, we normalize all nominal variables by p_2, so that DM trades occur at the real price $p = p_1/p_2$. In this manner, the timing of events in any period t can be described as follows.

An arbitrary agent of type $j = H, L$ enters the DM in period t with a portfolio $\omega_j = (m_j, a_j)$ listing $m_j = m(\beta_j)$ real money holdings and $a_j = a(\beta_j)$ loans (or savings) from the preceding period after experiencing a time-preference shock β_j. Trading shocks k and aggregate demand shocks ε are then realized and agents receive a lump-sum transfer $\tau_1(\varepsilon) = T_1(\varepsilon)M_{-1}$. After the DM closes, the agent enters the CM with portfolio $\omega_j^k = (m_j^k, a_j)$, where $m_j^k = m_j^k(\beta_j, \varepsilon)$ denotes money holdings carried over from the DM and $k = s, b, o$ denotes the trading shock experienced in the DM. Here, o identifies an idle agent, while b and s identify a buyer and a seller respectively. Thus, if we let $q_j = q(\beta_j, \varepsilon)$ denote consumption and $y_j = y(\beta_j, \varepsilon)$ production in the DM, individual real money holdings for an agent j evolve as follows:

$$\begin{aligned} m_j^b &= m_j + \tau_1 - pq_j \\ m_j^s &= m_j + \tau_1 + py_j \\ m_j^o &= m_j + \tau_1 \end{aligned} \tag{5}$$

That is, buyers deplete balances by pq_j while sellers increase them by py_j. Idiosyncratic time-preference shocks β_z are then realized at the beginning of the CM. Left-over cash is used to trade and settle bonds positions and x and n are respectively consumption bought and production sold in the CM. Note that bonds positions a_j at the beginning of the CM are not affected by trading shocks in the DM, since bonds can only be used in the CM. Agents also receive lump-sum transfers $\tau + \tau_2(\varepsilon)$, adjust their money balances $m_z' = m'(\beta_z, \varepsilon)$ and decide whether they want to borrow or lend $a_z' = a'(\beta_z, \varepsilon)$, where m_z' and a_z' denote real values of money holdings and loans (or savings if $a_z' < 0$) at the start of tomorrow's DM. Figure 1 displays the timeline of shocks and decisions within each period.

Since we focus on stationary equilibria where end-of-period real money balances are time and state invariant so that $M/p_2 = M'/p'_2$, we have that:

$$\frac{p'_2}{p_2} = \frac{M'}{M} = \pi \tag{6}$$

which implies the inflation rate equals the growth rate of money supply. The government budget constraint therefore is:

$$\tau = (\pi - 1)[\rho m_H + (1-\rho) m_L] \tag{7}$$

Note that the long-run inflation rate is deterministic since the per capita lump-sum transfers τ in the CM are not state dependent.

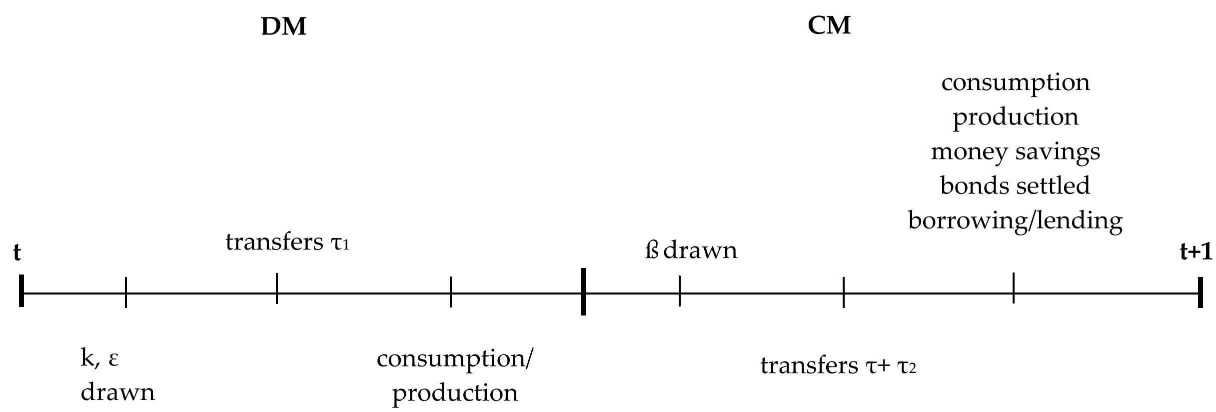

Figure 1: Timing of events within a period

5.1 The CM problem

Given the recursive nature of the problem, we use dynamic programming to analyze the problem of an agent j at any date, with $j = H, L$. We let $V(\omega_j)$ denote the expected lifetime utility for an agent entering the DM with portfolio ω_j before shocks are realized. We also let $W_z(\omega_j^k)$ denote the expected lifetime utility from entering the CM with portfolio ω_j^k and receiving a discount factor shock β_z at the beginning of the CM. The agent's problem at the start of the CM then is:

$$W_z(\omega_j^k) = \underset{x_{jz}^k, n_{jz}^k, a'_z, m'_z}{Max} \; U(x_{jz}^k) - n_{jz}^k + \beta_z V(\omega'_z) \tag{8}$$

$$\text{s.t. } x_{jz}^k + \pi m'_z = n_{jz}^k + m_j^k + p_a \pi a'_z - a_j + \tau + \tau_2$$

$$\text{s.t. } a'_z \leq A$$

$$\text{s.t. } m'_z \geq 0$$

where $A \geq 0$ is a constant denoting an exogenous borrowing constraint. The resources available to the agent in the CM depend on the realization of the DM trading shock k, as well as the aggregate and idiosyncratic shocks ε, β_j and β_z. Specifically, an agent has m_j^k real balances carried over from the DM and is able to borrow $\pi a'_z$ (or lend if $a'_z < 0$) at a price p_a. Other resources are n_{jz}^k receipts from current sales of goods and lump-sum transfers $\tau + \tau_2$. These resources can be used to finance current consumption x_{jz}^k, to pay back loans a_j and to carry $\pi m'_z$ real money balances into next period. The factor $\pi = p'_2/p_2$ multiplies a'_z and m'_z because the budget constraint is expressed in real terms. Rewriting the constraint in terms of n_{jz}^k and substituting into (8) yields:

$$\begin{aligned} W_z(\omega_j^k) = & \underset{x_{jz}^k, a'_z, m'_z}{Max} \; U(x_{jz}^k) - x_{jz}^k - \pi m'_z + \pi p_a a'_z - a_j + m_j^k + \tau + \tau_2 + \beta_z V(\omega'_z) \\ & \text{s.t. } \; a'_z \leq A \\ & \text{s.t. } \; m'_z \geq 0 \end{aligned}$$

Note that here we are focusing on a stationary equilibrium in which all agents provide a positive labor effort. Conditions for $n_{jz}^k > 0$ are in the Appendix, but the intuition is that agents will always choose to work in the CM if the borrowing limit A is tight enough. It follows that in a stationary monetary economy we must have:

$$1 = \frac{\partial W_z(\omega_j^k)}{\partial m_j^k} = -\frac{\partial W_z(\omega_j^k)}{\partial a_j} \tag{9}$$

This result depends on the quasi linearity of the CM preferences and the use of competitive pricing. It implies that the marginal valuation of real balances and bonds in the CM are identical and do not depend on the agent's current type z or past type j, wealth ω_j^k or trade shock k. This allows us to disentangle the agents' portfolio choices from their trading histories since:

$$W_z(\omega_j^k) = W_z(0) + m_j^k - a_j$$

i.e., the agent's expected value from having a portfolio ω_j^k at the start of a CM is the expected value from having no wealth, $W_z(0)$, letting $\omega_j = (0, 0) \equiv 0$, plus the current real value of net wealth $m_j^k - a_j$. Note also that everyone consumes identically in the CM since:

$$U'(x) = 1 \tag{10}$$

which also implies $x = x^*$. That is, everyone consumes the same amount x^* independent of current type and past shocks, the reason being that agents in the CM can produce any amount at constant marginal cost. Thus, goods market clearing in the CM requires:

$$x^* = \alpha_b N^b + \alpha_s N^s + (1 - \alpha_b - \alpha_s) N^o \tag{11}$$

where $N^k = \rho^2 n_{HH}^k + \rho(1-\rho)(n_{LH}^k + n_{HL}^k) + (1-\rho)^2 n_{LL}^k$ is labor effort for all agents who experienced a trading shock in the DM, with $k = b, s, o$. Let $\mu_z^m \geq 0$ denote the Kuhn-Tucker multipliers associated with the non-negativity constraint for money and λ_z^a denote the multiplier on the borrowing constraint. The first order conditions for the optimal portfolio choice then are:

$$1 = \frac{\beta_z}{\pi} \frac{\partial V(\omega_z')}{\partial m_z'} + \mu_z^m / \pi \tag{12}$$

$$-p_a = \frac{\beta_z}{\pi} \frac{\partial V(\omega_z')}{\partial a_z'} - \lambda_z^a / \pi \tag{13}$$

The left hand sides of the expressions above define the marginal cost of the assets. The right hand sides define the expected marginal benefit from holding the asset, either money or bonds, discounted according to time preferences and inflation. From (12) and (13) we know that money holdings m_z' and bonds a_z' are independent of trading histories and past demand shocks, but instead depend on the current type z and the expected marginal benefit of holding money and bonds in the DM, which may differ across types. We will study this next.

5.2 The DM problem

An agent with a portfolio ω_j at the opening of the DM before aggregate demand and trading shocks are realized has expected lifetime utility:

$$V(\omega_j) = \int_{\Omega} \left\{ \alpha_b V^b(\omega_j) + \alpha_s V^s(\omega_j) + (1 - \alpha_b - \alpha_s) V^o(\omega_j) \right\} f(\varepsilon) d\varepsilon \tag{14}$$

First, we determine y_j. The seller's problem depends on the current disutility of production and the expected continuation value. Specifically, the seller's problem can be written as:

$$V^s(\omega_j) = \underset{y_j}{Max} \quad -c(y_j) + \rho W_H(\omega_j^s) + (1 - \rho) W_L(\omega_j^s), \tag{15}$$

for which the first order conditions, together with (5) and (9), give:

$$c'(y_j) = p \tag{16}$$

Note that (16) implies production is not type dependent, i.e. $y_j = y$ for $j = H, L$.

Now, we determine q_j. A buyer's problem is:

$$V^b(\omega_j) = \underset{q_j}{Max} \quad \varepsilon u(q_j) + \rho W_H(\omega_j^b) + (1-\rho)W_L(\omega_j^b) \tag{17}$$

$$\text{s.t. } pq_j \leq m_j + \tau_1$$

The budget constraint reflects that consumption can be financed with both money holdings and DM transfers. Let λ_j^b denote the multiplier on the buyer's budget constraint. Using (5) and (9), the first order conditions for the buyer's problem imply:

$$\varepsilon u'(q_j) = p(1 + \lambda_j^b) \tag{18}$$

From (16) and (18) we know that if the buyer is constrained and $\lambda_j^b > 0$, then $\varepsilon u'[q_j(\varepsilon)] > c'[y(\varepsilon)]$. If instead the buyer is unconstrained and therefore $\lambda_j^b = 0$, then $\varepsilon u'(q_j(\varepsilon)) = c'(y(\varepsilon))$.

Last, an idle agent's problem is simply:

$$V^o(\omega_j) = \rho W_H(\omega_j^n) + (1-\rho)W_L(\omega_j^n)$$

Goods market clearing in the DM therefore requires:

$$\alpha_s y(\varepsilon) = \alpha_b[\rho q_H(\varepsilon) + (1-\rho)q_L(\varepsilon)] \qquad \text{for } \varepsilon \in \Omega \tag{19}$$

5.3 Monetary equilibrium

To find optimal savings for an agent j use (8), (14), (15), (16) and (17) to obtain:

$$V(\omega_j) = \int_\Omega \left\{ m_j - a_j + \tau_1 + \alpha_b[\varepsilon u(q_j) - pq_j] - \alpha_s[c(y) - py] + EW(0) \right\} f(\varepsilon) d\varepsilon$$

The expected lifetime utility $V(\omega_j)$ therefore depends on the agent's net wealth and income $m_j - a_j + \tau_1$ and two other elements: the expected continuation payoff $EW(0) = \rho W_L(0) + (1-\rho)W_L(0)$ and the expected surplus from trade in the DM. With probability α_b the agent spends pq_j on consumption deriving utility $\varepsilon u(q_j)$ and with probability α_s, instead, gets disutility $c(y)$ from production and earns py from his sales. Note that, unlike in the representative-agent case, the expected earnings $p(y - q_j)$ from DM trades might be different from zero since amounts produced and consumed by an agent $j = H, L$ may be mismatched. Hence, we have:

$$\frac{\partial V(\omega_j)}{\partial m_j} = \int_{\Omega} \left\{ 1 + \alpha_b \left[\frac{\varepsilon u'(q_j)}{p} - 1 \right] \right\} f(\varepsilon) d\varepsilon \tag{20}$$

and

$$\frac{\partial V(\omega_j)}{\partial a_j} = -1, \tag{21}$$

which imply money is valued dissimilarly by agents, whereas bonds are valued identically in the economy. Combining (12) with (20) and (13) with (21) one gets that in a monetary equilibrium the following Euler equations must hold:

$$\frac{\pi - \beta_z}{\beta_z} = \int_{\Omega} \left\{ \alpha_b \left[\frac{\varepsilon u'(q_z(\varepsilon))}{c'(y(\varepsilon))} - 1 \right] \right\} f(\varepsilon) d\varepsilon \tag{22}$$

and

$$\pi p_a = \beta_z + \lambda_z^a \tag{23}$$

The expression in (22) tells us that the choice of real balances depends on three components. The first two are standard: the discount factor β_z and the real yield on cash $1/\pi$. The third component is $\varepsilon u'(q_z)/c'(y)$. This can be interpreted as the expected liquidity premium from having cash available in the DM and it arises because money is needed to trade in that market. This premium grows with the severity of the cash constraint and the likelihood of a consumption shock α_b. The expression in (23), instead, refers to the choice of bonds, which depends on the discount factor β_z and the real yield $1/\pi p_a$. Note that (23) implies that bonds have no liquidity premium. This is because bonds are always held until maturity and cannot be used to buy consumption in the DM.

We can now define the equilibrium as the set of values of m_z and a_z for $z = H, L$ that solve (22) and (23). The reason is that once the equilibrium stocks of money and bonds are determined, all other endogenous variables can be derived.

Definition 1 *A symmetric stationary monetary equilibrium consists of m_z satisfying (22) and a_z satisfying (23) for $z = H, L$.*

We now want to investigate whether a CM to CM bond a_z for $z = H, L$ would indeed circulate in this economy. We find that the following result holds:

Lemma 1 *A stationary monetary equilibrium exists with impatient agents borrowing and patient agents lending at a price* $p_a = \beta_H/\pi$*. Specifically,* $a_L = A$ *and* $a_H = -(1-\rho)A/\rho$*.*

Why are agents interested in trading such a bond in equilibrium? This is somewhat puzzling since we know from (10) they always consume the efficient quantity x^* in the CM. This in turn implies that there is no reason for using bonds for consumption smoothing here due to the quasi linearity of preferences. Bonds, however, allow agents to smooth the labor effort across periods—H agents prefer to work more today and less in the future, whereas L agents would rather do the opposite.

Once we know the price at which these bonds circulate in equilibrium, we can pin down their net nominal yield, which is:

$$i = \frac{1}{p_a} - 1 \quad \Rightarrow \quad i = \frac{\pi}{\beta_H} - 1 \tag{24}$$

Note that i is the nominal interest rate on an illiquid asset and it is affected directly by long-term monetary policy through π. We will refer to i as the nominal interest rate in this economy. We now want to determine the returns on money and bonds that are consistent with equilibrium.

Lemma 2 *Any stationary monetary equilibrium must be such that* $\pi \geq \beta_H$*, i.e.* $i \geq 0$*.*

This result derives from a simple no-arbitrage condition—in a monetary equilibrium, the value of money cannot grow too fast with $\pi < \beta_H$ or else type H agents will not spend it.[5] This, together with (24), implies that to run the Friedman rule the monetary authority must let $\pi \to \beta_H$ and cannot target β_L instead. In what follows, we investigate how the result in Lemma 2 affects the central bank's choice for the optimal inflation rate.

6 Optimal inflation rate

At this point, we know that given the result in Lemma 2 the monetary authority is constrained in its ability to give a rate of return on money that is attractive for everyone. Given this result, one should expect inefficiencies will arise at $i = 0$[6] and therefore the Friedman rule might not be the optimal policy here. We investigate this next, and in this section we will focus on the optimal inflation rate in an economy without aggregate demand shocks, which we will then reintroduce in Section 6. We find that the following result holds:

[5] Other models with heterogeneous time preferences have analogous results, in that the rate of return on the asset cannot exceed the lowest rate of time preference. See for example Becker (1980) and Boel and Camera (2006). In those models, however, types are fixed.

[6] We know from (24) that it is equivalent to fix i or π.

Proposition 1 *Let $i = 0$. If $c''(y) = 0$, then $q_H = q^*$ and $q_L < q^*$. If $c''(y) > 0$, then $q_H > q^*$ and $q_L < q^*$.*

Proposition 1 implies that, since agents value future consumption differently, $i = 0$ fails to sustain the constrained-efficient allocation in a monetary equilibrium. Indeed, even if the Friedman rule eliminates the opportunity cost of holding money for type H agents, it still fails to provide incentives for everyone to save enough since $\pi > \beta_L$. That is because impatient agents are facing an effective nominal interest rate equal to $\beta_H/\beta_L - 1$, which is positive. We know from Lemma 2 that $\pi \geq \beta_H$ and therefore the central bank is limited in its ability to reduce interest rates even further. Thus, type L agents remain constrained even when $i = 0$.

Proposition 1 also implies that the nature of preferences has important consequences for the optimality of $i = 0$. Specifically, a convex disutility from labor generates a pecuniary externality induced by type L agents. This happens because impatient agents consume too little even at $i = 0$, thus driving down total output, marginal cost of production and relative price. The low price in turn induces type H agents to consume too much compared to the efficient allocation. This pecuniary externality disappears with linear costs, since in that case the marginal cost of production (and hence the relative price in the DM) is constant at any level of output. This externality also depends on the DM market structure. With competitive pricing, all sellers produce the same quantity. With bargaining, instead, individual sellers can produce different amounts according to the type of buyer they meet, as shown in the Appendix, and thus the externality disappears.[7]

In light of the results described in Proposition 1, one must wonder if setting $i = 0$ is still the optimal monetary policy in this economy. The following result clarifies when this is the case.

Proposition 2 *If $c''(y) = 0$, $i = 0$ is always the optimal policy. If $c''(y) > 0$, $i > 0$ is the optimal policy if $c'(y) < 1$ and $\beta_L u'(q_L) > \beta_H u'(q_H)$. Otherwise, $i = 0$ is optimal.*

The Friedman rule is always the optimal policy with linear costs. In that case, even if $i = 0$ fails to sustain the constrained-efficient allocation, such policy delivers a second best allocation that cannot be Pareto improved. With convex costs, however, the Friedman rule is not necessarily optimal. Why not? Because the policy maker needs to take into account the pecuniary externality induced by the underconsumption of impatient agents. In this case, $q_H > q^*$ and $q_L < q^*$ at

[7] As explained in Rocheteau and Wright (2005), considering a competitive DM does not make money inessential in this type of models as long as the double coincidence problem and anonimity assumptions are maintained. Indeed, there are numerous studies based on Lagos and Wright (2005) which use Walrasian markets (e.g. Rocheteau and Wright (2005) and Berentsen, Camera and Waller (2007) among the first ones).

$i = 0$. Increasing i will lower q_L further from q^*, which worsens welfare, but it moves q_H closer to q^*, which improves welfare. Thus, the optimality of the Friedman rule hinges on whether the welfare loss from moving q_L further from q^* outweighs the welfare gain from moving q_H closer to q^*. Proposition 2 implies that $i > 0$ is optimal if two conditions hold. First, $p = c'(y) < 1$, so that the relative price is so low that it leads to substantial overconsumption of type H agents. Second, $u'(q_L)/u'(q_H) < \beta_H/\beta_L$, meaning the disparity in consumption between impatient and patient agents, and thus the departure of DM of consumption from the first best, is sufficiently large.

Example 1: optimal inflation In order to derive intuition for the results in Proposition 2, we consider an example with the following functional forms:

$$u(q) = 1 - \exp^{-q} \quad \text{and} \quad c(y) = \exp^{y} - 1$$

In this case, the optimal inflation problem to be solved in a monetary equilibrium becomes:

$$\begin{aligned} \underset{\pi}{Max} \;\; & \alpha_b \left[(1-\rho)\varepsilon u(q_L) + \rho \varepsilon u(q_H)\right] - \alpha_s c(y) \\ \text{s.t.} \;\; & \frac{\pi - \beta_H}{\beta_H} = \alpha_b \left[\frac{\varepsilon \exp^{-q_H}}{\exp^{y}} - 1\right] \\ \text{s.t.} \;\; & \frac{\pi - \beta_L}{\beta_L} = \alpha_b \left[\frac{\varepsilon \exp^{-q_L}}{\exp^{y}} - 1\right] \\ \text{s.t.} \;\; & \alpha_s y = \alpha_b[\rho q_H + (1-\rho) q_L] \end{aligned} \tag{25}$$

If we differentiate the objective function in (25), we find that the optimal π must satisfy:

$$\alpha_b \left[(1-\rho)\varepsilon u'(q_L)\frac{dq_L}{d\pi} + \rho \varepsilon u'(q_H)\frac{dq_H}{d\pi}\right] - \alpha_s c'(y)\left[\rho \frac{dq_H}{d\pi} + (1-\rho)\frac{dq_L}{d\pi}\right] \leq 0 \tag{26}$$

In the Appendix, we derive expressions for $dq_L/d\pi$ and $dq_H/d\pi$ from the constraint in (26) and let $\varepsilon = \exp$ so that $\ln(\varepsilon) = 1$. We find that in order for $\pi = \beta_H$ to be optimal in this case it must be that:

$$\frac{\beta_H - \beta_L}{\beta_H} \leq \frac{\alpha_s}{\rho(1-\alpha_b)}$$

Intuitively, the condition above imposes an upper bound on time-preference heterogeneity. This will limit the pecuniary externality highlighted in Propositions 1 and 2, and $i = 0$ will be optimal with convex costs. Of course, if the condition above is not satisfied, then $i > 0$ must be optimal.

That would be the case, for example, with $\rho = 0.90$, $\beta_H = 0.99$, $\beta_L = 0.70$ and $\alpha_s = \alpha_b = 0.10$. In this case the central bank would choose $i = 1.46\%$.

7 Optimal stabilization policy at the ZLB

We now reintroduce aggregate demand shocks ε. We first investigate which inefficiencies arise when the central bank does not engage in stabilization policy, i.e. when $\tau_1(\varepsilon) = \tau_2(\varepsilon) = 0$. We call this passive policy and then compare it to active stabilization policy, i.e. the policy implemented by a central bank whose objective is to maximize the weighted welfare of the agents in the economy. We do so for $i = 0$. We find the following result holds with passive policy.

Proposition 3 *Let $i = 0$ and $\tau_1(\varepsilon) = \tau_2(\varepsilon) = 0$. A unique monetary equilibrium exists for $c''(y) \geq 0$ such that: with $c''(y) = 0$, then $q_L(\varepsilon) = q_H(\varepsilon) = q^*(\varepsilon)$ for $\varepsilon \leq \tilde{\varepsilon}$ and $q_L(\varepsilon) < q_H(\varepsilon) = q^*(\varepsilon)$ for $\varepsilon > \tilde{\varepsilon}$, where $\tilde{\varepsilon} \in [0, \bar{\varepsilon}]$; with $c''(y) > 0$, then $q_L(\varepsilon) = q_H(\varepsilon) = q^*(\varepsilon)$ for $\varepsilon \leq \hat{\varepsilon}$ and $q_L(\varepsilon) < q^*(\varepsilon) < q_H(\varepsilon)$ for $\varepsilon > \hat{\varepsilon}$, where $\hat{\varepsilon} \in [0, \bar{\varepsilon}]$.*

Proposition 3 implies that, with passive policy, impatient agents are unconstrained in low demand states and consume $q_L = q^*$. They are instead constrained in high demand states, when they consume $q_L < q^*$. Why don't type L ever consume $q_L > q^*$? Because the pecuniary externality outlined in Proposition 1 cannot arise when all agents are unconstrained, and therefore $q_L = q_H = q^*$ with $c''(y) \geq 0$ in the low demand states. Agents will use the extra cash to work less in the CM. Type H agents, instead, are never constrained, and overconsume only in high demand states if $c''(y) > 0$.

We now move on to studying the problem of a central bank engaged in stabilization policy and thus maximizing welfare by choosing the quantities consumed and produced by each type $j = H, L$ in each state subject to the constraint that the chosen quantities satisfy the conditions characterizing a competitive equilibrium. The policy is then implemented by choosing state-contingent injections

$\tau_1(\varepsilon)$ and $\tau_2(\varepsilon)$ accordingly. The primal Ramsey problem faced by the central bank is:[8]

$$\begin{aligned}
&\underset{q_L(\varepsilon),q_H(\varepsilon),y(\varepsilon)}{Max} \int_\Omega \{\alpha_b\varepsilon\left[\rho u(q_H(\varepsilon))+(1-\rho)u\left(q_L(\varepsilon)\right)\right]-\alpha_s c(y(\varepsilon))\} f(\varepsilon)d\varepsilon \\
&\text{s.t. } \frac{\pi-\beta_H}{\beta_H}=\int_\Omega\left\{\alpha_b\left[\frac{\varepsilon u'(q_H(\varepsilon))}{c'(y(\varepsilon))}-1\right]\right\} f(\varepsilon)d\varepsilon \\
&\text{s.t. } \frac{\pi-\beta_L}{\beta_L}=\int_\Omega\left\{\alpha_b\left[\frac{\varepsilon u'(q_L(\varepsilon))}{c'(y(\varepsilon))}-1\right]\right\} f(\varepsilon)d\varepsilon \\
&\text{s.t. } \alpha_s y(\varepsilon)=\alpha_b[\rho q_H(\varepsilon)+(1-\rho)q_L(\varepsilon)]
\end{aligned}$$

Note that we are focusing on a monetary equilibrium such that $m_j>0$ for $j=H,L$. This explains why the first two constraints in the Ramsey problem must hold with equality. Moreover, since $i=0$, then (22) implies that $\varepsilon u'(q_H(\varepsilon))=c'(y(\varepsilon))$ in every state, and the Ramsey planner simply solves the following problem:

$$\begin{aligned}
&\underset{q_L(\varepsilon),y(\varepsilon)}{Max} \int_\Omega \{\alpha_b\varepsilon\left[\rho u(q_H(\varepsilon))+(1-\rho)u\left(q_L(\varepsilon)\right)\right]-\alpha_s c(y(\varepsilon))\} f(\varepsilon)d\varepsilon \\
&\text{s.t. } \frac{\pi-\beta_L}{\beta_L}\geq\int_\Omega\left\{\alpha_b\left[\frac{\varepsilon u'(q_L(\varepsilon))}{c'(y(\varepsilon))}-1\right]\right\} f(\varepsilon)d\varepsilon \\
&\text{s.t. } \alpha_s y(\varepsilon)=\alpha_b[\rho q_H(\varepsilon)+(1-\rho)q_L(\varepsilon)]
\end{aligned} \tag{27}$$

We find the following result holds:

Proposition 4 *If $i=0$, in a monetary equilibrium with $m_L>0$, $m_H>0$ and $c''(y)\geq 0$ the optimal policy is $q_L(\varepsilon)<q^*(\varepsilon)$ for all states. This implies $q_H(\varepsilon)=q^*(\varepsilon)$ if $c''(y)=0$ and $q_H(\varepsilon)>q^*(\varepsilon)$ if $c''(y)>0$ instead.*

Proposition 4 implies that the central bank is able to temporarily relax liquidity constraints on impatient agents at $i=0$. It can do so simply engaging in repo arrangements that are undone at a later date, i.e. $\tau_1(\varepsilon)+\tau_2(\varepsilon)=0$. But why does the Ramsey planner choose $q_L<q^*$ in all states? Because we know from Proposition 3 that without policy intervention agents L would have enough cash to buy q^* in low demand states, but in high demand states their cash holdings would constrain their spending to $q_L<q^*$. This would create an inefficiency of consumption across

[8]The objective function of the Ramsey problem faced by the central bank is:

$$\rho(1-\beta_H)V(\omega_H)+(1-\rho)(1-\beta_L)V(\omega_L)$$

where $V(\omega_j)$ is defined in (14). We know that trades are efficient in the CM. Moreover, $\tau_1(\varepsilon)+\tau_2(\varepsilon)=0$ and $\bar{m}=\rho m_H+(1-\rho)m_L$. Therefore, the central bank has to worry only about maximizing $\int_\Omega \{\alpha_b\varepsilon\left[\rho u(q_H(\varepsilon))+(1-\rho)u\left(q_L(\varepsilon)\right)\right]-\alpha_s c(y(\varepsilon))\} f(\varepsilon)d\varepsilon$.

states that can be overcome by stabilization policy. Note also that the central bank is not actively trying to stabilize consumption of patient agents. However, the short-term monetary policy aimed at stabilizing q_L generates an externality on q_H. Since $q_L < q^*$ in all states with convex costs, type H agents always consume more than q^* in light of Proposition 1. Moreover, we know from Proposition 3 that without policy intervention agents H would buy q^* in some states and $q_H > q^*$ in others. Stabilization policy addresses this discontinuity indirectly and consumption is smoothed so that $q_H > q^*$ in all states. Example 2 illustrates.

It is worth emphasizing that our aim here is to determine how stabilization policy can improve welfare compared to the allocation that would be achieved with a passive policy in which agents would only be able to rely on their money balances to finance consumption in the DM. That is why Proposition 4 focuses on an equilibrium where money balances are positive for both L and H agents in all states. Of course, the central bank could also provide L agents with enough liquidity to finance q^* in every state. In that case, however, we would have $m_L = 0$ in all states since L agents would have no incentive to bring any money.

We also need to mention that we are not focusing on informational problems and that is why we study a case in which all agents receive the same state-contingent cash injections. We don't view this as problematic since patient agents are unconstrained at the ZLB and thus transfers to them are irrelevant, thus implying such agents have no incentive to mimic the impatient ones. Moreover, there is no additional inflation induced by our proposed stabilization policy since cash injections are undone at the end of the same period. However, our framework also allows for type-specific cash injection since impatient agents are the only borrowers and patient ones the only savers. This implies types are revealed through the market as long as the central bank can track their bond position.

Example 2: stabilization policy We now want to get some intuition on optimal stabilization policy at the zero lower bound. We focus on convex costs and consider the following functional forms:

$$u(q) = 1 - \exp^{-q} \quad \text{and} \quad c(y) = \exp^{y} - 1$$

Derivations are in the Appendix. Figure 2 illustrates the effects of active and passive policies for impatient (type L) agents with the specified cost function and assuming the following parameter values: $\alpha_s = \alpha_b = 0.3$, $\beta_H = 0.99$, $\beta_H = 0.95$, $\rho = 0.5$. The values for β_H and β_L are consistent with the evidence in Lawrence (1991), Carroll and Samwick (1997) and Samwick (1998) who provide empirical estimates of distributions of discount factors.

The curve "efficient $q_L(\varepsilon)$" represents the constrained-efficient allocation at which $q^*(\varepsilon) = q_H^*(\varepsilon) = q_L^*(\varepsilon)$. The curve "passive $q_L(\varepsilon)$" represents equilibrium consumption for type L under a passive policy, whereas the curve "active $q_L(\varepsilon)$" denotes consumption for the same agents when the central bank engages in stabilization policy. Consistent with the results in Proposition 1, $q_L < q^*$ in all states with stabilization policy. The curve "monetary injections" denotes the lump-sum injection $\tau_1(\varepsilon)$ implemented by the central bank to conduct stabilization policy. Note that the higher the demand for the DM good, the higher the injection needed to finance the increase in consumption. The important thing to notice here is that the central bank's optimal choice is strictly increasing in ε—the central bank chooses to reduce consumption from the first best in low demand states in order to increase it in higher demand states.

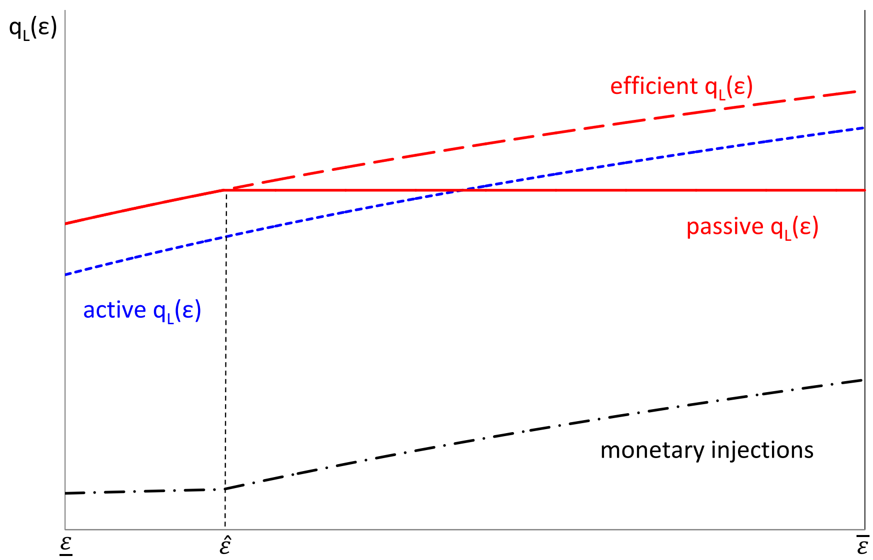

Figure 2: Stabilization versus passive policy for impatient (type L) agents

Figure 3 illustrates the effects of both active and passive policies for patient (type H) agents. The curve "active $q_H(\varepsilon)$" illustrates that, consistent with the results in Proposition 4, type H agents always consume more than q^* with stabilization policy. That's because $q_L < q^*$ in all states in that case, and thus $q_H > q^*$ in light of the externality in Proposition 1.

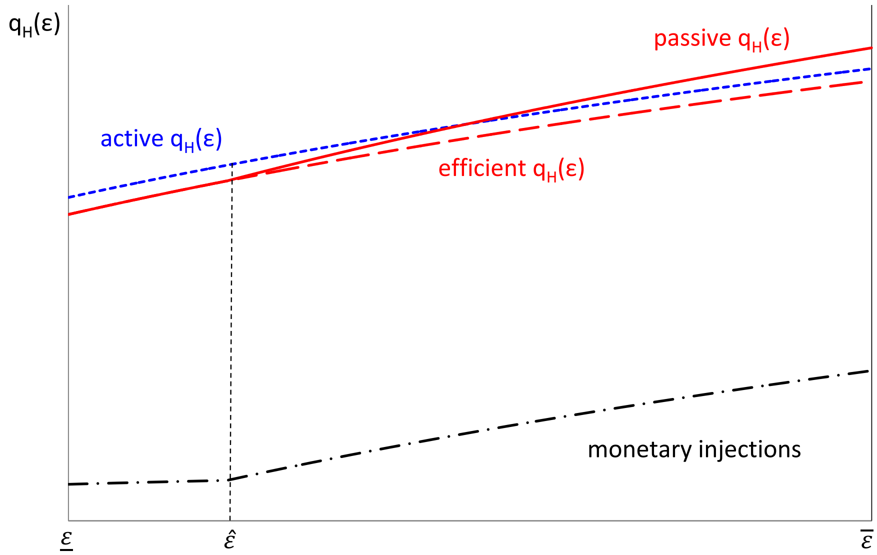

Figure 3: Stabilization versus passive policy for patient (type H) agents

In sum, we show that, with stabilization policy in response to demand shocks at the ZLB, the central bank is able to temporarily relax the liquidity constraint on impatient agents. This improves their welfare without harming the patient agents. Furthermore, we demonstrate that due to the pecuniary externality illustrated in Proposition 1, stabilization policy may also have beneficial general equilibrium effects for the patient agents even if they are unconstrained in their holdings of real balances.[9]

Interestingly, standard open market operations are neutral in this environment, even with discount factor heterogeneity. If $i = 0$, the patient agents' liquidity premium $[u'(q_H)/c'(y) - 1]$ is zero and then it must be that these agents are satiated with money, meaning that they don't earn any extra marginal benefit from holding money versus bonds. This means that if the central bank trades money for bonds in a situation in which $i = 0$, the patient agents are willing to absorb all the trade, leaving all other agents at the same allocations. So, traditional open market operations are neutral.

8 Conclusion

We construct a New Monetarist model characterized by aggregate liquidity shocks and heterogeneous idiosyncratic preference shocks. Agents are heterogeneous with respect to the idiosyncratic

[9] Berentsen, Huber and Marchesiani (2014), Chiu, Dong and Shao (2018) and Rojas Breu (2013) get similar pecuniary externalities when agents have differing abilities to pay for goods. However, at the ZLB the pecuniary externality disappears in their models whereas that is not the case in our environment.

shock, so that in every period some agents are more patient than others. This heterogeneity generates a distribution in asset holdings. In this environment, we then study the optimal stabilization response of the central bank to aggregate demand shocks that hit the economy.

Several results hold in this environment. First, a zero-interest-rate policy is not necessarily the best a central bank can implement due to a pecuniary externality. However, even when this is the best policy, not all agents are satiated at the ZLB and therefore there is scope for central bank policies of liquidity provision. Second, we study a form of stabilization policy whereby the central bank engages in repo arrangements in response to aggregate demand shocks. We find such a policy is welfare improving even at the ZLB since it can relax the liquidity constraint of impatient agents without harming the patient ones. This is true as long as the central bank commits to undo the injections at a later date. As a result, the liquidity injections are temporary. If they were permanent, instead, the only effect would be a jump in nominal prices and there would be no real effects on the economy. Third, due to a pecuniary externality, stabilization policy can be welfare improving for patient agents even if they are unconstrained at the ZLB.

References

Adam, K., Billi, R., (2006). "Optimal Monetary Policy under Commitment with a Zero Bound on Nominal Interest Rates." *Journal of Money, Credit, and Banking*, 38(7): 1877–1905.

Adam, K., Billi, R., (2008). "Monetary Conservatism and Fiscal Policy." *Journal of Monetary Economics*, 55: 1376–1388.

Aruoba, S.B., Rocheteau, G., Waller, C.J., (2007). "Bargaining and the Value of Money." *Journal of Monetary Economics*, 54(8): 2636–2655.

Auclert, A., (2017). "Monetary Policy and the Redistribution Channel." NBER Working Paper 23451.

Auerbach, A.J., Obstfeld, M., (2005). "The Case for Open-Market Purchases in a Liquidity Trap." *American Economic Review*, 95(1): 110-137.

Bayer, C., Lütticke, R., Pham-Dao, L., Tjaden, V., (2018). "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk." *Econometrica*, forthcoming.

Becker, R., (1980). "On the Long-Run Steady State in a Simple Dynamic Model of Equilibrium with Heterogeneous Households." *Quarterly Journal of Economics*, 95(2): 375–382.

Berentsen, A., Camera, G., Waller, C.J., (2007). "Money, Credit and Banking." *Journal of Economic Theory*, 135(1): 171–195.

Berentsen, A., Waller, C.J., (2011). "Price-Level Targeting and Stabilization Policy." *Journal of Money, Credit and Banking*, 43(7): 559–580.

Berentsen, A., Huber, S., Marchesiani, A., (2014). "Degreasing the Wheels of Finance." *International Economic Review*, 55(3): 735–763.

Billi, R., (2011). "Optimal Inflation for the US Economy." *American Economic Journal: Macroeconomics*, 3: 29–52.

Boel, P., Camera, G., (2006). "Efficient Monetary Allocations and the Illiquidity of Bonds." *Journal of Monetary Economics*, 53(7): 1693–1715.

Campbell, J.R., Fisher, J.D.M., Justiniano, A., Melosi, L. (2016). "Forward Guidance and Macroeconomic Outcomes Since the Financial Crisis". In NBER Macroeconomics Annual 2016, 31: 283-357. M. Eichenbaum and J.A. Parker, editors.

Carroll, C.D., Samwick, A.A., (1997). "The Nature of Precautionary Wealth." *Journal of Monetary Economics*, 40(1): 41–71.

Chiu, J., Dong, M., Shao, E., (2018). "On the Welfare Effects of Credit Arrangements." *International Economic Review*, 59(3): 1621–1651.

Christiano, L., Eichenbaum, M., Rebelo, S., (2011). "When Is the Government Spending Multiplier Large?" *Journal of Political Economy*, 119(1): 78–121.

Eggertsson, G.B., Woodford, M., (2003). "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity*, Vol. 1: 139–211.

Eggertsson, G.B., Juelsrud, R.E., Wold, E. (2017). "Are Negative Nominal Interest Rates Expansionary?" NBER Working Paper 4038.

Galí, J. (2017). "The Effects of a Money-Financed Fiscal Stimulus." CEPR Discussion Paper 10165.

Gertler, M., Karadi, P., (2013). "QE 1 vs. 2 vs. 3...: A Framework to Analyzing Large Scale Asset Purchases as a Monetary Policy Tool." *International Journal of Central Banking* 9:5–53.

Gornemann, N., Kuester, K., Nakajima, K., (2016). "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy." International Finance Discussion Papers 1167, Board of Governors of the Federal Reserve System (U.S.).

Gu, C., Han, H., Wright, R., (2016). "The Effects of Monetary Policy and Other Announcements." Working Papers 1621, Department of Economics, University of Missouri.

Guerrieri, V., Lorenzoni, G., (2017). "Credit Crises, Precautionary Savings, and the Liquidity Trap." *Quarterly Journal of Economics* 132(3): 1427–67.

Jung, T., Teranishi, Y., Watanabe, T., (2005). "Optimal Monetary Policy at the Zero-Interest-Rate Bound." *Journal of Money, Credit, and Banking*, 37(5): 813–835.

Kaplan, G., Moll, B., Violante, G.L., (2018). "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

Kaplan, G., Violante, G.L., Weidner, J., (2014). "The Wealthy Hand-to-Mouth." *Brookings Papers on Economic Activity* 48 (Spring): 77–153.

Kocherlakota, N., (1998). "Money is Memory." *Journal of Economic Theory*, 81(2): 232-251.

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

Lawrence, E.C., (1991). "Poverty and the Rate of Time Preference: Evidence from Panel Data." *Journal of Political Economy*, 99(1): 54–77.

McKay, A., Nakamura, E., Steinsson, J., (2016). "The Power of Forward Guidance Revisited." *American Economic Review*, 106(10): 3133–58.

McKay, A., Reis, R., (2016). "The Role of Automatic Stabilizers in the U.S. Business Cycle." *Econometrica*, 84(1): 141–194.

Ramsey, F.P., (1927). "A Contribution to the Theory of Taxation." *The Economic Journal*, 37: 47–61.

Rocheteau, G., Wright, R., (2005). "Money in Search Equilibrium, in Competitive Equilibrium and in Competitive Search Equilibrium." *Econometrica*, 73(1): 175–202.

Rocheteau, G., Wright, R., Xiao, S., (2018). "Open Market Operations." *Journal of Monetary Economics*, 98: 114–128 .

Rojas Breu, M., (2013). "The Welfare Effect of Access to Credit." *Economic Inquiry*, 51(1): 235–247.

Samwick, A.A., (1998). "Discount Rate Heterogeneity and Social Security Reform." *Journal of Development Economics*, 57(1): 117–46.

Werning, I., (2015). "Incomplete Markets and Aggregate Demand." NBER Working Paper 21448.

Williamson, S., (2016). "Scarce Collateral, the Term Premium, and Quantitative Easing." textitJournal of Economic Theory, 164: 136–165.

Woodford, M., (2012). "Methods of Policy Accommodation at the Interest-Rate Lower Bound." In *The Changing Policy Landscape*, Proceedings - Economic Policy Symposium - Jackson Hole, Federal Reserve Bank of Kansas City, pages 185-288.

Appendix 1: Proofs

Conditions for $n^k_{jz} > 0$. We now want to provide conditions that guarantee $n^k_{j,z} \geq 0$ in the constrained-efficient equilibrium with $i = 0$. Note that if $n^s_{HL} > 0$, then $n^k_{jz} \geq 0$ in all other cases. We know that $x^k_{jz} = x^*$ for all j, z. This, together with the budget constraint in (8), implies:

$$n^s_{HL} = x^* - m^s_H + \pi m_L - \pi p_a a_L + a_H - \tau - \tau_2$$

From (5), Lemma 1 and (7) the expression above becomes:

$$n^s_{HL} = x^* - m_H - \tau_1 - py + \beta_H m_L - A[\beta_H + (1-\rho)/\rho] - (\beta_H - 1)(\rho m_H + (1-\rho)m_L) - \tau_2$$

Since $\tau_1 + \tau_2 = 0$ and $py = \rho(m_H + \tau_1) + (1-\rho)(m_L + \tau_1)$, rearranging we get:

$$n^s_{HL} = x^* - A[\beta_H + (1-\rho)/\rho] - m_H - \tau_1 - \rho\beta_H[\rho m_H + (1-\rho)m_L]$$

Note that for $\pi = \beta_H$ we have that $m_H - \tau_1 = q^*$. Let

$$K = \rho\beta_H[\rho m_H + (1-\rho)m_L] + A[\beta_H + (1-\rho)/\rho] \tag{28}$$

Then, in order to have $n^s_{HL} > 0$ it must be that $x^* - q^* > K$. Since $K > 0$, then x^* must be sufficiently bigger than q^* in order to have $n^s_{HL} > 0$. Note that from (28) the necessary difference between q^* and x^* will depend on A—a tighter borrowing constraint will generate an incentive to work.

Proof of Lemma 1 From the Euler equation in (23) we have that the following must hold:

$$\beta_L + \lambda_L = \beta_H + \lambda_H$$

Since $\beta_H > \beta_L$, it must be that $\lambda_L > \lambda_H \geq 0$. If $\lambda_L > \lambda_H > 0$, then there is no borrowing or lending. If instead $\lambda_L > \lambda_H = 0$, then $a_L = A$ and given the bonds market clearing condition:

$$\rho a_H + (1-\rho)a_L = 0 \tag{29}$$

Thus, we have that $a_H = -A(1-\rho)/\rho$. Since $\pi p_a = \beta_H$ from (23), then $p_a = \pi/\beta_H$. ■

Proof of Lemma 2 We know from (16) and (18) that $\varepsilon u'(q_j) \geq c'(y)$ for $j = H, L$. This, together with (22), implies that $\pi \geq \beta_H$. ■

Proof of Proposition 1 From (22) and (18) we know that if $\pi = \beta_H$ then $\varepsilon u'(q_L) > c'(y)$, thus implying type L agents are constrained and $q_L < q^*$ for $c''(y) \geq 0$. From (22) and (18) we also know that if $\pi = \beta_H$ then $\varepsilon u'(q_H) = c'(y)$, thus implying type H agents cannot be constrained and $q_H \geq q^*$. Assume $q_H = q^*$. Since $q_L < q^*$, then we have that $y < y^*$ where $y^* = (\alpha_b/\alpha_s)q^*$. With $c''(y) = 0$ this would imply $\varepsilon u'(q^*) = c'(y) = c'(y^*)$ since $c'(y)$ is constant and therefore $q_H = q^*$. With $c''(y) > 0$, instead, $y < y^*$ implies $\varepsilon u'(q^*) = c'(y) < c'(y^*)$. Therefore, it cannot be that $q_H = q^*$ and it must be $q_H > q^*$. ■

Bargaining Suppose buyers and sellers are pairwise matched in the DM. Suppose also that buyer and seller bargain on the price and quantity of goods to be delivered using a Nash bargaining protocol with threat points given by the continuation values. Let θ and $1-\theta$ denote the buyer's and the seller's bargaining powers respectively and let d_j denote the dollars the buyer of type j pays the seller. Then, the terms of trade (q_j, d_j) between a buyer of type j with real money holdings m_j and a seller of any type must satisfy:

$$\max_{q_j, d_j} \ [u(q_j) - d_j]^{\theta} [-c(y) + d_j]^{1-\theta} \tag{30}$$

subject to $d_j \leq m_j$. Note that the terms of trade do not depend on the seller's type but only on the buyer's real balances. This is due to the linearity of the payoffs, as shown in Aruoba, Rocheteau and Waller (2007). Moreover, as in Lagos and Wright (2005), it can be shown that in any equilibrium it must be the case that $d_j = m_j$. Thus, in order to find q_j, we take the partial derivative of the expression in (30) with respect to q_j and set it equal to zero. This implies q_j is the solution to:

$$d_j = \frac{\theta u'(q_j)c(y) + (1-\theta)c'(y)u(q_j)}{\theta u'(q_j) + (1-\theta)c'(y)}$$

Implicit differentiation implies:

$$q_j'(\varepsilon) = \frac{[\theta u'(q_j) + (1-\theta)c'(y)]^2}{u'(q_j)c'(y)\,[\theta u'(q_j) + (1-\theta)c'(y)] + \theta(1-\theta)\,[u(q_j) - c(y)]\,[u'(q_j)c''(y) - c'(y)u''(q_j)]} \tag{31}$$

One can prove that equilibrium consumption in the DM satisfies the Euler equation:

$$\frac{\pi - \beta_j}{\beta_j} = \int_{\Omega} \left\{ \alpha_b \left[\varepsilon u'(q_j(\varepsilon)) q_j'(\varepsilon) - 1 \right] \right\} f(\varepsilon) d\varepsilon, \tag{32}$$

If the buyer has all the bargaining power and $\theta = 1$ the expression in (31) reduces to $c'(y)$ so that (32) is analogous to (22) we derived for competitive pricing. In each match, however, $y_L = q_L$ and $y_H = q_H$ with bargaining. ■

Proof of Proposition 2 The optimal π maximizes:

$$\underset{\pi}{Max} \;\; \alpha_b \left[(1-\rho) \varepsilon u(q_L) + \rho \varepsilon u(q_H) \right] - \alpha_s c(y) \tag{33}$$

subject to the constraints

$$\begin{aligned} \text{s.t.} \;\; \frac{\pi - \beta_H}{\beta_H} &= \alpha_b \left[\frac{\varepsilon u'(q_H)}{c'(y)} - 1 \right] \\ \text{s.t.} \;\; \frac{\pi - \beta_L}{\beta_L} &= \alpha_b \left[\frac{\varepsilon u'(q_L)}{c'(y)} - 1 \right] \end{aligned} \tag{34}$$

and the DM market clearing condition (19). By differentiating the objective function in (33), we know that in order for $\pi = \beta_H$ to be the optimal policy, the following condition must hold:

$$(1-\rho) \left[\varepsilon u'(q_L) - c'(y) \right] \left. \frac{dq_L}{d\pi} \right|_{\pi = \beta_H} + \rho \left[\varepsilon u'(q_H) - c'(y) \right] \left. \frac{dq_H}{d\pi} \right|_{\pi = \beta_H} \leq 0 \tag{35}$$

From the Euler equations we know that $\varepsilon u'(q_H) - c'(y) = 0$ and $\varepsilon u'(q_L) - c'(y) > 0$ at the Friedman rule. Therefore, (35) becomes:

$$(1-\rho) \left[\varepsilon u'(q_L) - c'(y) \right] \left. \frac{dq_L}{d\pi} \right|_{\pi = \beta_H} \leq 0$$

which implies the Friedman rule will only be optimal if $\left. \frac{dq_L}{d\pi} \right|_{\pi = \beta_H} \leq 0$. Now, if we totally differentiate the constraints in (34), we find the following system of equations has to hold:

$$\begin{bmatrix} n\frac{\varepsilon u''(q_H)}{c'(y)} - n\frac{\varepsilon u'(q_H)c''(y)}{c'(y)^2}\frac{n}{s}\rho & -n\frac{\varepsilon u'(q_H)c''(y)}{c'(y)^2}\frac{n}{s}(1-\rho) \\ -n\frac{\varepsilon u'(q_L)c''(y)}{c'(y)^2}\frac{n}{s}\rho & n\frac{\varepsilon u''(q_L)}{c'(y)} - n\frac{\varepsilon u'(q_L)c''(y)}{c'(y)^2}\frac{n}{s}(1-\rho) \end{bmatrix} \begin{bmatrix} \frac{dq_H}{d\pi} \\ \frac{dq_L}{d\pi} \end{bmatrix} = \begin{bmatrix} \frac{1}{\beta_H} \\ \frac{1}{\beta_L} \end{bmatrix}$$

where the determinant D is:

$$D = \frac{\alpha_b^2}{c'(y)^2}\left[\varepsilon u''(q_H) - \kappa_H\right]\left[\varepsilon u''(q_L) - \kappa_L\right] - \frac{\alpha_b^2}{c'(y)^2}\kappa_L\kappa_H$$

with $\kappa_H = (\alpha_b/\alpha_s)\rho\varepsilon u'(q_H)c''(y)/c'(y)$ and $\kappa_L = (\alpha_b/\alpha_s)(1-\rho)\varepsilon u'(q_L)c''(y)/c'(y)$. Using Cramer's rule we have:

$$\frac{dq_L}{d\pi} = \frac{\alpha_b}{c'(y)D}\left\{\left[\varepsilon u''(q_H) - \frac{\varepsilon u'(q_H)c''(y)}{c'(y)}\frac{\alpha_b}{\alpha_s}\rho\right]\frac{1}{\beta_L} + \left[\frac{\varepsilon u'(q_L)c''(y)}{c'(y)}\frac{\alpha_b}{\alpha_s}\rho\right]\frac{1}{\beta_H}\right\}$$

and

$$\left.\frac{dq_L}{d\pi}\right|_{\pi=\beta_H} = \frac{\alpha_b}{c'(y)D}\frac{1}{\beta_L}\left\{\varepsilon u''(q_H) + \left[\frac{u'(q_L)}{u'(q_H)}\frac{\beta_L}{\beta_H} - 1\right]c''(y)\frac{\alpha_b}{\alpha_s}\rho\right\}$$

Since

$$D|_{\pi=\beta_H} = \frac{\alpha_b^2}{c'(y)}[(\varepsilon u''(q_H) - \kappa_H)(\varepsilon u''(q_L) - \kappa_L) - \alpha_b^2\kappa_L\frac{\alpha_b}{\alpha_s}\rho c''(y)/c'(y)], \tag{36}$$

then at $\pi = \beta_H$ it must be that

$$\frac{dq_L}{d\pi} = \frac{\varepsilon u''(q_H) + \left[\frac{u'(q_L)}{u'(q_H)}\frac{\beta_L}{\beta_H} - 1\right]\frac{\alpha_b}{\alpha_s}\rho c''(y)}{\alpha_b\beta_L\left\{\left[\varepsilon u''(q_H) - c''(y)\frac{\alpha_b}{\alpha_s}\rho\right]\left[\varepsilon u''(q_L) - \frac{u'(q_L)c''(y)}{u'(q_H)}\frac{\alpha_b}{\alpha_s}(1-\rho)\right] - \left[\frac{\varepsilon u'(q_L)}{c'(y)^2}\frac{\alpha_b}{\alpha_s}\rho\right]\left[\frac{\alpha_b}{\alpha_s}(1-\rho)\right]c''(y)^2\right\}}$$

Note that if $c''(y) = 0$ then $\left.\frac{dq_L}{d\pi}\right|_{\pi=\beta_H} < 0$, which implies the Friedman rule is always the optimal policy with linear costs. If instead if $c''(y) > 0$, then the sign of $\left.\frac{dq_L}{d\pi}\right|_{\pi=\beta_H} < 0$ is indeterminate and therefore the Friedman rule is not necessarily optimal. Note also that (36) can be simplified as:

$$\begin{aligned} D|_{\pi=\beta_H} &= -\frac{\alpha_b^2}{\alpha_s^2 c'(y)^3}\varepsilon(u'(q_L)\alpha_b^2 c''(y)^2\rho(1-\rho)\left[1 - c'(y)\right] - \alpha_s^2 u''(q_H)c'(y)^2 u''(q_L)\varepsilon) \\ &\quad -\frac{\alpha_b^2}{\alpha_s^2 c'(y)^3}\varepsilon c''(y)\alpha_s c'(y)\varepsilon\left[u'(q_L)u''(q_H)(1-\rho) + \alpha_b u'(q_H)u''(q_L)\rho\right] \end{aligned}$$

Therefore, if $c''(y) > 0$ then $D|_{\pi=\beta_H} > 0$ if and only if $c'(y) \geq 1$. If $c'(y) \geq 1$, then $\left.\frac{dq_L}{d\pi}\right|_{\pi=\beta_H} < 0$ if $u'(q_L)/u'(q_H) < \beta_H/\beta_L$. ■

Proof of Proposition 3 From (22) we have:

$$\frac{\pi - \beta_L}{\beta_L} = \int_{\underline{\varepsilon}}^{\overline{\varepsilon}}\left\{\alpha_b\left[\frac{\varepsilon u'(q_L(\varepsilon))}{c'(y(\varepsilon))} - 1\right]\right\}f(\varepsilon)d\varepsilon \tag{37}$$

Let $g_L(\varepsilon)$ denote real aggregate spending of type L agents when their trades are efficient, i.e.

$g_L(\varepsilon) = \alpha_b(1-\rho)p(\varepsilon)q^*(\varepsilon)$. We want to understand how changes in ε affect $g_L(\varepsilon)$:

$$dg_L(\varepsilon) = \alpha_b(1-\rho)\left[q^*(\varepsilon)dp + p(\varepsilon)dq^*\right]$$

The first term denotes the change in the relative price $p(\varepsilon)$ and the second one changes in the efficient quantity $q^*(\varepsilon)$. We can rewrite the expression for $dg_L(\varepsilon)$ as follows:

$$dg_L(\varepsilon) = \alpha_b(1-\rho)pq^*\left[\frac{dp}{p} + \frac{dq^*}{q^*}\right]$$

From (16) we derive that:

$$\frac{dp}{p} = 0$$

The term dq^*/q^*, instead, can be derived from $\varepsilon u'(q^*) = c'[(\alpha_b/\alpha_s)q^*]$:

$$\frac{dq^*}{q^*} = -\frac{\varepsilon u'(q^*)}{\varepsilon u''(q^*) - c''[(\alpha_b/\alpha_s)q^*](\alpha_b/\alpha_s)}\frac{d\varepsilon}{\varepsilon}$$

so that:

$$\frac{dg_L(\varepsilon)}{d\varepsilon} = -\frac{\alpha_b(1-\rho)c'[(\alpha_b/\alpha_s)q^*]q^*u'(q^*)}{\varepsilon u''(q^*) - c''[(\alpha_b/\alpha_s)q^*](\alpha_b/\alpha_s)} > 0 \text{ for } c''(y) \geq 0$$

Let's first consider the case $c''(y) = 0$. If $g_L(\varepsilon) > m_L$, then agents are constrained in all states. If $g_L(\bar{\varepsilon}) < m_L$, then agents are never constrained. If $g_L(\bar{\varepsilon}) \geq m_L \geq g_L(\varepsilon)$, for a given value of m_L there exists a critical value $\tilde{\varepsilon}$ such that $g_L(\tilde{\varepsilon}) = m_L$. This implies that $q_L(\varepsilon) = q^*(\varepsilon) = q_H(\varepsilon)$ for $\varepsilon \leq \tilde{\varepsilon}$ and $q_L(\varepsilon) < q^*(\varepsilon) = q_H(\varepsilon)$ for $\varepsilon > \tilde{\varepsilon}$. The right-hand side (RHS) of (22) is a function of m_L. Note that $\lim_{m_L \to 0} RHS = \infty$ and, for $\bar{m}_L = g(\bar{\varepsilon})$, $RHS|_{\bar{m}_L} = 0 \leq (\pi - \beta_L)/\beta_L$. Since RHS is continuous in m_L then an equilibrium exists. The RHS of (22) is also monotonically decreasing in m_L. To see this use Leibnitz's rule and note that by construction $q_L(\check{\varepsilon}) = q^*(\tilde{\varepsilon})$ to get:

$$\frac{\partial RHS}{\partial m_L} = \int_{\check{\varepsilon}}^{\bar{\varepsilon}} \left\{\alpha_b \left[\frac{\varepsilon\left[u''c' - u'c''(\alpha_b/\alpha_s)(1-\rho)\right]}{(c')^2}\frac{\partial q_L}{\partial m_L}\right]\right\} f(\varepsilon)d\varepsilon < 0$$

Since the RHS is strictly decreasing in m_L, we have a unique m_L that solves (22). Consequently, we have $q_L(\varepsilon) = q^*(\varepsilon)$ if $\varepsilon \leq \tilde{\varepsilon}$ and $q_L(\varepsilon) < q^*(\varepsilon)$ otherwise.

The argument is analogous for the case $c''(y) > 0$ and there exists a unique critical value $\hat{\varepsilon}$ such that $g_L(\hat{\varepsilon}) = m_L$. However, we know from Proposition 1 that when type L agents are constrained type H ones consume more than q^*. This implies that $q_L(\varepsilon) = q^*(\varepsilon) = q_H(\varepsilon)$ for $\varepsilon \leq \hat{\varepsilon}$ and $q_L(\varepsilon) < q^*(\varepsilon) < q_H(\varepsilon)$ for $\varepsilon > \hat{\varepsilon}$. ■

Proof of Proposition 4 The Lagrangian for (27) is:

$$\mathcal{L} = \underset{q_L(\varepsilon), y(\varepsilon)}{Max} \quad \int_{\Omega} \{\alpha_b \varepsilon [(1-\rho) u(q_L(\varepsilon)) + \rho u(q_H(\varepsilon))] - \alpha_s c\left(y(\varepsilon)\right)\} f(\varepsilon) d\varepsilon$$

$$+\lambda_R \left[\int_{\Omega} \left\{ \alpha_b \left[\frac{\varepsilon u'(q_L(\varepsilon))}{c'\left(y(\varepsilon)\right)} - 1 \right] \right\} f(\varepsilon) d\varepsilon - \frac{\pi - \beta_L}{\beta_L} \right] + \mu(\varepsilon) \left[y(\varepsilon) - \frac{\alpha_b}{\alpha_s} \left(\rho q_H(\varepsilon) + (1-\rho) q_L(\varepsilon) \right) \right]$$

Note that μ is a function of ε because the resource constraint varies state by state. The first-order condition for $q_L(\varepsilon)$ implies:

$$\mu(\varepsilon) = \varepsilon \alpha_s \frac{(1-\rho) u'(q_L(\varepsilon)) + \lambda_R u''\left[q_L(\varepsilon)\right] / c'\left(y(\varepsilon)\right)}{1-\rho} \tag{38}$$

The first order condition with respect to $y(\varepsilon)$ instead yields:

$$\alpha_b \rho \varepsilon u'(q_H(\varepsilon)) \frac{dq_H(\varepsilon)}{dy(\varepsilon)} - \alpha_s c'\left(y(\varepsilon)\right) - \lambda_R \frac{\alpha_b \varepsilon u'(q_L(\varepsilon)) c''\left(y(\varepsilon)\right)}{c'\left(y(\varepsilon)\right)^2} + \mu(\varepsilon) \left[1 - \frac{\alpha_b}{\alpha_s} \rho \frac{dq_H(\varepsilon)}{dy(\varepsilon)} \right] = 0 \tag{39}$$

Note that, since $\varepsilon u'(q_H(\varepsilon)) = c'(y(\varepsilon))$ at $i = 0$, from the implicit function theorem we know that

$$\frac{dq_H(\varepsilon)}{dy(\varepsilon)} = \frac{c''\left(y(\varepsilon)\right)}{\varepsilon u''(q_H(\varepsilon))} \tag{40}$$

Therefore, combining (38), (39) and (40) we find that the following expression holds for $c''(y) \geq 0$:

$$\varepsilon u'(q_L(\varepsilon)) - c'\left(y(\varepsilon)\right) = \frac{\lambda_R}{(1-\rho)\, c'\left(y(\varepsilon)\right)^2} \frac{\alpha_b \varepsilon u'(q_L(\varepsilon)) c''\left(y(\varepsilon)\right) - \alpha_b \alpha_s u''(q_L(\varepsilon))}{\alpha_s - \alpha_b \rho c''\left(y(\varepsilon)\right) / \varepsilon u''(q_H(\varepsilon))} \tag{41}$$

At this point we need to consider two cases. In the first case, $m_L > 0$ and therefore the first constraint in (27) holds with equality so that $\lambda_R > 0$. Therefore, $\varepsilon u'(q_L(\varepsilon)) > c'\left(y(\varepsilon)\right)$ in (41), which implies $q_L(\varepsilon) < q^*(\varepsilon)$. In the second case, $m_L = 0$ and therefore $\lambda_R = 0$. This in turn implies $\varepsilon u'(q_L(\varepsilon)) = c'\left(y(\varepsilon)\right)$ in (41), so that $q_L(\varepsilon) = q_H(\varepsilon) = q^*(\varepsilon)$.

Since $q_L(\varepsilon) < q^*(\varepsilon)$ when $m_L > 0$, then we know from Propositions 1 and 3 that in all states $q_H(\varepsilon) = q^*(\varepsilon)$ if $c''\left(y\right) = 0$ and $q_H(\varepsilon) > q^*(\varepsilon)$ if $c''\left(y\right) > 0$ instead. ■

Appendix 2: Examples derivations

Derivations for Example 1. We consider the following functional forms:

$$u(q) = 1 - \exp^{-q} \quad \text{and} \quad c(y) = \exp^{y} - 1$$

In this case, the constraints in the optimal inflation problem become:

$$\begin{aligned}
\frac{\pi - \beta_H}{\beta_H} &= \alpha_b \left[\frac{\varepsilon \exp^{-q_H}}{\exp^{y}} - 1 \right] \\
\frac{\pi - \beta_L}{\beta_L} &= \alpha_b \left[\frac{\varepsilon \exp^{-q_L}}{\exp^{y}} - 1 \right] \\
\alpha_s y &= \alpha_b [\rho q_H + (1 - \rho) q_L]
\end{aligned} \tag{42}$$

The derivative of the objective function with respect to π yields:

$$\alpha_b \varepsilon \left[(1 - \rho) u'(q_L) \frac{dq_L}{d\pi} + \rho u'(q_H) \frac{dq_H}{d\pi} \right] - \alpha_s c'(y) \left[\rho \frac{dq_H}{d\pi} + (1 - \rho) \frac{dq_L}{d\pi} \right] \leq 0 \tag{43}$$

We now need to find expressions for $dq_L/d\pi$ and $dq_H/d\pi$. If we simplify the expressions in the constraints in (42) and take logs, we find that:

$$\begin{aligned}
q_H &= \frac{\alpha_s Z_H + \alpha_b (1 - \rho) (Z_H - Z_L)}{\alpha_s + \alpha_b} \\
q_L &= \frac{\alpha_s Z_L - \alpha_b \rho (Z_H - Z_L)}{\alpha_s + \alpha_b}
\end{aligned}$$

where $Z_H = \ln(\varepsilon) + \ln[\alpha_b \beta_H / (\pi - \beta_H + \alpha_b \beta_H)]$ and $Z_L = \ln(\varepsilon) + \ln[\alpha_b \beta_L / (\pi - \beta_L + \alpha_b \beta_L)]$. We let $\varepsilon = \exp$ so that $\ln(\varepsilon) = 1$. Then:

$$\frac{dq_L}{d\pi} = \frac{\alpha_b \rho}{(\alpha_s + \alpha_b)(\pi - \beta_H + \alpha_b \beta_H)} - \frac{\alpha_s + \alpha_b \rho}{(\alpha_s + \alpha_b)(\pi - \beta_L + \alpha_b \beta_L)}$$

and

$$\frac{dq_H}{d\pi} = \frac{\alpha_b (1 - \rho)}{(\alpha_s + \alpha_b)(\pi - \beta_L + \alpha_b \beta_L)} - \frac{\alpha_s + \alpha_b (1 - \rho)}{(\alpha_s + \alpha_b)(\pi - \beta_H + \alpha_b \beta_H)}$$

By plugging the expressions for $dq_L/d\pi$ and $dq_H/d\pi$ into (43), we find that in order for $\pi = \beta_H$ to be optimal it must be that $(1-\rho)\left[-(\alpha_s + \alpha_b \rho)(\alpha_b \beta_H) + \alpha_b \rho (\beta_H - \beta_L + \alpha_b \beta_L)\right]\left[\exp^{1 - Z_L} - 1\right] \leq 0$. Therefore, the following condition must hold:

$$\frac{\beta_H - \beta_L}{\beta_H} \leq \frac{\alpha_s}{\rho(1 - \alpha_b)}$$

Derivations for Example 2. Consider the following functional forms:

$$u(q) = 1 - \exp^{-q} \quad \text{and} \quad c(y) = \exp^{y} - 1$$

The central bank's problem is as in (27) and the first-order condition for $q_L(\varepsilon)$ and $y(\varepsilon)$ are as in (38) and (39) respectively. Combining (38) with (39) and the fact that with convex costs $dq_H(\varepsilon)/dy(\varepsilon) = c''[y(\varepsilon)]/\varepsilon u''[q_H(\varepsilon)]$ and solving for λ_R, we find:

$$\lambda_R = \frac{\left\{\varepsilon u'[q_L(\varepsilon)] - c'[y(\varepsilon)]\right\}\left[1 - \frac{\alpha_b}{\alpha_s}\rho\frac{c''[y(\varepsilon)]}{\varepsilon u''[q_H(\varepsilon)]}\right]}{\left\{\frac{\alpha_b \varepsilon u'[q_L(\varepsilon)]\, c''[y(\varepsilon)]}{\alpha_s c'[y(\varepsilon)]^2} - \frac{\varepsilon u''[q_L(\varepsilon)]}{(1-\rho)c'[y(\varepsilon)]}\left[1 - \frac{\alpha_b}{\alpha_s}\rho\frac{c''[y(\varepsilon)]}{\varepsilon u''[q_H(\varepsilon)]}\right]\right\}} \tag{44}$$

We now consider a uniform distribution with $\underline{\varepsilon} = \exp$ and we proceed as follows. First, we use the first constraint in the Ramsey problem in (27) for the case when $m_L(\varepsilon) > 0$ in all states to solve for $y(\underline{\varepsilon})$:

$$\frac{\pi - \beta_L}{\beta_L} = \int_{\varepsilon}^{\bar{\varepsilon}} \left\{\alpha_b \left[\varepsilon \exp^{\frac{\alpha_b \rho \ln(\varepsilon) - (\alpha_s + \alpha_b) y(\varepsilon)}{\alpha_b (1-\rho)}} - 1\right]\right\} f(\varepsilon) d\varepsilon$$

Second, since λ_R in (44) does not depend on any state ε, the following condition must hold for all ε such that $\underline{\varepsilon} \leq \varepsilon \leq \bar{\varepsilon}$ given an arbitrary state ε:

$$\lambda_R = \lambda_R\big|_{\underline{\varepsilon}} \text{ with } \lambda_R = \frac{\left[\varepsilon \exp^{-q_L(\varepsilon)} - \exp^{y(\varepsilon)}\right]\left[1 - \frac{\alpha_b}{\alpha_s}\rho\frac{\exp^{y(\varepsilon)}}{-\varepsilon \exp^{q_H(\varepsilon)}}\right]}{\left\{\frac{\alpha_b \varepsilon \exp^{-q_L(\varepsilon)} \exp^{y(\varepsilon)}}{\alpha_s [\exp^{y(\varepsilon)}]^2} - \frac{-\varepsilon \exp^{-q_L(\varepsilon)}}{(1-\rho)\exp^{y(\varepsilon)}}\left[1 - \frac{\alpha_b}{\alpha_s}\rho\frac{\exp^{y(\varepsilon)}}{-\varepsilon \exp^{q_H(\varepsilon)}}\right]\right\}}$$

We use this to solve for $y(\varepsilon)$ in terms of $y(\underline{\varepsilon})$ for all ε. Then, we use the resource constraint (19) to solve for $q_L(\varepsilon)$ as a function of $y(\varepsilon)$:

$$q_L(\varepsilon) = \frac{y(\varepsilon)(\alpha_s + \alpha_b \rho) - \alpha_b \rho \ln(\varepsilon)}{\alpha_b (1-\rho)}$$

Last, from $\varepsilon u'[q_H(\varepsilon)] = c'[y(\varepsilon)]$ we find an expression for $q_H(\varepsilon)$ as a function of $y(\varepsilon)$. With the functional forms we chose, the condition is $q_H(\varepsilon) = \ln(\varepsilon) - y(\varepsilon)$.