Optimal Monetary Policy and Liquidity with Heterogeneous Households[I]

Florin O. Bilbiie[II] Xavier Ragot[III]

PSE and CEPR *Sciences Po and OFCE*

September 2017 (First draft February 2016)

Abstract

A novel *liquidity-insurance* motive for monetary policy implies optimal deviations from price stability when heterogeneous households who participate infrequently in financial markets use liquidity to insure idiosyncratic risk. In our tractable sticky-price model that can be solved in closed form, aggregate demand depends on liquidity. The liquidity-insurance motive changes the central bank's trade-off, which is nevertheless still described by a second-order approximation to aggregate welfare. Price stability has significant welfare costs because inflation volatility hinders the consumption volatility of constrained households as a side-effect of liquidity-insuring them. Helicopter drops are a better way to achieve this insurance than open-market operations.

JEL Codes: D14, D31, E21, E3, E4, E5

Keywords: *optimal (Ramsey) monetary policy; heterogenous households; incomplete markets; liquidity constraints; limited participation; money; insurance, inequality; helicopter drops; open market operations.*

[I]Previously circulated as part of "Monetary Policy and Inequality when Aggregate Demand Depends on Liquidity", then "Inequality, Liquidity, and Optimal Monetary Policy" and presented since March 2016. We are especially grateful to our first discussant Benjamin Moll and to Mirko Wiederholt for extensive comments, to Klaus Adam, Andrew Atkeson, Adrien Auclert, Paul Beaudry, Jess Benhabib, Edouard Challe, Wei Cui, Ester Faia, Nobu Kiyotaki, Jordi Galí, Alejandro Justiniano, Dirk Krueger, John Leahy, Eric Leeper, Francesco Lippi, Roland Meeks, Cyril Monnet, Galo Nuno, Morten Ravn, Pontus Rendahl, Gilles Saint-Paul, Vincent Sterk, Paolo Surico, Andrea Tambalotti, Gianluca Violante, Pierre-Olivier Weill, and Michael Woodford for comments and suggestions, and to other participants in several conferences and seminars. Bilbiie thanks (without implicating) Banque de France for financial support through the eponymous Chair at PSE, Institut Universitaire de France, and NYU and CREi for hospitality during part of writing this paper.

[II]Paris School of Economics, 48 Boulevard Jourdan 75014 Paris; florin.bilbiie@parisschoolofeconomics.eu; http://florin.bilbiie.googlepages.com.

[III]Sciences Po, CNRS, and OFCE; xavier.ragot@gmail.com; xavier-ragot.fr

1 Introduction

Leading central bankers recently focused extensively on the monetary policy implications of such considerations as *insurance*, redistribution and inequality—so much so that the last two Chairs of the Federal Reserve, the President and other board members of the European Central Bank dedicated entire speeches to the issue and explicitly called for more research on it: Bernanke (2007, 2015); Yellen (2014); Draghi (2016); Cœuré (2012); Mersch (2014).

The aftermath of the 2008 financial crisis and the ensuing Great Recession also saw an unprecedented *liquidity expansion*: to take one example, the year-on-year growth rate of M1 quadrupled (from 2.5 to 11 percent on average) in the post-crisis period as compared to the 2000-2008 interval; Figure 1 illustrates this, along with nominal GDP growth.[1] And among possible responses to the crisis, *helicopter drops* (HD) returned to the policy debate as an actual policy option.[2]

Yet a framework for the analysis of *optimal* monetary policy when all of the aforementioned issues matter (the implications of insurance, or inequality, over the cycle, its link with liquidity provision, and the means to provide that liquidity) is hitherto lacking. This is what this paper does: in a model where aggregate demand depends on liquidity, we identify a novel channel that we label the *liquidity-insurance motive* of optimal policy. This changes the standard stabilization objectives (of inflation and real activity): quantitatively, it implies significant optimal deviations from price stability in response to shocks that in standard sticky-price models generate no such trade-off.

Our goal is to contribute to the understanding of *optimal monetary policy* in a model that belongs to a new synthesis that is under way at the time of our writing. A very recent quantitative literature that we review below analyzes monetary policy transmission in incomplete-market, heterogeneous-agent New Keynesian models—abbreviated "HANK" by Kaplan, Moll, and Violante (2014). These contributions can speak to empirical findings documenting the link between expansionary monetary policy and redistribution,[3] and are also consistent with recent microeconometric evidence on the

[1]Since nominal GDP has actually *fallen* during the crisis and growth thereafter does not nearly match money growth, it follows that velocity sank during the crisis and kept falling. The picture is even more extreme when considering base money, whose growth rate during the three QE episodes is off the charts: above 100 percent in the crisis and its aftermath (for over a year), and around 40 percent in late 2011 and late 2013 episodes respectively.

[2]Much confusion surrounds this notion; we hope that our paper adds, as a side effect, to the clarification—one scope of academic literature on policy-relevant topics. "Helicopter drops" were proposed by Friedman (1969), although he in fact attributes the idea to Haberler (1952) and provides the following quote "Suppose the quantity of money is increased by tax reductions or government transfer payments, and the resulting deficit is financed by borrowing from the central bank or simply printing money". Notice that there is an inherent fiscal dimension to this that is stated very clearly, as it is in Wallace (1981, p. 267): "unbacked government liabilities, liabilities that I call fiat money". Leeper (1991) and Woodford (1996) among others provide more modern treatments on fiscal-monetary interactions.

[3]A general and robust conclusion of several papers (using a variety of methods and data sets) seems to be that looser monetary policy is associated with less inequality (e.g. through the redistribution effects of inflation). Starting from Doepke and Schneider (2006), these include i.a. Adam and Zhu (2014), Coibion, Gorodnichenko, Kueng, and

heterogeneity of marginal propensities to consume MPCs, its relation to liquidity constraints, and income and wealth distributions.[4]

We build a tractable general equilibrium model that belongs to this vintage, in order to revisit standard New Keynesian optimal monetary policy analysis. In our model, heterogeneous households are subject to liquidity constraints, and liquidity is used to self-insure against uninsurable risk: financial markets are incomplete as in Bewley, and participation is limited (infrequent) in the Baumol-Tobin tradition. In equilibrium, aggregate demand depends on *liquidity*, which we define as the nominal asset used by households to self-insure; we call it "money", but it can be any asset whose return is affected by monetary policy.[5] Liquidity is thus used in equilibrium as long as there is a need for insurance, or *inequality*—understood following the Bewely-Huggett-Aiyagari (heterogeneous-agent) literature as the endogenous outcome of uninsurable shocks combined with households' ability to self-insure.[6] We thus focus on the notion of liquidity that has a long tradition, going back at least to Friedman's (1969) analysis of the redistributive effect of monetary policy and to Bewley's (1983) formalization of that analysis.

Like many others, we consider that monetary policy is the relevant tool at business cycle (quarterly) frequency to improve the distorted market outcome. Thus, we analyze the residual trade-offs for monetary policy after the (imperfect) use of any fiscal tools (without considering time-varying fiscal tools as a policy instrument). But we do let fiscal policy do much in our model: in the baseline, it takes care of the monopolistic distortion by sales subsidies, which it finances by an implicit redistribution of profit income. Indeed, liquidity is an equilibrium phenomenon in our economy precisely because imperfect insurance subsists (fiscal policy does not undo inequality perfectly); the amount of liquidity demanded is thus an indirect metric of the insurance job left undone by fiscal policy. We in fact calibrate the degree of imperfect insurance in the model—which as we shall see is the main determinant of optimal inflation—to match one plausible data counterpart of this object: the fall in consumption at unemployment, which takes into account any fiscal transfers.[7]

Silvia (2013), Adam and Tzamourani (2016), Deutsche Bundesbank (2016), and Furceri, Loungani, and Zdzienicka (2016).

[4]Recent empirical evidence using micro data from various sources supports the hypothesis that high MPCs correspond to households who are liquidity constrained (rather than, say, income-poor); see Kaplan and Violante (2014), Cloyne et al (2016), Jappelli and Pistaferri (2014), Gorea and Midrigan (2015), Misra and Surico (2014) and Surico and Trezzi (2016).

[5]Our framework can hence accomodate nominal bonds, if they are used to self-insure and thus have a liquidity premium. See also the discussion of alternative choices to model liquidity in Section 2.1.

[6]Admittedly, by focusing on short-run business cycles and stabilization policy, our framework does not capture other important aspects of inequality, such as human capital accumulation, inequality along the age-dimension, and others—some of which operate in the richer models reviewed below. See Krueger, Mitman and Perri (2017) for a recent contribution and review of what is now a vast literature.

[7]Note that this is the standard in monetary policy analysis: even in the baseline, textbook NK model, if the fiscal authority had enough lump-sum instruments and the ability to use them at quarterly frequency, any cost-push

We study Ramsey-optimal monetary policy in this framework, and unveil a—to the best of our knowledge—novel channel that we call the *liquidity-insurance motive*, for short: with imperfect insurance (inequality) there is a rationale for providing liquidity, whose inflationary consequences' costs are generically dominated by its insurance benefits. In other words, the trade-off faced by a central bank changes: providing insurance through liquidity is consistent with its standard objectives of stabilizing inflation and aggregate demand—but our novel channel implies that inflation stabilization take a back seat.

We illustrate this analytically by providing a second-order approximation to the aggregate welfare function à la Woodford (2003, Ch. 6). There is scope for a planner to provide consumption insurance, an objective that is costly to achieve through inflation when prices are sticky (and absent a full set of fiscal instruments). This trade-off operates in the long-run, as in any monetary model, making deflation optimal by shrinking liquidity (as prescribed by the Friedman rule and its incomplete-market variants). But more importantly, and unlike other monetary sticky-price frameworks, the trade-off *also* operates in the short run in response to shocks: insofar as there is long-run inequality making the liquidity-insurance motive operative, optimal policy requires volatile inflation. What is more, this inflation volatility matters for welfare: a central bank that stabilizes inflation, albeit around an optimal long-run target, incurs a large welfare cost—consumers would pay (around 0.1 to 0.5 percent of consumption) to live in the economy with volatile inflation. Such deviations and welfare effects are larger than those encountered in existing monetary models with nominal rigidities.

Inflation volatility is beneficial in our economy because it dampens the consumption volatility of constrained households without much affecting the unconstrained, who can self-insure. The optimal policy consists of providing liquidity, which insures the constrained, and inflating away some of its value, in order to give the unconstrained the right intertemporal incentives to hold this liquidity for precautionary purposes.

Since the optimal policy consists of providing liquidity to insure in face of aggregate shocks, it is only natural that more direct ways of injecting this liquidity (such as helicopter drops HD) are preferable to indirect ways (such as open market operations OM). The former consist of injecting liquidity during the period so that it reaches all households, but hence also—most importantly—constrained ones, with unit marginal propensity to consume. While the latter (OM) consists of exchanging liquidity for other assets and transferring the proceeds only later through the consolidated budget—thus depriving the central bank of a within-the-period transfer. In the latter case, optimal policy thus needs to rely more on the Pigou effect, or on a distortionary tax: using (costly) inflation to influence the value of real balances of constrained households.[8] We provide a rigorous welfare comparison of

type shocks could be accommodated through variations in, e.g., labor-tax rates or sale subsidies—thus redistributing from firms to consumers. Similarly, the zero lower bound could be avoided by appropriate saving taxes (consumption subsidies). Standard analysis assumes that such perfect redistribution is unfeasible, which is what we also assume.

[8] The two means of money creation are equivalent for welfare when prices are flexible (and inflation is like a

the two policy arrangements by calculating Ramsey-optimal policy for each and find that, for a same change in government liabilities, implementing optimal policy through HD is preferable to the most favorable OM (whereby liquidity is transferred to households after one period only).

1.1 Related Literature

Our paper is related to several *literatures*. The model (that we then use for studying optimal policy) integrates two streams of monetary economics that evolved divergently over the past two decades: New Keynesian (NK) models with nominal rigidities, and microfounded models of money demand with flexible prices.[9] Within these frameworks, we connect their two *subsets* that focus on heterogeneity, market incompleteness and limited participation. One stream consists of **monetary theory models with limited participation and incomplete markets** in the Bewley and Baumol-Tobin tradition. In our model, money is used to self-insure against idiosyncratic shocks as in Bewley models, but only for non-participating agents as in the Baumol-Tobin literature. Some of the key contributions, *all with flexible prices*, include Bewley (1983); Scheinkman and Weiss (1986); Lucas, (1990); Kehoe, Levine, and Woodford (1992); Algan, Challe, and Ragot (2010); Alvarez and Lippi (2014); Khan and Thomas (2015); Cao et al (2016); Lippi, Ragni and Trachter (2015); Gottlieb (2015); Rocheteau, Weill and Wong (2015, 2016); and Ragot (2016).[10] Drawing on this literature, two assumptions are key to deliver our model's tractability. First, households participating in financial markets have a high income and join a family where risk is pooled—an extension of Lucas (1990), also used more recently by i.a. Challe et al. (2017). Second, a family head chooses the allocations of all households (including those not participating in financial markets who have a low income), under liquidity contraints. In the equilibrium that we focus on, non-participating households consume all their liquid wealth, and there are only two wealth states—instead of a whole distribution of wealth as in a fully-fledged Bewley model. This delivers Euler equations that preserve self-insurance (here, through liquidity or money demand), while capturing heterogeneity in a simplified manner.

The other stream of literature studies **heterogeneous agents in NK models**; an early, 2000s literature introduced "hand-to-mouth" consumers (or limited participation in asset markets) to study aggregate demand and monetary policy—one could call this "first-generation HANK". Galí, Lopez-Salido and Valles (2007) and Bilbiie (2008) are two early examples of such models, where a subset of agents are (employed) hand-to-mouth and have unit MPC.[11] Compared to these models, we allow for

non-distortionary tax): they just deliver different inflation and money balances paths, for a same real allocation.

[9] Money demand in the NK model is generically residual when money is introduced in the utility function, through a cash-in-advance constraint, or through shopping-time distortions. This has nevertheless important consequences for optimal policy, which we review in due course.

[10] Recent empirical work argues that such frictions are needed to explain money demand, including the distribution of money holdings across agents (i.a. Alvarez and Lippi 2009; Cao et al 2012; Ragot 2014).

[11] Gali et al (2004, 2007) distinguish households according to whether they hold physical capital or not and solve

temporarily-binding credit constraints and allow constrained agents to self-insure. So do some of the more recent, 2010s-vintage models referred to as HANK above: quantitative models with household heterogeneity and incomplete markets that are consistent with microeconomic heterogeneity and data on household finances, and replicate plausible distributions of wealth and MPCs. Among these, Kaplan, Violante, and Moll (2014, hereinafter KMV) revisit the transmission mechanism of monetary policy in such a model with liquid and illiquid assets. In contrast to representative-agent NK models where monetary policy works mainly through intertemporal substitution, in their model monetary policy works mainly through what they label an "indirect effect" (the endogenous, general-equilibrium response of output). Gornemann, Kuester, and Nakajima (2012) also studied monetary transmission when markets are incomplete and unemployment risk endogenous, focusing on the distributional welfare effects on households with different wealth levels.[12] McKay, Nakamura, and Steinsson (2015, hereinafter MNS) use a similar model to those mentioned above, but with exogenous unemployment risk, to show that forward guidance is less powerful than in the standard model—mostly because an incomplete-markets model implies a form of "discounting" of aggregate demand. Auclert (2015) analyzes the role of redistribution for the transmission mechanism and decomposes it into three channels that are related to households' asset positions, but in a model with one asset only.

Our simplified framework captures some key features and mechanisms of recent quantitative HANK models, yet it trades off some (relevant and important, but thoroughly analyzed elsewhere) complexity for analytical tractability. This allows us to analyze the transmission and design of optimal monetary policy, which are integral parts of the state-of-the-art NK framework.[13]

Because we use this model to look at optimal monetary policy, we owe much debt to the literature

the model numerically to study determinacy and fiscal multipliers. Bilbiie (2004, 2008) derives for the first time an analytical aggregate demand IS curve emphasizing the Keynesian amplification with hand-to-mouth agents, and studies optimal policy. Eggertsson and Krugman (2012) combine a very similar aggregate IS curve with a particular theory of the natural interest rate in order to build a fascinating story of deleveraging, debt deflation, and the liquidity trap. Nistico (2015) allows households to switch stochastically between the two states, and also computes optimal monetary policy. Yet another, separate but related stream studies "financial accelerator" models—see Kiyotaki and Moore (1997) and Iacoviello (2005). Gertler and Kiyotaki (2010) review this literature.

[12] Several other models combine incomplete markets, nominal rigidities, and search and matching frictions. Ravn and Sterk (2013) focus on unemployment risk and show that job uncertainty generates deep and lasting recessions through aggregate demand amplification. Den Haan, Rendhal and Riegler (2016) show that such a model with sticky wages delivers a deflationary spiral, a key element of which is (precautionary) demand for money (which enters the utility function); there is a role for unemployment insurance in that model, as in McKay and Reis (2015). Challe, Matheron, Ragot, and Rubio-Ramirez (2015) estimate a model of this vintage using Bayesian methods, and assess the quantitative importance of the link between precautionary saving and aggregate demand. Bayer, Luetticke, Pham-Dao, and Tjaden (2015) look at the effect of idiosyncratic uncertainty shocks.

[13] To fix ideas, one could argue that while the existing literature in this realm puts more emphasis on the "heterogeneous-agent" part of **HA**NK, our framework does the opposite—it simplifies heterogeneity to put more emphasis on the latter part of HA**NK**. In our opinion, the two approaches are complementary.

that, building on the seminal paper of Lucas and Stokey (1983), shaped our understanding of *optimal policy in NK models.* Some of the key contributions include Khan, King, and Wollman (2003), Adao, Correia, and Teles (2003), Woodford (2003, Ch. 6), Benigno and Woodford (2005, 2012), and Schmitt-Grohe and Uribe (2004, 2007).

Significant deviations from price stability are optimal in our framework, and not only in the long run—the cited papers also imply, when relying on (other, different) money demand theories, some convex combination between the Friedman rule and a zero inflation long-run prescription. More surprisingly, our framework also gives rise to significant optimal deviations from price stability *over the cycle*—in response to shocks that in existing frameworks do not generate such deviations. A welfare-maximizing central bank relies on inflation volatility optimally, as this inflation volatility is associated with providing liquidity for insurance and contributes to reducing inequality—even though, as we shall see, inflation is unconditionally "bad" for constrained households because it reduces the real value of their money balances. Renouncing this volatility (by adopting a policy of constant deflation at the optimal asymptotic rate) thus has a large welfare cost in our model, whereas it is innocuous in the NK models with money demand reviewed above. The key to this difference is inequality, and the liquidity-insurance motive mentioned above.

Ours is not the only paper to study **optimal monetary policy with heterogeneous households and sticky prices** (in HANK-type models). Several earlier studies analyzed optimal monetary policy in different heterogeneous-agents models, focusing on other channels. In the realm of two-agent models, Bilbiie (2008) derives optimal policy in a model with hand-to-mouth agents, and Curdia and Woodford (2009) and Nistico (2016) in models with infrequent participation and borrowers and savers. The setup of these last two papers shares similarities to ours, in particular concerning the "infrequent participation" structure that draws on an earlier monetary theory literature; but in the domain of optimal policy, these studies focus on the case where there is perfect insurance in steady state, thus abstracting from the liquidity-insurance, or inequality channel that gives rise to the novel trade-off we emphasize.[14]

Lastly, several more recent and independent papers deal broadly with the same topic but differ substantially and in several key respects: assumptions about the environment, solution techniques, results, and economic intuition and mechanisms. Differently from Bhandari, Evans, Golosov and Sargent (2017), we consider an economy with two assets to analyze the role of liquidity injections when liquidity constraints bind occasionally; this is also different relative to Challe (2017). The difference with respect to Nuno and Thomas (2017) is that we consider aggregate shocks, and infrequent

[14]An important difference between our model and Curida and Woodford's is also that our non-participant agents are liquidity constrained, and consume a liquidity injection that relaxes this constraint. Whereas in their setup, non-participant borrowers can borrow but subject to a spread. Braun and Nakajima (2012) prove an aggregation result, showing the conditions under which the optimal policy in an incomplete-market economy is the same as that of a representative-agent model.

participation in financial markets. In short, the key mechanism we focus on, differently from these papers, is precisely self-insurance through liquidity, in an economy with two types of assets liquid and illiquid, and limited participation.[15]

What distinguishes our framework is the introduction of limited participation in financial markets as a microfoundation for liquidity (money); we thus consider two assets, at the cost of a simplification of the cross-sectional distribution.[16] We therefore focus on and isolate a novel trade-off between liquidity-insurance (inequality), and standard stabilization. We find closed-form solutions for the case of exogenous policy and when solving for optimal policy, we summarize the trade-offs through a "loss function"—which allows a transparent illustration of the mechanism at work. The Ramsey problem that we solve is simple and transparent: it implies that imperfect long-run consumption insurance (an aspect of inequality), is enough—and hence essential—to motivate large optimal deviations from price stability stemming from a motive to provide liquidity.

2 A Monetary NK Model with Heterogeneous Agents

We build a simple, tractable, heterogeneous-agent, New Keynesian model with money: heterogenous households hold money to self-insure against idiosyncratic risk, markets are incomplete, participation is limited (infrequent), and price adjustment is costly. The main idea, following contributions reviewed in the introduction and referred to below, is to introduce partial insurance among a subgroup of households to reduce heterogeneity while preserving the self-insurance motive.

Households. There is a mass 1 of households, indexed by $j \in [0,1]$, who discount the future at rate β and derive utility from consumption c_t^j and disutility from labor supply l_t^j. The period utility function is:

$$u\left(c_t^j\right) - \chi \frac{\left(l_t^j\right)^{1+\varphi}}{1+\varphi},$$

with $u(c) = \left(c^{1-\gamma} - 1\right) / (1-\gamma)$. Households have access to three assets: money (with zero nominal return), public debt (with nominal return $i_t > 0$), and shares in monopolistically competitive firms. Money is held despite being a dominated asset because financial frictions give it a consumption-smoothing, insurance role. These frictions are: uninsurable idiosyncratic risks and infrequent participation in financial markets. Such frictions customarily generate a large amount of heterogeneity: the economy is characterized by a continuous distribution of wealth, which is very hard to study with

[15] In both Bhandari et al and Nuno and Thomas, the main channel (absent in our paper) is instead "Fisherian": inflation redistributes from savers to borrowers by reducing the value of debt. This mechanism is absent in our paper. While in Challe (2017), there are no deviations from price stability under optimal policy when the uninsurable idiosyncratic risk is endogenous unemployment (in the absence of equilibrium trade and endogenous liquidity).

[16] Limited participation being a pervasive fact in the US data, the heterogeneity of returns that we consider is likely important for the link between monetary policy and liquidity-insurance.

aggregate shocks and sticky prices.

To simplify the problem (and thus enable us to perform the analysis previewed in the Introduction), we use tools developed in the incomplete-markets literature to reduce the amount of heterogeneity. These simplifications keep the essence of intertemporal trade-offs and of redistributive effects of monetary policy in general equilibrium, and can be viewed as a simple generalization of the Lucas (1990) multiple-member household metaphor. As we shall see, in our economy the key intertemporal trade-offs are captured by households' Euler equations for money and other assets; at the same time, a relevant but limited amount of heterogeneity captures the redistributive effects of inflation and money creation.[17] The gain of this modeling strategy is that one can use standard techniques used in representative-agent (New Keynesian or otherwise) models. In particular, we can solve a version of the model in closed-form by standard local approximation, and compute Ramsey-optimal policy with aggregate shocks.[18]

Households participate infrequently in financial markets. When they do, they can freely adjust their portfolio and receive dividends from firms. When they do not, they can use only money to smooth consumption. Denote by α the probability to keep participating in period $t+1$, conditional upon participating at t (hence, the probability to switch to not participating is $1-\alpha$). Likewise, call ρ the probability to keep non-participating in period $t+1$, conditional upon not participating at t (hence, the probability to become a participant is $1-\rho$). The fraction of *participating* households is $n = (1-\rho)/(2-\alpha-\rho)$, and the fraction $1-n = (1-\alpha)/(2-\alpha-\rho)$ does not participate.

Furthermore, households belong to a family whose head maximizes the intertemporal welfare of family members using a utilitarian welfare criterion (all households are equally weighted), but faces some limits to the amount of risk sharing that it can do. Households can be thought of as being in two states or "islands"[19]. All households who are participating in financial markets are on the same island, called P. All households who are not participating in financial markets are on the same island, called N. The family head can transfer *all* resources across households *within* the island, but cannot transfer *some* resources *between* islands.

Households in the participating island work at real wage w_t. To simplify the exposition, we assume as a benchmark that non-participating households work to get a fixed exogenous income, a home-production amount δ (which is also their fixed labor supply) that is low enough to induce them

[17]See also Curdia and Woodford (2009) and Nistico (2016) for other applications of the "infrequent participation" structure in different contexts with sticky prices.

[18]As will become clear conceptually below, because of the extreme truncation of the state space that we use, the environment can be extended to more states with little increase in computational difficulty—we use Dynare to compute Ramsey-optimal policy.

[19]The use of the family head and island metaphors builds on Challe et al (2017); this is generalized further, in a diferent context, by Le Grand and Ragot (2017). Khan and Thomas (2011) provide a decentralization of the family head assumption with limited participation. We use limited participation in financial markets to introduce a demand for liquidity for self-insurance.

to self-insure. This isolates the channel that we want to emphasize: self-insurance through money in face of uninsurable risk. In this version, a natural interpretation of the idiosyncratic risk is related to unemployment, but our framework is more general and can accommodate several others: broadly speaking, we can think of these shocks as "liquidity shocks", i.e. any shock that makes households want to consume and increase their demand for liquidity. We then relax this assumption and study a version of the model where non-participating households also work at the market wage.

The *timing* is the following. At the beginning of the period, the family head pools resources within the island. The aggregate shocks are revealed and the family head determines the consumption/saving choice for each household in each island. Then households learn their next-period participation status and have to move to the corresponding island accordingly, taking *only money* with them. The key assumption is that the family head cannot make transfers to households *after* the idiosyncratic shock is revealed, and will take this as a constraint for the consumption/saving choice.

The *flows across islands* are as follows. The total measure of households leaving the N island each period is the number of households who participate next period: $(1-n)(1-\rho)$. The measure of households staying on the island is thus $(1-n)\rho$. In addition, a measure $(1-\alpha)n$ leaves the P island for the N island at the end of each period.

Total welfare maximization implies that the family head pools resources at the beginning of the period in a given island and implements symmetric consumption/saving choices for all households in that island. Denote as b_{t+1}^P and M_{t+1}^P the per-capita period t bonds and money balances respectively, in the P island, after the consumption-saving choice. The real money balances are $m_{t+1}^P = M_{t+1}^P/\mathcal{P}_t$, where $\mathcal{P}_t$ is the price level. The end-of-period per capita real values (*after* the consumption/saving choice but *before* agents move across islands) are $\tilde{b}_{t+1}^P$ and $\tilde{m}_{t+1}^P$. Denote as m_t^N the per capita beginning-of-period capital money in the N island (where the only asset is money). The end-of-period values (*before* agents move across islands) are $\tilde{m}_{t+1}^N$. We have the following relations, after simplification (as bonds do not leave the P island, we have $b_{t+1}^P = \tilde{b}_{t+1}^P$):

$$\begin{aligned} m_{t+1}^P &= \alpha \tilde{m}_{t+1}^P + (1-\alpha)\,\tilde{m}_{t+1}^N \\ m_{t+1}^N &= (1-\rho)\,\tilde{m}_{t+1}^P + \rho \tilde{m}_{t+1}^N. \end{aligned} \tag{1}$$

The *program of the family head* is (with $\pi_t = (\mathcal{P}_t - \mathcal{P}_{t-1})/\mathcal{P}_{t-1}$ denoting the net inflation rate):

$$\begin{aligned} W\left(b_t^P, m_t^P, m_t^N, X_t\right) \quad = \quad & \max_{\substack{\{c_t^P, \tilde{b}_{t+1}^P, \\ \tilde{m}_{t+1}^P, \tilde{m}_{t+1}^N, c_t^N, l_t^P\}}} n\left[u\left(c_t^P\right) - \chi \frac{\left(l_t^P\right)^{1+\varphi}}{1+\varphi}\right] + (1-n)\left[u\left(c_t^N\right) - \chi \frac{\delta^{1+\varphi}}{1+\varphi}\right] \\ & + \beta E W\left(b_{t+1}^P, m_{t+1}^P, m_{t+1}^N, X_{t+1}\right) \end{aligned}$$

subject to:

$$c_t^P + \tilde{b}_{t+1}^P + \tilde{m}_{t+1}^P = w_t l_t^P - \tau_t^P \qquad (2)$$
$$+ \frac{1+i_{t-1}}{1+\pi_t} b_t^P + \frac{m_t^P}{1+\pi_t} + \frac{1}{n} d_t,$$
$$\tilde{m}_{t+1}^N + c_t^N = \delta - \tau_t^N + \frac{m_t^N}{1+\pi_t} \qquad (3)$$
$$\tilde{m}_{t+1}^P, \tilde{m}_{t+1}^N \geq 0 \qquad (4)$$

and the laws of motion for money flows relating m_{t+1}^j to $\tilde{m}_{t+1}^j$ (1). Equation (2) is the per capita budget constraint in the P island: P-households (who own all the firms) receive dividends d_t/n, and the real return on money and bond holdings. With these resources they consume and save in money in bonds, and pay taxes/receive transfers τ_t^P (lump-sum taxes include any new money created or destroyed). Equation (3) is the budget constraint in the N island. Finally (4) are positive constraints on money holdings and are akin to credit constraints in the heterogeneous-agent literature. The variable X_t in the value function refers to all relevant period t information necessary to form rational expectations. Using the first-order and envelope conditions, we have:

$$u'\left(c_t^P\right) \geq \beta E \frac{1+i_t}{1+\pi_{t+1}} u'\left(c_{t+1}^P\right) \text{ and } \tilde{b}_{t+1}^P = 0 \qquad (5)$$
$$u'\left(c_t^P\right) \geq \beta E \left[\alpha u'\left(c_{t+1}^P\right) + (1-\alpha) u'\left(c_{t+1}^N\right)\right] \frac{1}{1+\pi_{t+1}} \text{ or } \tilde{m}_{t+1}^P = 0 \qquad (6)$$
$$u'\left(c_t^N\right) \geq \beta E \left[(1-\rho) u'\left(c_{t+1}^P\right) + \rho u'\left(c_{t+1}^N\right)\right] \frac{1}{1+\pi_{t+1}} \text{ or } \tilde{m}_{t+1}^N = 0 \qquad (7)$$
$$w_t u'\left(c_t^P\right) = \chi \left(l_t^P\right)^{\varphi} \qquad (8)$$

The first Euler equation corresponds to the choice of bonds: there is no self-insurance motive, for they cannot be carried to the N island: the equation is the same as with a representative agent.[20]

The money choice of P-island agents is governed by (6), which takes into account that money can be used when moving to the N island. The third equation (7) determines the money choice of agents in the N island, and the last equation labor supply.

The important implication of this market structure is that the Euler equations (6) and (7) have the same form as in a fully-fledged incomplete-markets model of the Bewely-Huggett-Aiyagari type. In particular, the probability $1-\alpha$ measures the uninsurable risk to switch to "low income" (unemployment) next period, risk for which money is the only means to self-insure. This is why money is held in equilibrium for self-insurance purposes, despite being a dominated asset.

[20] An intuition for the underlying market structure is as follows. As agents pool resources when participating (which would be optimal with symmetric agents in time 0 and time 0 trading), they perceive a return conditional on participating next period in financial markets. This exactly compensates for the probability of not participating next period, thus generating the same Euler equation as with a representative agent.

Production and Price Setting. The final good is produced by a firm using intermediate goods as inputs. The final sector production function is $Y_t = \left(\int_0^1 (y_t(z))^{1-\frac{1}{\varepsilon}} dz\right)^{\frac{\varepsilon}{\varepsilon-1}}$, where y_t is the amount of intermediate good z used in production. Denote as $\mathcal{P}_t(z)$ the price of intermediate goods z. Demand for an individual product is $Y_t(z) = (\mathcal{P}_t(z)/\mathcal{P}_t)^{-\varepsilon} Y_t$ with the welfare-based price index $\mathcal{P}_t = \left(\int_0^1 \mathcal{P}_t(z)^{1-\varepsilon} dz\right)^{\frac{1}{1-\varepsilon}}$. Each individual good is produced by a monopolistic competitive firm, indexed by z, using a technology given by: $Y_t(z) = A_t l_t(z)$. Cost minimization, taking the wage as given, implies that the real marginal cost is $W_t/(A_t\mathcal{P}_t)$. The problem of producer z is to maximize the present value of future profits, discounted using the stochastic discount factor of their shareholders, the participants.

When price adjustment is frictionless, prices of all firms are equal to a constant markup over the nominal marginal cost—the real marginal cost is constant $W_t/(A_t\mathcal{P}_t) = (\varepsilon-1)/\varepsilon$. We assume that firms are subject to nominal rigidities as in Rotemberg (1982): to change their prices, firms incur a quadratic adjustment cost that is homogenous across firms. Profits of each firm are thus given by $d_t = \left(1 - \frac{w_t}{A_t} - \frac{\nu}{2}\pi_t^2\right) Y_t$, anticipating that the equilibrium is symmetric. Maximization of their present discounted value gives rise to the nonlinear forward-looking "New Keynesian Phillips curve", whose derivation is described in detail in the Appendix—where we replaced the labor supply schedule $w_t = \chi\left(l_t^P\right)^{\varphi}\left(c_t^P\right)^{\gamma}$:

$$\pi_t(1+\pi_t) = \beta E_t\left[\left(\frac{c_t^P}{c_{t+1}^P}\right)^{\gamma} \frac{Y_{t+1}}{Y_t}\pi_{t+1}(1+\pi_{t+1})\right] + \frac{\varepsilon}{\nu}\left(\frac{\chi\left(l_t^P\right)^{\varphi}\left(c_t^P\right)^{\gamma}}{A_t} + \Phi - 1\right), \tag{9}$$

where $\Phi \equiv 1 - (\varepsilon-1)(1+\sigma)/\varepsilon$ captures the steady-state distortion and σ is a corrective sales subsidy. In particular, when the subsidy is equal to the desired net markup $\sigma = (\varepsilon-1)^{-1}$, there is no steady-state distortion associated with monopolistic competition and elastic labor, $\Phi = 0$. These considerations will be useful when studying the Ramsey policy below.

Money Creation and the Government Budget. To start with, we assume that money is created through "helicopter drops", although we also look at the implications of open-market operations later. Furthermore, we focus on uniform taxation $\tau_t^P = \tau_t^N = \tau_t$.[21]. Denote by x_t the (real value of) new money created in period t, and by M_{t+1}^{tot} the total nominal quantity of money in circulation at the end of each period. In nominal terms, $M_{t+1}^{tot} = M_t^{tot} + \mathcal{P}_t x_t$, and in real terms:

$$m_{t+1}^{tot} = \frac{m_t^{tot}}{1+\pi_t} + x_t \tag{10}$$

[21] We abstract from the possibility of exogenous redistribution by choosing type-specific transfers: $\tau_t^P = \frac{\omega}{n}\tau_t; \tau_t^N = \frac{1-\omega}{1-n}\tau_t$ where ω is the share of total taxes paid by all type-N agents (we focus here on $\omega = n$, entirely lump-sum transfers). But it can be easily shown that there exist (i) processes τ_t^j that (redistribute money so as to) replicate Woodford's cashless limit; and (ii) a value of ω that restores neutrality and Wallace's 1981 logic—i.e. "keeping fiscal policy constant" in the sense of finding an (exogenous) redistribution that un-does the endogenous redistribution triggered by a monetary policy shock in our framework.

Hence, the total period t net taxes/transfers are $\tau_t = -x_t$.

Market clearing and equilibrium. Since there is no public debt, the period t market for bonds is $nb^p_{t+1} = 0$. The money market clears $m^{tot}_{t+1} = (1-n)\,\tilde{m}^N_{t+1} + n\tilde{m}^P_{t+1}$ and so does the labor market $l_t = nl^P_t$. Denoting by c_t total consumption and $Y_t = A_t l_t$ market-produced output (or earned total income), we have that the goods market will also clear, by Walras' Law:

$$c_t \equiv nc^P_t + (1-n)\,c^N_t = \left(1 - \frac{\nu}{2}\pi^2_t\right) Y_t + (1-n)\,\delta. \tag{11}$$

Note for further use that there is a resource cost of changing prices (inflation), which is isomorphic to the welfare cost of relative price dispersion in a Calvo-type model, see e.g. Woodford (2003). In Appendix A we provide the summary of model equations and the equilibrium definition.

Steady state. The analysis of the model's steady state (defined as an allocation where real variables are constant and nominal variables grow at a constant rate π) provides a series of first insights into its monetary structure. The Euler equation for bonds implies that their real return is always equal to the inverse of the discount factor:

$$\frac{1+i}{1+\pi} = \beta^{-1}.$$

Defining $q_t \equiv c^P_t / c^N_t$, as consumption *inequality (imperfect insurance),* the self-insurance Euler equation delivers:

$$q \equiv \frac{c^P}{c^N} = \left(\frac{\frac{1+\pi}{\beta} - \alpha}{1-\alpha}\right)^{\frac{1}{\gamma}} > 1.$$

Letting the steady-state share of exogenous income of N in average consumption be $\delta_c \equiv \delta/c$ (recall this is home production, or unemployment benefits when interpreting idiosyncratic risk as unemployment risk), and the share of N households' consumption in total be h (with the share of P's consumption in total similarly denoted by p):

$$h \equiv \frac{c^N}{c} = \frac{1}{1 + n\,(q-1)},$$

we find (as long as it is positive) the steady-state money demand share, or inverse consumption velocity of money:[22]

$$\mu \equiv \frac{m^{tot}}{c} = \frac{(h - \delta_c)\,(1+\pi)}{2 - \rho - \alpha + \pi}.$$

Subject to a caveat of existence of a monetary equilibrium, discussed in detail in Appendix B, steady-state money demand is equal to the share of (non-home-produced) consumption of N (adjusted for inflation), divided by a parameter capturing the degree of overall churning, the sum of the transition probabilities from one state to another. Under the restriction $\alpha + \rho > 1$ (which we return to below), this parameter is between 0 and 1. For a given level of home production, this

[22] Appendix B.7 provides the expression for the model variant where N are employed at the market wage.

expression implicitly defines upper bounds on the degree of market incompleteness (as described by α and ρ) so that steady-state money demand is positive.[23] Conversely, for given α and ρ there exists a threshold δ beyond which P choose not to hold money: the outside option is too good and there is no need to self-insure. Thus, δ_c captures the degree of insurance provided by (un-modelled) fiscal transfers: were it high enough, no liquidity would be traded $\mu = 0$ and there would be no role for monetary policy in this model beyond its standard role in cashless models. We will focus on the case with equilibrium liquidity and inequality, $\delta_c < h < 1$.

2.1 Simple Monetary NK Model with Heterogeneous Agents

It is instructive to pause and compare the household side of our model with that of the seminal HANK papers reviewed in the introduction. This helps understand how ours is a simplified version of that framework—what mechanisms it still captures, and what it leaves out in order to gain tractability. Take first our concept of *liquidity*, which differs from KVM's, where bonds are liquid and equity and housing illiquid. In assuming that bonds and equity are illiquid while money is liquid, we follow the definition of the monetary theory that we reviewed.[24] Second, our constrained unit-MPC households *are wealthy hand-to-mouth*, similarly to KMV's—their wealth is located just on the P island, where they have a positive probability of going (back). Third, unlike in KMV, our constrained households in the baseline have *exogenous income*. An earlier literature already clarified the amplifying, Keynesian effect on monetary transmission of hand-to-mouth households who have endogenous income because they are employed (see Bilbiie, 2004; 2008 and the discussion in the Introduction). We first abstract from that well-understood general equilibrium channel to isolate and better understand another, which we emphasize below—endogenous movements in liquidity; this is also consistent with an interpretation of the uninsurable risk being related to unemployment, as in MNS.[25] We then introduce this channel by studying a version of our model where constrained households are employed and have endogenous income. Lastly, the assumptions we used to *reduce heterogeneity and history-dependence* have a close counterpart in the sticky-price literature that is probably clear to readers well-seasoned in NK models: our participation/insurance scheme is conceptually similar to the Calvo model of price stickiness. Whereas KMV's portfolio decision based on a quadratic transaction cost for illiquid assets generate endogenous participation in liquidity; since

[23] The formal restriction is, for the case of zero steady-state inflation and treating n as a parameter: $\alpha < 1 - n\frac{\beta^{-1}-1}{\delta_c^{-1}-1} < \beta^{-1}$. In terms of the original parameter ρ we have $\frac{1-\alpha}{1-\rho} > \sqrt{\frac{1}{4} + \frac{(1-\beta)\delta_c}{(1-\rho)\beta(1-\delta_c)}} - \frac{1}{2}$.

[24] Tongue in cheek, one may label this a MONK model, as in "monetarist New Keynesian". See Weill (2007), Rocheteau and Weill (2011), Kiyotaki and Moore (2012), and Cui and Sadde (2016) for recent sophisticatd refinements (as well as reviews) of the concept of liquidity in recent monetary theory, including a different view of liquidity based on asset resalability (while our is on limited participation).

[25] Unemployment risk is exogenous in MNS, but endogenous through search and matching in other HANK models reviewed in the introduction.

state variables enter this decision (generating complex distributional dynamics that our simplification abstracts from) this is closer to state-dependent models of price stickiness.

Thus, our model *cannot* fit the detailed distribution of asset holdings and wealth, nor reproduce movements in portfolio shares or realistic idiosyncratic income processes—it does not capture the rich household wealth dynamics of fully-fledged Bewley-Aiyagari-Huggett models; in particular, it does not capture tails of the distribution—households who have a long stream of good (or bad) luck. But it does captures market incompleteness by one parameter through which, as we shall see, "history matters"—even though for just one period. The simplifications "buy" us the ability to compute optimal policy transparently.[26]

3 Inspecting HANK Transmission: Liquidity and Aggregate Demand

In this section, we use our simple and tractable model to shed light on some of its properties that are key for understanding optimal monetary policy. First, we assume that liquidity provision is exogenous—the central bank follows a money supply (growth) rule—and study the effect and transmission of a liquidity injection. Then, we study whether (endogenous) liquidity provision can be used to provide insurance, i.e. neutralize the effect of aggregate shocks on inequality—and if so, with what inflationary consequences? To explore these questions, we use a local approximation of the model around a steady state with zero inflation $\pi = 0$ for ease of illustration (a summary of all loglinearized equilibrium conditions around an arbitrary inflation rate is in Appendix B). Denote log-deviations of any variable by a hat, unless specified otherwise.

The Euler equation of participants and the self-insurance equation are given by, respectively:

$$\begin{aligned} \hat{c}_t^P &= E_t \hat{c}_{t+1}^P - \gamma^{-1} \left(i_t - E_t \hat{\pi}_{t+1} \right), & (12) \\ \hat{c}_t^P &= \alpha\beta E_t \hat{c}_{t+1}^P + (1-\alpha\beta) E_t \hat{c}_{t+1}^N + \gamma^{-1} E_t \hat{\pi}_{t+1}. & (13) \end{aligned}$$

Let $\hat{x}_t \equiv (x_t - x)/m^{tot}$ be the deviation of new money issued by the central bank today, as a fraction of steady-state total money. The equation governing money growth is hence:

$$\hat{x}_t = \hat{m}_{t+1}^{tot} - \hat{m}_t^{tot} + \hat{\pi}_t \quad (14)$$

[26] In a separate paper, we concentrate on the positive implications: we loglinearize this model, solve it in closed-form, and analyze the monetary transmission mechanism. We show that the Taylor principle fails dramatically in this economy: the central bank cannot stick to a Taylor rule that is otherwise reasonable in the representative-agent model. Augmenting the rule with inequality or a measure of liquidity restores determinacy, while a money growth rule is even better.

The linearized budget constraint of non-participants is:

$$\hat{c}_t^N = \frac{1-\alpha}{1-n}\frac{\mu}{h}\left(\hat{m}_t^{tot} - \hat{\pi}_t\right) + \frac{\mu}{h}\hat{x}_t. \tag{15}$$

New money $\hat{x}_t$ reaches N agents within the period (because money is issued through helicopter drops), and the Pigou effect reduces the value of their outstanding real balances. Finally, the price-setting equation is the loglinearized version of (9):[27]

$$\hat{\pi}_t = \beta E_t \hat{\pi}_{t+1} + \psi\left[(\bar{\varphi} np + \gamma)\,\hat{c}_t^P + \bar{\varphi}\,(1-n)\,h\hat{c}_t^N - (1+\varphi)\,\hat{a}_t\right]. \tag{16}$$

with $\psi \equiv \frac{\varepsilon-1}{\nu}$ ranging from 0 (fixed prices) to ∞ (flexible prices) and $\bar{\varphi} = \varphi/\left(1-(1-n)\,\delta_c\right)$.[28]

A local rational expectations equilibrium consists of a vector of processes $\hat{c}_t^N, \hat{c}_t^P, \hat{\imath}_t, \hat{\pi}_t, \hat{x}_t,\ \hat{m}_{t+1}^{tot}$ that satisfy the equations (12) to (16). The reduced-form model, while capturing key elements of modern HANK models, is thus reminiscent of both "old" monetarist and Keynesian "dynamic ISLM" models—such as Sargent and Wallace (1975), or Sargent (1987). To close the model, we need to specify how monetary policy is conducted. In this section, we sketch the effects of monetary injections and analyze their transmission under the assumption that the central bank chooses money growth (under which, as in Sargent and Wallace, the equilibrium is determinate). Subsequently, we solve for the optimal policy of the central bank: liquidity will then be determined endogenously.[29]

3.1 Liquidity-Insurance, Aggregate Demand, and Inflation

Combining (13) and (12), we obtain a core equation of our model, which captures the link between interest rates (the price of liquidity) and lack of consumption insurance, or "inequality" defined as $\hat{q}_t \equiv \hat{c}_t^P - \hat{c}_t^N$:

$$E_t\hat{q}_{t+1} = E_t\hat{c}_{t+1}^P - E_t\hat{c}_{t+1}^N = \frac{\gamma^{-1}}{1-\alpha\beta}\hat{\imath}_t \tag{17}$$

This illustrates the insurance role of monetary policy in our model: as the opportunity cost of holding liquidity (i) falls, P hold more of it, leading to higher consumption for N (and lower for

[27] We used that aggregate labor supply is proportional to participants' labor supply $\hat{l}_t = \hat{l}_t^P$, the linearized labor supply equation $\varphi\hat{l}_t = \hat{w}_t - \gamma\hat{c}_t^P$, the economy resource constraint $\hat{c}_t = (1-(1-n)\,\delta_c)\left(\hat{l}_t + \hat{a}_t\right)$with aggregate consumption (denoting $p \equiv c^P/c = qh$): $\hat{c}_t = np\hat{c}_t^P + (1-n)\,h\hat{c}_t^N$. Finally, $\bar{\varphi} = \varphi/\left(1-(1-n)\,\delta_c\right)$.

[28] Both agents consume the output, so movements along the labor supply curve concern them both: thus, the real wage depends on *total* consumption with elasticity $\bar{\varphi}$. However, since only participants work, they are the only ones subject to the income effect: thus, the real wage depends only on consumption of participants with elasticity equal to the income effect γ. This generates an asymmetry in the inflationary effects of consumption of the two agents.

[29] In a separate paper, we analyze the other case: monetary policy via Taylor-type interest rate rules, and ask: how does a central bank ensure equilibrium determinacy in an incomplete-markets economy where liquidity (money creation) $\hat{x}_t$ is *endogenous*. It turns out that endogenous fluctuations in precautionary liquidity seriously challenge the central bank's control of aggregate demand and question the appropriateness of Taylor rules. For moderate market incompleteness, the Taylor coefficients required for determinacy are in the *double digits*—but responding to inequality or liquidity can restore conventional wisdom in the form of the "Taylor principle".

P) agents tomorrow. Hence, more liquidity (lower interest rates) leads to more insurance (lower future inequality). This effect is stronger, the more intertemporal substitution there is (higher γ^{-1}) and the higher is α. This equilibrium outcome of our model is consistent with the empirical findings documenting a positive correlation between expansionary, inflationary monetary policy and redistribution; see for example Doepke and Schneider (2006), Adam and Zhu (2014), and Coibion et al (2013).

Monetary-NK IS curve. Aggregate demand in our economy is made of the demand of the two types, participants and nonparticipants. The demand of participants is determined by an Euler equation, but in contrast to the standard RA model (and to models with hand-to-mouth agents) this Euler equation includes an insurance/precautionary saving motive (13). That equation thus links the two components of aggregate demand: participants' and non-participants'. The latter is determined by the previous accumulation of money balances, and by the money transfer received, as in (15). Inflation has an impact on both households' demand: *realized* inflation reduces the real value of money balances (and hence, the income and consumption) of N, while *expected future* inflation influences the insurance decision of P.

Combining (15), and (14), we obtain an equation linking the aggregate demand of N to money transfers and inflation:

$$\begin{aligned}(1-n)\,h\hat{c}_t^N &= \mu(\alpha-n)\,\hat{x}_t+\mu(1-\alpha)\,\hat{m}_{t+1}^{tot}\\ &= \mu(1-n)\,\hat{m}_{t+1}^{tot}-\mu(\alpha-n)\left(\hat{m}_t^{tot}-\hat{\pi}_t\right)\end{aligned} \tag{18}$$

The key measure of market incompleteness in our model is:

$$\alpha-n=(1-n)(\alpha+\rho-1)>0,$$

which captures the direct effect on non-participants' demand of an increase in liquidity x_t . Indeed, $\alpha-n$ captures the idea that the *conditional* probability of *remaining* P is higher than the *unconditional* probability of *becoming* P, i.e. the share of P in the total population. The parameter thus measures the incumbents' advantage, the "memory" of the process, or the trials' not being independent: $\alpha>1-\rho$ implies that it is more likely for a P household to stay P than it is for an N household to become P (with the labor-risk interpretation, it implies that is is more likely for an employed agent to keep their job than for an unemployed agent to become employed, which is a natural restriction).

In equilibrium, $\alpha-n$ is hence the elasticity (integrated across all) of N agents' consumption to a monetary transfer (for given future real money balances): for while an $1-n$ fraction is consumed by N, a $1-\alpha$ fraction is saved for self-insurance purposes by P. The same parameter captures *also* the elasticity of N's aggregate demand to inflation, for given real money balances—that is, the Pigou effect discussed previously.

The aggregate IS curve of our economy is obtained by using (18) twice (evaluated at t and $t+1$), together with the Euler equation of participants (12), the expression of aggregate consumption $\hat{c}_t = np\hat{c}_t^P + (1-n)\, h\hat{c}_t^N$, and money growth (14):

$$\begin{aligned}\hat{c}_t &= E_t\hat{c}_{t+1} - np\gamma^{-1}\left(\hat{\imath}_t - E_t\hat{\pi}_{t+1}\right) + \mu\,(1-\alpha)\, E_t\hat{\pi}_{t+1} \\ &\quad + \mu\,(\alpha - n)\,\hat{x}_t - \mu\,(1-n)\, E_t\hat{x}_{t+1}\end{aligned} \tag{19}$$

Through what we could call the *monetary-New Keynesian IS curve* (19), aggregate demand depends on money (liquidity), interest, and prices (inflation)—hat tip to Patinkin (1956). There are three main differences with respect to the aggregate IS curve of a standard representative-agent economy, corresponding to these three components.

First, *money* (liquidity) creation affects aggregate demand directly, through its impact on aggregate demand of N agents discussed in detail above. This effect is proportional to $\alpha - n > 0$, which captures market incompleteness in our model as explained above: while a fraction $1-n$ of the liquidity injection gets consumed by the hand-to-mouth, constrained N, a fraction $1-\alpha$ is held as insurance by the precautionary P—which gives the net effect of $1-n-(1-\alpha) = \alpha - n$.[30]

Second, expected *inflation* matters for aggregate demand over and above its effect through the ex-ante real interest rate (our next point). Higher expected inflation creates more demand today at given real interest rates (through $\mu\,(1-\alpha)\, E_t\hat{\pi}_{t+1}$) by intertemporal substitution, because it diminishes the real value of liquidity tomorrow. This expected inflation channel is "as if" N were at the zero lower bound permanently.

Lastly, the *interest*-elasticity of aggregate demand is lower than in a representative-agent economy: $np\gamma^{-1} < \gamma^{-1}$ and *decreasing* with the share of constrained households (as in MNS). This is the opposite with respect to a model in which nonparticipants have *endogenous labor income (*for instance, *employed* at the market wage*).*[31] In that model (that we also study below), the interest elasticity of aggregate demand is *increasing* with the share of hand-to-mouth nonparticipants: in response to a cut in interest rates, demand expands, labor demand shifts, and the wage increases; the income of the constrained increases, leading to a further amplification on demand. We first abstract from this to focus in isolation on the role of liquidity (money) for self-insurance against

[30] A particular exogenous redistribution through transwers $n\tau^P = \alpha\tau$ is one instance of a Wallace (1981)-type "constant fiscal policy" that will undo the effect of monetary injections.

[31] See Bilbiie (2008) for a full analysis of a cashless model with employed nonparticipants (including the case where at high values of the share of non-participants, the interest elasticity of aggregate demand changes sign). See Gali, Lopez-Salido, and Valles (2004, 2007) for related models with hand-to-mouth agents focusing on different issues. Eggertsson and Krugman (2012) use a similar aggregate demand structure to analyze deleveraging and liquidity traps. See Bilbiie (2017) for further discussion of the difference between the two aggregate demand models and their different implications for forward guidance.

idiosyncratic risk, by assuming that all nonparticipants have exogenous income. We then relax this assumption below to also study the role of the "endogenous income", New Keynesian cross channel. The aggregate IS curve for that model (see Appendix B.7) shows that the same (NK cross) amplification mechanism discussed in Bilbiie (2008, 2017) applies here—both with respect to interest rate changes but also, something novel here, with respect to the other aggregate demand determinants: liquidity, and expected inflation. Yet because bonds are illiquid here, in contrast to MNS and other contributions discussed in the Appendix, there is no "discounting" in the aggregate Euler equation, and no interaction between amplification and discounting.

Reduced form, 3-equation model. Rewriting the self-insurance equation using the definition of inequality and replacing the budget constraint of N, we obtain an equation that links inequality (imperfect insurance) to present and future liquidity, and expected inflation:

$$q_t = \alpha\beta E_t q_{t+1} - \frac{\alpha - n}{1-n}\frac{\mu}{h}\hat{x}_t + \frac{\mu}{h}E_t\hat{x}_{t+1} + \left(\gamma^{-1} - \frac{1-\alpha}{1-n}\frac{\mu}{h}\right)E_t\hat{\pi}_{t+1}, \tag{20}$$

Expected inflation has two opposing effects on present inequality, keeping future inequality (and hence the nominal interest rate) fixed. On the one hand, it tells P to consume more today, for money will have a lower payoff tomorrow—this is driven by intertemporal substitution. On the other hand, the Pigou effect on N tomorrow tells P (who might become N tomorrow) to save more for precautionary reasons, i.e. hold more liquidity and consume less—an income effect that gives more insurance today. With log utility the elasticity to inflation is positive and less than unity, namely $0 < \frac{\delta_c}{h} < 1$ (as required by positive steady-state money demand).

Since under a money growth rule liquidity is exogenous, we can solve for the entire path of inequality and inflation using equations (20) and (16), appropriately rewritten as:

$$\hat{\pi}_t = \beta E_t\hat{\pi}_{t+1} + \psi\left[\left(\bar{\varphi}np + \gamma\right)q_t + \left(\bar{\varphi} + \gamma\right)\left(\frac{1-\alpha}{1-n}\frac{\mu}{h}\left(\hat{m}_t^{tot} - \hat{\pi}_t\right) + \frac{\mu}{h}\hat{x}_t\right) - \left(1+\varphi\right)\hat{a}_t\right],$$

where $\hat{a}_t$ is the log-deviation of the technology level A_t. The money equation (14) then determines the path of real money balances, while the nominal interest rate is proportional to expected future inequality as explained above in (17). Note that there is a liquidity effect if expected inequality falls when issuing money.

Solving the model in closed form is not possible in the general case under a money growth rule (14)—a property shared with even the simplest textbook NK model with money, e.g. Galí (2008). To obtain closed-form solutions that help our understanding of the model, we consider two instructive special cases.

3.2 Closed-form solution with horizontal aggregate supply (fixed prices)

Consider the extreme case of **fixed prices** $\psi = 0$.[32] Under this assumption, inflation does not move $\hat{\pi}_t^f = 0$, so one equation (20) is enough to determine equilibrium, which is locally unique because $\alpha\beta < 1$; solving it forward under the assumption that money growth is AR(1), $E_t \hat{x}_{t+1} = \rho_x \hat{x}_t$, we obtain:

$$q_t^f = -\frac{\frac{\alpha - n}{1-n} - \rho_x}{1 - \alpha\beta\rho_x} \frac{\mu}{h} \hat{x}_t \tag{21}$$

The response depends on our key parameter, $\alpha - n$: the larger it is, the larger the effect of liquidity on aggregate demand (through demand of the constrained), and the larger the ensuing fall in inequality. If the shock is "too" persistent, inequality can *increase* as agents correctly anticipate the future windfall and self-insure less. The path of nominal interest rates is immediately determined through (17): in particular, there is a *liquidity effect* (interest rates fall) if and only if (i). the increase in liquidity is persistent but (ii). not too persistent (so that inequality goes down): $0 < \rho_x < \frac{\alpha - n}{1-n}$.[33]

3.3 Perfect insurance and inflation

In order to help our intuition for the full optimal policy considered next, it is instructive to consider the mechanics of *endogenous liquidity*, given an allocation. In particular, we consider the endogenous path of liquidity $\hat{x}_t$ necessary to implement two specific allocations when the economy is hit by aggregate shocks a_t. The first allocation we consider is the *perfect-insurance* benchmark—as we show formally below, this is the first-best limit of our economy. We compare this with a policy of perfectly stabilizing inflation. Subsequently, we conduct a rigorous Ramsey-optimal policy exercise—but the purpose here is to elucidate the mechanism at work using simple closed-form expressions allowed by our model.

Consider first the policy implementing perfect insurance ($q_t = 0$) under *flexible prices* and starting from a steady state with $q = 1$ ($h = p = 1$).[34] Assuming further log utility, the solution for inequality in the simplest iid case is, defining $\Omega \equiv \frac{1+\bar{\varphi}}{1+\varphi}(1 - \delta_c) > 0$:

$$q_t = -\frac{\alpha - n}{1 - \alpha}\Omega \hat{x}_t + \delta_c \hat{a}_t.$$

Denoting with a double star the economy with no inequality variations $q_t^{**} = 0$, the level of endogenous

[32] A closed-form solution can also be obtained in the other polar case of flexible prices, $\psi \to \infty$, but without obtaining much additional intuition for our purposes in the case of exogenous liquidity.

[33] The result that sticky-price models deliver a liquidity effect is emphasized by Christiano, Eichenbaum, and Evans (2005). The same authors compared sticky-price and limited-participation models' ability to deliver a liquidity effect in previous work—see Christiano (1991), Christiano and Eichenbaum (1992, 1995); see also Fuerst (1992).

[34] Notice that the welfare objective is *not* merely perfect insurance in deviations; as our objective function derived below shows clearly, there is a rationale for increasing c^N in *levels*.

liquidity injection that achieves this is:

$$\hat{x}_t^{**} = \frac{1-\alpha}{\alpha - n}\frac{\delta_c}{\Omega}\hat{a}_t,$$

which is always positive because $1 > \delta_c > 0$. It is decreasing with $\alpha - n$ because, as shown above, $\alpha - n$ captures the elasticity of aggregate demand (and of insurance) to liquidity; the higher this elasticity, the lower the necessary liquidity injection.

One key implication of this policy is that inflation varies, namely:

$$\begin{aligned}
\frac{d\hat{\pi}_t^{**}}{da_t} &= \frac{1}{\Omega}\left(\frac{1-n}{\alpha - n}\delta_c - 1\right) \\
\frac{dE_t\hat{\pi}_{t+1}^{**}}{d\hat{a}_t} &= \frac{1+\varphi}{1+\bar{\varphi}}
\end{aligned}$$

In particular, expected inflation gives agents the right intertemporal incentives to hold the extra liquidity for self-insurance purposes. Thus, the equilibrium is one with *inflation volatility*: the consumption of N is increased by the liquidity injection but *decreased* by current inflation through the Pigou effect.

Consider now the other extreme, of **strict inflation targeting**: a policy (denoted by superscript 0) of stabilizing inflation around the perfect-insurance steady-state (note that the degree of price stickiness plays no role as inflation is constant). Imposing $\frac{d\hat{\pi}_t^0}{da_t} = 0$ at all times, we obtain that inequality is inversely directly related to liquidity, $q_t^0 = -\frac{\mu}{h}\hat{x}_t^0$, and liquidity and the consumption of N in this equilibrium are:

$$\frac{d\hat{c}_t^{N0}}{da_t} = \mu\frac{d\hat{x}_t^0}{\hat{a}_t} = \frac{1+\varphi}{\bar{\varphi}(1-n)}.$$

It then follows immediately by direct inspection that $d\hat{c}_t^{N0} > d\hat{c}_t^{N**}$: consumption of N responds *more*, and is thus more volatile under the zero-inflation policy. In other words, inflation is a means to insure constrained agents against aggregate risk: even though, in levels, inflation decreases consumption of N, inflation volatility reduces N's volatility of consumption. We will see that this general insight holds more broadly when we analyze the policy trade-offs rigorously by means of a Ramsey-optimal policy analysis.

4 Optimal Monetary Policy: Insurance, Liquidity, and Inflation

To understand the role of inflation for redistribution and providing insurance, it is useful to start by looking at the first-best benchmark (the planner solution) and compare it to our economy with flexible prices. The first best allocation is obtained when the planner chooses quantities to maximize ex-ante aggregate welfare:

$$\max_{c_t^P c_t^N, l_t^P} E_0 \sum_{t=0}^{\infty} \beta^t \left(n \left[u\left(c_t^P\right) - \chi \frac{\left(l_t^P\right)^{1+\varphi}}{1+\varphi} \right] + (1-n) \left[u\left(c_t^N\right) - \chi \frac{\delta^{1+\varphi}}{1+\varphi} \right] \right) \tag{22}$$

subject to the resource constraint $nc_t^P + (1-n)\, c_t^N = nA_t l_t^P + (1-n)\,\delta$. Effectively, there is no intertemporal problem: the first-best equilibrium is one with **perfect insurance,** given by the two conditions $c_t^P = c_t^N = c_t$ and $c_t = \chi^{-1/\gamma} \left(l_t^P\right)^{-\varphi/\gamma} A_t^{1/\gamma}$. Consider now the **market economy with flexible prices**, $\nu = 0$. The first-best allocation can be implemented at the Friedman rule, i.e. when the nominal interest rate is $i = 0$, and the inflation rate $\pi = \beta - 1$ (because the real interest rate is β): the return on money is equal to the real interest rate and there is no opportunity cost to self-insure. But our framework exhibits a difficulty which is well known in this class of monetary models starting from Bewley (1983): that, at the Friedman rule, monetary variables are indeterminate—whatever the nominal quantity of money, the price of the final good is indeterminate when the real quantity of money allows households to self-insure;[35] furthermore, there exist examples of Bewley economies in which the Friedman rule is not optimal because of a redistribution effect (when interest payments are paid through lump-sum taxes), see Mehrling (1995).[36]

In a general monetary equilibrium with sticky prices ($\nu > 0$), a novel trade-off occurs: inflation fluctuations (generated by liquidity movements) allow households to self-insure, but generate price adjustment costs. A zero-inflation policy minimizes price adjustment costs, but decrease the ability of households to self-insure. Optimal monetary policy needs to find the balance between these two distortions: *inequality*, or a scope for providing *liquidity-insurance* (specific to an incomplete-markets, limited-participation setup like ours), and *costly price adjustment*: the standard distortion that operates in a representative-agent NK model. This section analyzes how this trade-off is resolved in our model. We solve the full Ramsey-optimal policy, provide a second-order approximation à la Woodford (2003) that is useful to understand the policy trade-offs, and analyze optimal policy quantitatively. The general theme is that inequality (understood as imperfect insurance) triggers a liquidity-insurance motive that implies large optimal deviations from price stability in response to shocks that, absent liquidity, are innocuous.

[35] The indeterminacy property is not specific to our formalization of money demand. It is also found in simple cash-in-advance or money-in-the-utility function models, when no satiation point is introduced in the utility function. See for exemple Correia and Teles (1999) for a discussion. In our model version, this should not be taken as too serious a critique of the Friedman Rule, for it can be shown that when the policy rule converges to it $i \longrightarrow 0^+$ the allocation is well-defined and converges smoothly to the first-best allocation—see Appendix.

[36] Mehrling shows that if taxes are independent of wealth or income, but interest is proportional to money holdings, higher taxes redistribute away from households with little money balances; households eventually get this back once their money holdings increased, but are made worse off because this is the opposite of insurance: it redistributes from high-marginal-utility to low-marginal-utility periods.

4.1 Ramsey-Optimal Policy

Following a long tradition started by Lucas and Stokey (1983), we assume that the central bank acts as a Ramsey planner who maximizes aggregate welfare. In our economy, this entails calculating the welfare of the two agents and weighting them by their shares in the population. The constraints of the planner are the rearranged private equilibrium conditions: self-insurance (6), the Phillips curve (9), and the economy resource constraint (11).[37] We denote the system of these three constraints by $\Gamma_t\left(c_t^P, c_t^N, l_t^P, \pi_t\right)$. As it is by now well understood, the optimal policy problem of the central bank can be written as choosing the allocation $\{c_t^P, c_t^N, l_t^P, \pi_t\}$ to maximize the following Lagrangian:

$$\max_{\{c_t^P, c_t^N, l_t^P, \pi_t\}} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ n \left[u\left(c_t^P\right) - \chi \frac{\left(l_t^P\right)^{1+\varphi}}{1+\varphi} \right] + (1-n) \left[u\left(c_t^N\right) - \chi \frac{\delta^{1+\varphi}}{1+\varphi} \right] + \omega_t \Gamma_t \right\} \tag{23}$$

where ω_t is the vector of three costate, Lagrange multipliers, one for each constraint in Γ_t. The first-order conditions of this problem are outlined in Appendix B.4, as is the proof of the following Proposition.

Proposition 1 *The optimal long-run inflation rate is such that* $\beta - 1 \leq \pi^* \leq 0$.

As in other NK models incorporating different theories of money demand (e.g. Khan et al, 2003; Schmitt-Grohe and Uribe, 2004, 2007), the long-run inflation rate ranges from the Friedman rule under flexible prices and optimal subsidy, to zero inflation under sticky prices and inelastic labor. A low inflation rate allows households to self-insure, but generates price adjustment costs. An inflation rate close to 0 minimizes price adjustment costs, but decreases the ability of households to self-insure, as the return on money decreases.

But unlike in other NK models, including those incorporating money demand, in our economy the central bank *also* uses inflation optimally over the cycle, as a by-product of using liquidity to provide insurance and decrease inequality. We first illustrate formally the trade-off faced by a central bank by deriving a second-order approximation to the aggregate utility function, which contains a *liquidity-insurance* motive, and we then explore the quantitative significance of this novel trade-off.

4.2 A second-order approximation to welfare

To understand the relevant policy trade-offs, we derive a second-order approximation à la Woodford (2003, Ch. 6) to the aggregate welfare function, around a steady-state with imperfect insurance

[37] Since the nominal interest rate enters only the Euler equation for bonds, the problem can be regarded as one where the planner chooses the allocation directly; once the consumption of participants and inflation are known, the optimal interest rate is determined by the Euler equation. By similar reasoning, once the consumption of non-participants is determined, along with inflation, the quantity of real money balances is fully determined too. These simplifications apply only when money is issued via helicopter drops; see Appendix B.4 for the case of optimal policy under open-market operations.

($p > 1 > h$), an optimal subsidy inducing marginal-cost pricing in steady state ($\Phi = 0$ in (9)), and arbitrary steady-state inflation.[38] Furthermore, we assume for simplicity of exposition but with no loss of substance that utility is logarithmic in consumption—Appendix B.6 presents the more general CRRA case and the proof of the following Proposition.

Proposition 2 *Solving the welfare maximization problem is equivalent to solving:*

$$\min_{\{c_t^P, c_t^N, q_t, \pi_t\}} \frac{1}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \Lambda_\pi \tilde{\pi}_t^2 + \Lambda_c \hat{c}_t^2 - \Lambda \hat{c}_t^N \right\},$$

where the optimal relative weights are:

$$\Lambda_\pi = \frac{\nu}{1 - \frac{\nu}{2}\pi^2}; \; \Lambda_c = \frac{1+\varphi}{1 - (1-n)\,\delta_c};$$
$$\Lambda = 2\,(1-n)\,h\,(q-1)\,.$$

The Proposition transparently illustrates the novel "liquidity-insurance" motive implied by our framework: a trade-off between insurance (inequality) and stabilization of inflation and aggregate demand.[39] The last (linear!) term pertains to lack of insurance: it shows the first-order welfare benefit of increasing N's welfare by increasing their consumption level. Intuitively, the weight Λ is proportional to the steady-state distortion captured by $(q-1)$—the first-order benefit exists only insofar as the steady state is distorted to start with. This is analogous to the linear benefit of increasing output above the natural rate when the steady-state is first-order distorted in the standard New Keynesian model (see the next footnote). The distortion vanishes when the steady-state is egalitarian ($p = h$, perhaps through a steady-state insurance scheme, if enough fiscal lump-sum instruments are available to undertake such policy) or, trivially, when $n = 1$ (the standard cashless representative-agent NK model). Replacing the equilibrium q reveals that Λ is proportional to $\beta^{-1} - 1 + \pi$: the distortion also becomes arbitrarily small when the steady state tends toward the Friedman rule, $\pi \to 1 - \beta^{-1}$. Our result thus isolates the role of this monetary distortion, assuming a *second-best* world whereby *fiscal policy does not achieve perfect insurance*. Recall, however, that fiscal policy is already doing much redistribution along two dimensions here: it subsidizes firms to induce marginal cost pricing, and it taxes participants only to finance that subsidy—thus redistributing de facto profit income. The only long-run difference left between the two types is proportional to the return on nominal assets—an inherently monetary distortion, in our view.

[38] Bilbiie (2008) also derives a quadratic loss function in a cashless economy with hand-to-mouth agents. Curdia and Woodford (2009) and Nistico (2015) also do this for cashless models with infrequent access to credit markets; unlike us, they focus on an efficient equilibrium with insurance when calculating optimal policy.

[39] Note that $\tilde{\pi}_t = \pi_t + \pi$ is the inflation *level*, so the function is written so that target inflation is zero absent aggregate shocks, $\tilde{\pi}_t = 0$. The optimal target is the optimal long-run inflation found in the Ramsey problem above, the equivalent of which is here the steady state of the solution of the relevant linear-quadratic problem.

Because of the linear term in the loss function, second-order terms in the private constraints matter for welfare[40]: as long as there is steady-state inequality, inflation and aggregate demand volatility matter for welfare beyond their direct effects through Λ_π and Λ_c. The reason is by now intuitively clear: when the steady state has $q > 1$, increasing the consumption of non-participants provides a first-order welfare benefit; the only way to achieve this benefit, absent fiscal instruments, is monetary. As we show next, a quantitative analysis of optimal policy in a calibrated version of our model suggests that inflation volatility is desirable in this framework. Pursuing price stability instead, even around an optimally chosen inflation target, has large welfare costs.

As will become clear below, the optimal policy prescription is not to simply increase N's consumption; indeed, as we shall see other policies that imply higher consumption levels for N are suboptimal because they imply too much volatility—like in our analytical example above.

A second clarification is that inflation, because of the Pigou effect, is "bad" for N —it erodes the value of their outstanding money balances. We shall see that inflation can nevertheless be optimal, despite this direct harmful effect, when it is a side-effect of liquidity used for insurance.

In Appendix B.7 we outline the main ingredients and implications of the model with endogenous income of participants as in Bilbiie (2008, 2017), which delivers an additional "New Keynesian cross" amplification channel. Therein, we show that the loss function in that model contains a linear term in both c^N and l^N, namely $hc_t^N - l_t^N$. Thus, while there is a benefit in that model to increasing consumption of N, there is also a cost insofar as this insurance-expansion is sustained by the same households working more hours. We will see that quantitatively this intuition implies that there will be less incentives to accommodate inflation, and less deviations from price stability in that model.[41]

5 Optimal Liquidity and Inflation: a Quantitative Evaluation

We calibrate the model at quarterly frequency and follow, for common parameters pertaining to preferences and the supply side, the classic papers in optimal policy in NK models, Khan, King, and Wollman (2003) and Schmitt-Grohé and Uribe (2007): the inverse elasticity of labor supply is $\varphi = 0.25$, and $\gamma = 1$. The elasticity of substitution between goods is $\varepsilon = 6$, and we introduce the

[40] This is analogous to the linear benefit of increasing output above the natural rate when the steady-state is first-order distorted in the standard New Keynesian model. See Woodford (2003; Ch. 6), Benigno and Woodford (2005, 2012) and Schmitt-Grohe and Uribe (2007) for an analysis of this when the distortion pertains to monopolistic distortion, i.e. $\Phi > 0$, including explanations of the second-order corrections that are necessary to correctly evaluate welfare.

[41] Other changes are that the weight on consumption volatility becomes $\varphi + \gamma$ and the expression determining h and q (and hence the size of the distortion) is more involved, as fiscal redistribution matters—see Appendix.

steady-state subsidy $\sigma = 1/(\varepsilon - 1)$ to avoid steady-state distortions due to monopolistic competition—thus isolating our novel channel as a motivation for deviations from price stability. Both cited papers use different models of staggered pricing and assume that prices stay unchanged on average for 5 periods; this implies a Phillips curve slope (our ψ) of around 0.05. Given our ε, the price adjustment cost parameter that delivers the same ψ is $\nu = 100$. The discount factor is $\beta = 0.98$, as in other studies with heterogeneous agents (Eggertsson and Krugman, 2012; Curdia and Woodford, 2009); we consider larger values for robustness below. We use the same labor productivity process as Khan et al, with autocorrelation 0.95 and standard deviation 1%.

Three parameters pertain to market incompleteness and money demand: the probabilities to keep participating (α) and non-participating (ρ), and home production when non-participating (or unemployment benefits) δ. Since we perfectly correlated financial market and labor market participation to obtain our tractable model, two calibrations are possible: one that targets financial market participation and money demand, and the other labor market variables. We use the former as a benchmark and report the latter for robustness.

We target three data features in our benchmark calibration. First, the *number of participants* n: in the US economy roughly half of the population participates in financial markets, either directly or indirectly (Bricker et al, 2014), and this is stable over time. We thus take $n = 0.5$, which implies the restriction $\alpha = \rho$. Second, the *velocity of money* (roughly speaking, μ^{-1} in our notation): considering a broad money aggregate, the quarterly velocity ($GDP/M2$) is around 2 over the period 1982—2007 (chosen to avoid the zero lower bound period). Third, *consumption inequality* q between participating and non-participating agents captures the lack of insurance due to market incompleteness. Since agents participate infrequently in financial markets (Vissing-Jorgensen, 2002) and one cannot keep track of their participation status, it is hard to find an exact empirical counterpart to q. We take as a proxy the fall of nondurable consumption when becoming unemployed, which is estimated between 10% and 20% (see e.g. Chodorow-Reich and Karabarbounis, 2014) and target the conservative value of 10% for this object ($q^{-1} - 1$ in our model). These three targets jointly imply $\alpha = \rho = 0.9$, and $\delta = 0.783$. Table 1 presents our parameters and the implied Ramsey steady-state values for our target variables—which are determined by the exact Ramsey equilibrium conditions outlined in the Appendix.

5.1 Optimal long-run deviations from price stability

The optimal asymptotic (steady-state) inflation rate is $\pi = -0.79\%$. As expected, this is higher than the inflation implied by the Friedman Rule (which is -2%), because prices are sticky, just as in standard monetary models with sticky prices, e.g. Khan et al (2003).[42] More equilibrium deflation

[42] Brunnemeier and Sannikov (2016) provide an example of a flexible-price monetary model where the Friedman rule is not optimal, because there is a distorted portfolio decision between money and physical capital.

occurs if prices are more flexible, labor is more elastic, and α is higher. The first two elements are standard (the former was first noticed by Chari Christiano Kehoe, 1997; see also Schmitt-Grohe and Uribe, 2004). The last part has a standard interpretation too: at given n, higher α implies more elastic money demand. As we will show below, less elastic money demand (lower $\alpha - n$) implies less optimal deflation—as in Khan, King and Wollman, although for a different theory of money demand.

Preferences			Production and price setting				Heterogeneity		
β	γ	φ	ε	ν	ρ^a	σ_a	α	ρ	δ
0.98	1	0.25	6	100	0.95	0.01	0.9	0.9	0.78
Model outcome									
GDP/M2	n	$(c^N - c^P)/c^P$	π	c^p	c^N	l	m^{tot}	q	
2	50%	-11%	-0.79%	0.98	0.87	1.07	0.46	1.12	

Table 1: Baseline calibration

What is the welfare cost of (*steady-state*) inflation? This is a classic question in monetary economics, going back at least to Bailey's 1956 calculation.[43] We calculate this (in the Lucas 1987 tradition): in our economy, moving from a steady-state annualized inflation rate of 2% (0.5% quarterly) to the optimal rate of -3.2% (-0.79% quarterly) is equivalent to a permanent increase in consumption of 0.61%—in line with (although slightly larger numbers than) e.g. Lucas (2000) and Imorohoglu (1992), although slightly larger.

Our model's implications for optimal policy in the long run are thus rather standard. But in the short run, things are different and this is intimately related to the lack of insurance in our model: with *long-run* inequality ($q = 1.12$), the steady state is distorted and this has implications for *short-run* optimal policy.

5.2 Optimal short-run deviations from price stability

The liquidity-insurance channel requires the central bank to accommodate some inflation volatility, as doing otherwise leads to large welfare losses. This is true in our economy even when the source of business cycles is a shock that, in the standard NK model *with* money demand but no inequality (insurance), generates no such trade-off: a plain-vanilla labor productivity shock.

Recall what happens in the baseline NK model in response to this shock: not much. A welfare-maximizing central bank keeps prices unchanged and inflation at zero, as this shock creates no

[43] A large literature analyzed this question using a variety of frameworks. To cite just some prominent examples, Lucas (2000) found that reducing inflation from 10 to 0 percent annually results in a 1 percent increase in consumption. Analyzing a monetary framework closer in spirit to the one our model embeds (based on the Bewley model), Imorohoglu (1992) showed that the welfare effects of inflation are larger in incomplete-markets economies. See Doepke and Schneider (2006), Erosa and Ventura (2002) and Ragot (2014) for reviews.

trade-off: the central bank can close the output gap costlessly, a well-known result labeled "divine coincidence" by Blanchard and Galí (2007). This result changes but only slightly when the steady state is distorted ($\Phi > 0$ in our notation), as analyzed in detail by Benigno and Woodford (2005): productivity shocks then have a "cost-push" dimension, creating a trade-off. Quantitatively, however, this is moot—subject to one caveat mentioned in the next footnote. The same is true in models incorporating a variety of other frictions—in particular, in models with monetary frictions such as Khan, King, and Wollman (2003) and Schmitt-Grohe and Uribe (2004, 2007): price stability is a robust policy prescription. Even though these models do imply inflation volatility under the optimal Ramsey policy, the welfare cost of eliminating such volatility is generically negligible.[44]

This is no longer the case in our model: optimal Ramsey policy requires volatile inflation, *and* this volatility matters for welfare. To see the first part of this argument, consider the impulse responses to a productivity shock presented in Figure 1, for three economies. With a black solid line, we have our economy under optimal policy—obtained by solving (23). With a blue dashed line, we have our monetary economy under what we label "Strict inflation targeting" (SIT): the central bank perfectly stabilizes inflation around the Ramsey-*optimal steady state inflation* (this is implemented by a Taylor rule with large ϕ_π and the optimal π^* target). Finally, we show with a red circle line optimal policy in a standard cashless equilibrium,[45] a comparison with which illustrates the extent of risk-sharing provided by money in our model. All variables are in percentage deviation from steady state, except the inflation and interest rates, which are in deviation from steady state.

[44] See for instance Table 2 in Schmitt-Grohé and Uribe (2007); see also Bilbiie, Fujiwara, and Ghironi (2014) for a result on optimal short-run price stability in a model with entry and variety, and a review of the literature using other distortions. As Benigno nd Woodford (2005) show analytically, this result changes—price stability ceases to be optimal—if, on top of $\Phi > 0$, the share of government spending in steady-state output is also non-zero.

[45] Since in the non-monetary equilibrium the steady-state inflation rate is 0, we recalibrated it to have the same steady state allocation. In particular, we reduce output by $\frac{\nu}{2}\pi^2$ and introduce a transfer between N and P households, such that the steady-state consumption and labor supply are the same in the monetary and non-monetary equilibrium, and only the steady-state inflation is different.

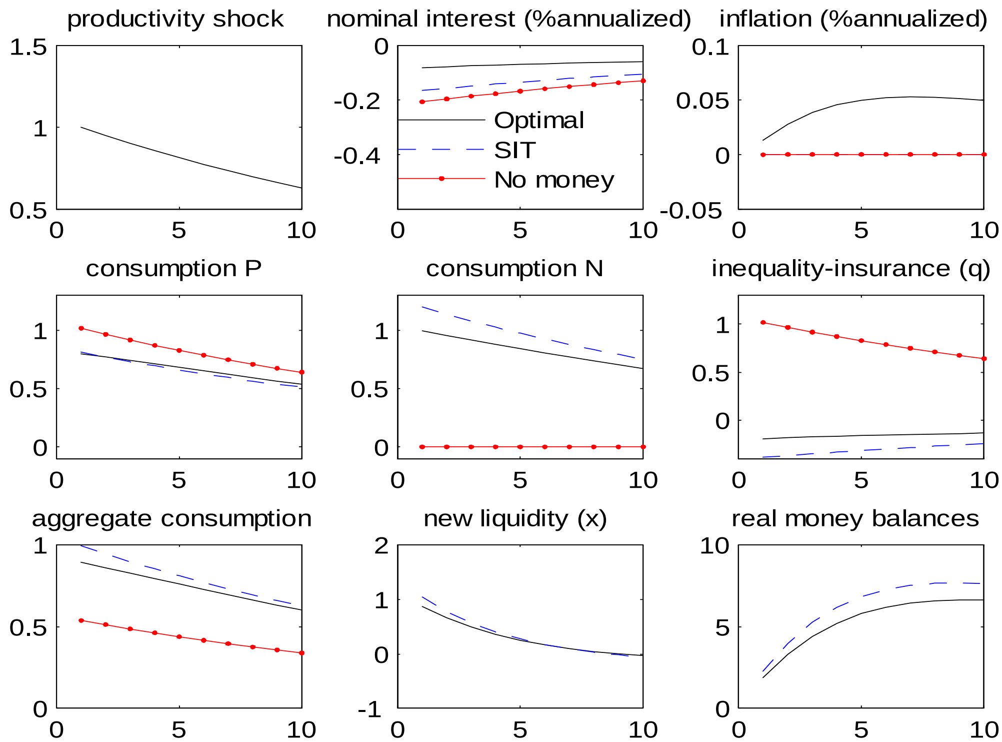

Figure 1: Responses to a labor productivity shock under optimal Ramsey policy in our model (solid black), strict inflation targeting in our model (blue dash), and optimal policy in cashless model (red circles).

The responses of the cashless model are standard: inflation does not move, and output is equal to its natural rate. Since labor productivity affects only P , their consumption increases (and so does inequality), and the nominal interest rate goes down.

In our monetary economy, the planner provides insurance: compared to the red circle line, the black solid line shows that the consumption of N increases (inequality decreases). The planner provides liquidity and interest rates fall; the result is inflation (due to the demand effect on firms), which erodes N's purchasing power (money balances) via the Pigou effect.

Consider now the allocation when this inflation is absent (blue dashed line): more liquidity is issued, and the real value of balances is much higher: thus, the consumption of N responds more, and is more volatile. Since the consumption of P is largely unchanged, the same is true for inequality-insurance. We will now show that this extra volatility is costly in terms of aggregate welfare.[46]

[46] It is by now well known, starting with the influential paper of King and Wolman (1999), that welfare calculations depend crucially upon the initial values of the Lagrange multipliers—which can be set to 0, or to their Ramsey steady-state values.Under the former choice, policy is not timeless-optimal: initial period t_0 inflation has no consequence for prior expectations, thus the policy chosen in any later period is not a continuation of t_0 policy. In the second case, policy is timeless-optimal in the sense of King and Wolman (1999) and Khan et al.(2003) (Woodford 2003 uses a different definition). The numers we report are for the former, t_0-optimal case; in the timeless-optimal case, the

Table 2 reports the standard deviations of the main variables for the (Ramsey-)optimal and SIT allocation. The volatility of inflation is comparable to that obtained by Khan et al (2003). Because of limited risk-sharing, N's consumption volatility is higher than P 's. More importantly, N's consumption volatility is higher under strict inflation targeting than under optimal policy—as a result, the volatility of our inequality measure is twice as large. This difference in volatilities translates into a large welfare cost of price stability (around the optimal asymptotic inflation rate): households need to be compensated by 0.08% of consumption every period in order to live in an economy with stable prices, rather than in one with optimal policy and inflation volatility—where we calculate these welfare costs following closely the method detailed in Schmitt-Grohe and Uribe (2007).[47]

Why do we qualify the welfare cost as "large"? The number should be compared with the welfare cost of eliminating business cycles, that is of providing the household with a certain (zero-volatility) consumption path, instead of the Ramsey-optimal but volatile path of consumption. That number, in our economy, is very small: that is less than 0.001%, even though consumption volatility is comparable to the data and to Lucas (1987), i.e. around 3% standard deviation.[48] In other words, the welfare cost of price stability is about 100 (one hundred) times—or two orders of magnitude—larger than the welfare cost of business cycle volatility in our model.

The reason for these high welfare costs of price stability is by now, we hope, clear: our long-run equilibrium is one with imperfect insurance (inequality), which for a planner is a distortion and implies a motive to provide liquidity. This distortion suffices to generate significant costs of price stability in our model, because volatility has a first-order welfare effect through the level of N's consumption. In terms of our second-order approximation, this effect makes it "as if" the weight on inflation volatility in true Ramsey loss function were smaller than Λ_π.

welfare losses are very close to zero in all cases; ses also Bilbiie, Fujiwara, and Ghironi (2014) for further discussion in a different context.

[47] The welfare losses are similar for the two types of agents: thus, our simplified heterogeneity misses some of the distributional effects emphasized by Krusell and Smith (1998)—in their framework and the subsequent literature (see Lucas 2003 for a review) the welfare benefits of *eliminating uncertainty* are asymmetric among the poor, the rich, and the middle class. One would expect that such an asymmetry occurs in a richer model of the wealth distribution also for the welfare costs that we calculate.

[48] Recall that the standard Lucas (1987, 2003) calculation delivering a cost of business cycles of 0.05% (still smaller than the cost of price stability here) is performed in a competitive, real model with exogenous labor. We argue that the right metric for assessing the benefit of eliminating business cycles is the same measure calculated for our Ramsey economy (any other assumption on policy will a fortiori be arbitrary). Recall also that the cost of business cycle measure is very sensitive to labor supply elasticity—indeed, business cycle volatility can even be beneficial in a standard RBC model with elastic labor. The welfare cost of fluctuations in a competitive version of our economy with inelastic labor and fixed money supply is 0.024%.

	Standard deviation (%)				Welfare (%)
Economies	$\hat{c}^P$	$\hat{c}^N$	$\hat{q}$	π	Δ^W
Baseline					
Ramsey	2.6	2.9	0.7	0.05	–
SIT	2.5	3.3	1.3	0	0.08
No SS inequality ($\beta \to 1$)					
Ramsey	2.5	3.1	0.6	0.05	–
SIT	2.5	3.8	1.3	0	0.00

Table 2: Standard deviations and welfare losses (percent)

That small inflation volatility translates into high welfare gains in our model is due not so much to volatility itself as to imperfect insurance (inequality). To illustrate this, consider an economy where the steady-state distortion vanishes, $q \to 1$, which amounts to taking $\beta \to 1$ and re-calibrating δ to get the same steady-state money velocity; evidently, the optimal long-run inflation rate converges to 0. As the bottom panel of Table 2 illustrates, the *volatility* of inflation under Ramsey policy in this economy is unchanged. Nevertheless, this volatility no longer means welfare: without a liquidity-insurance motive, the central bank can safely and costlessly pursue price stability (just as in Khan et al, 2003, and Schmitt-Grohe and Uribe, 2007).

Robustness 1: Different Supply Calibrations

As a first robustness check we report the same outcomes for economies with more flexible prices ($\nu = 50$) and less elastic labor ($\varphi = 1$).[49] The upper panel of Table 3 contains the results. Both the inflation volatility and its welfare benefit increase as prices become more flexible and labor supply more elastic. The reason is that with more flexible prices (lower ν), the cost of using inflation is lower: in the limit, as prices become flexible, inflation essentially becomes a lump-sum tax—an insight originally due to Chari, Christiano, and Kehoe (1997) and also discussed by Schmitt-Grohe and Uribe (2004).

[49] For each economy, in order to perform meaningful welfare comparisons we calibrate the discount factor β and home production δ, to start from the same steady state: this gives 0.973 and 0.79 for the first and 0.982 and 0.765 for the second calibration (results are similar when we keep these parameters unchanged).

			Standard deviation (%)				Welfare (%) Δ^W
Economies	$\pi^{SS}(\%)$		$\hat{c}^P$	$\hat{c}^N$	$\hat{q}$	π	$HD \rightarrow SIT$
$\nu = 50$	-1.54	Ramsey	2.7	3.2	0.5	0.08	–
		SIT	2.6	3.8	1.2	0	0.13
$\varphi = 1$	-0.6	Ramsey	2.1	2.7	0.6	0.04	–
		SIT	2.0	3.0	0.9	0	0.03
Labor market calibration	-0.36	Ramsey	3.0	5.1	2.1	0.02	–
($\alpha = 0.95$; $\rho = 0.5$; $\delta = 0.5$)		SIT	3.1	5.5	2.4	0	0.06
New Keynesian Cross	-0.37	Ramsey	2.8	3.3	1.1	0.02	–
		SIT	2.8	3.5	1.4	0	0.02

Table 3: Robustness analysis

Robustness 2: Different Demand-Side Calibration

The second alternative calibration we consider is based on labor market risk. Instead of matching financial market variables (n, μ, and q) as in our previous calibration, we draw on the labor market literature, in particular Shimer (2005) to find parameter values for α, ρ, and δ. At quarterly frequency, the job loss probability is 5% and the average job finding probability 50% for the post-war period—these two numbers imply $\alpha = 0.95$, $\rho = 0.5$ and thus $n = 0.94$; the gross replacement ratio is set to $\delta/w = 50\%$ (see also Challe and Ragot, 2014). The middle panel of Table 3 contains the results, assuming that all other parameters are as in the baseline; apart from the reported numbers, it is worth mentioning that the quarterly velocity of money is somewhat higher (2.33), and the fall in consumption when becoming unemployed is now 24%—in the upper range of the empirical estimates discussed above.

The optimal steady-state inflation rate is -0.36%: there is less deflation than in the baseline calibration, for there is less money in circulation. This is similar to the optimal deflation rate obtained by Khan et al for their calibration with low money demand elasticity (obtained by estimating money demand over a shorter sample); indeed, since α is very close to n, our calibration also implies low money demand elasticity. The similarities go further: as in that model, optimal policy also implies lower inflation volatility under this calibration; but the parallel stops here, for this smaller volatility is still associated with a large welfare cost in our model. households are willing to sacrifice 0.06% of consumption every period in order to live in an economy with optimally volatile inflation, rather than in an economy with stable prices. Our result thus survives even in this economy with very low idiosyncratic risk calibrated to labor market data.

Robustness 3: Adding the New Keynesian Cross Channel

The bottom panel of Table 3 adds the New Keynesian cross channel by making labor supply of N elastic and their income endogenous; constrained households thus have an additional margin to self-insure in face of shocks. Notice that the subsidy to firms is paid by taxing all households uniformly (otherwise, there is no reason to self-insure through money in the steady-state equilibrium) In this economy, there is the additional amplification effect through the "New Keynesian cross" mentioned above (see Appendix B.7) and described in detail in Bilbiie (2008, 2017). Furthermore, steady-state inequality is higher because there is less fiscal redistribution (under our calibration, the consumption difference is 16%)—so the distortion is larger. Monetary policy is more powerful, and there is also more insurance through the labor margin and hence less need to self-insure. In addition, as emphasized above, a linear term in the welfare function now penalizes expansions that are labor-driven, even when they provide insurance. In equilibrium, there is thus less inflation and inflation volatility because the liquidity-insurance motive is weaker. Consequently, the welfare costs of price stability are smaller—but they are still much larger (30 times larger) than the welfare costs of Ramsey cycles (which are again very small, smaller than 0.001%).

Robustness 4: Adding the Markup Distortion

All our previous calibrations assumed that there is an optimal subsidy that undoes the steady-state monopolistic distortion, $\Phi = 0$; this allows isolating the novel channel that operates in our framework. We now report one last set of robustness checks, assuming that there is no such subsidy $\sigma = 0$.

Economies	$\pi^{SS}(\%)$	$sd(\pi)(\%)$	Δ_0^W
Baseline	-1.54	0.06	0.48
$\nu = 50$	-2.6	0.1	0.55
$\varphi = 1$	-0.1	0.04	0.15
$\beta \to 1$	-0.7	0.05	0
Labor Market	-0.6	0.02	0.33
NK Cross	-0.4	0.02	0.16

Table 4: A distorted steady state, $\sigma = 0$

As Table 4 shows, the welfare losses are now *much* larger, about five to seven times. The notable exception is the case when there is long-run insurance (no steady-state inequality): the welfare loss is, again, zero—as our second-order approximation showed, the linear term in the loss function disappears in this case. The two long-run distortions are thus *complementary* in generating significant losses from price stability. This result is related to Benigno and Woodford (2005), who showed that a distorted steady state implies significant deviations from price stability *only* when the steady-state government spending share is non-zero. Our framework thus identifies another channel—which, when shut off, makes price stability again optimal even when the monopolistic distortion is large. But when our liquidity-insurance channel is at work ($q > 1$), the optimal deviations from price stability are very large indeed if supply-side distortions are also an issue ($\Phi > 0$).

6 Open-Market Operations

In our model, it makes a difference whether monetary policy is conducted via helicopter drops or open market operations. What are the implications for optimal monetary policy, and which means of issuing money is preferable from a welfare standpoint? To answer this question, we now assume that instead of money being injected through a transfer, it is exchanged for bonds through open market operations OM. We maintain a consolidated budget constraint for the government:

$$m_{t+1}^{tot} + b_{t+1} + \tau_t = \frac{1}{1+\pi_t} m_t^{tot} + \frac{1+i_{t-1}}{1+\pi_t} b_t \tag{24}$$

where τ_t are taxes, and b_{t+1} debt (when negative, these are assets). Using that money created by the central bank (money growth, or seigniorage) is given as before by (10) and replacing, we obtain:

$$b_{t+1} + \tau_t + x_t = \frac{1+i_{t-1}}{1+\pi_t} b_t \tag{25}$$

The way of money creation is dictated by how taxes/transfers adjust when money is introduced $x_t > 0$. At one extreme we have the previous case of HD (within-period transfer $\tau_t = -x_t$). As another extreme, consider OM with one-period repayment: no transfer within the period $\tau_t = 0$ but all debt gets repaid next period (to be precise, when money is issued and exchanged for bonds, this amounts not to repayment but to collecting, and transferring the proceeds):

$$\tau_{t+1} = \frac{1+i_t}{1+\pi_{t+1}} \left(-x_t + \frac{1+i_{t-1}}{1+\pi_t} b_t \right). \tag{26}$$

The general case consists of a "tax rule" that ensures that the intertemporal budget equation holds as a constraint for any price level (Leeper, 1991; Woodford, 1996; see Leeper and Leith, 2016 for a review)—we outline this in the Appendix for completion. Insofar as optimal policy is concerned, however, the one-period-repayment case happens to also be the one that delivers the highest welfare with the class of OM policies; this is intuitive, because as we shall see OM imposes further constraints on the amount of insurance that monetary policy can achieve, and the faster the repayment, the better the insurance properties.

To understand the key difference between HD and OM, consider the budget constraint of N under OM with one-period repayment:

$$c_t^N = \delta + \frac{1+i_{t-1}}{1+\pi_t} x_{t-1} + \frac{1-\alpha}{n} \frac{m_t^{tot}}{1+\pi_t} \tag{27}$$

Monetary policy affects demand of N through two channels. First, the *real balance, Pigou effect* operates, regardless of the means of money creation: inflation erodes the real value of money balances, although the equilibrium inflation *is* different under HD and OM. Second, there is the key difference between HD and OM: period t money creation (x_t) in the HD case relaxes the budget constraint of N within the period, whereas in the OM case it affects the budget constraint of N-households only starting from the following period.

This implies that the Pigou effect does more work under OM, for it needs to compensate for the lack of a transfer: equilibrium inflation volatility is larger than under HD. When inflation volatility is costless (under flexible prices), this is largely irrelevant: both OM and HD allow reaching the same optimal allocation. But when prices are sticky and inflation is costly, a clear difference between OM and HD emerges, with important welfare consequences.

For a welfare-maximizing central bank, this implies that there are additional instruments and constraints. In particular, the Ramsey problem now ought to include both new money x_t and nominal interest i_t as instruments; the additional constraints for the Ramsey planner are the budget constraint of N under OM (27), the Euler equation for bonds (5) and the definition of money growth (10).

We compute Ramsey policy when monetary policy is conducted via OM with one-period repayment and assess the optimal way of money creation.[50] Figure 2 illustrates the responses to a TFP shock under OM Ramsey policy, comparing them to the responses under HD Ramsey already illustrated in Figure 1.

[50] Numerical results support our previous intuition that this policy arrangement (with one-period repayment, i.e. $\phi_b = 1$) delivers the highest welfare within the class of OM policies; this biases the results in favor of OM, that is it provides a lower bound on its welfare costs relative to HD.

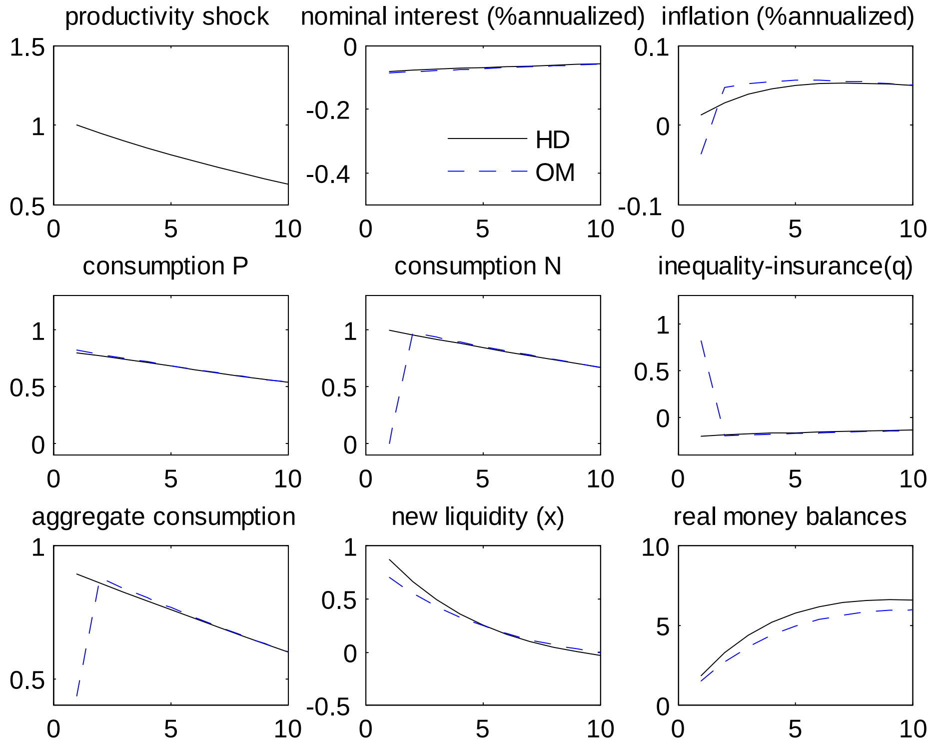

Figure 2: Responses to a labor productivity shock under optimal Ramsey policy with (solid black), versus OM with one-period repayment (blue dash).

A first main difference between HD and OM (illustrating the more general discussion above) is that when money is created by OM c^N does not move on impact, but increases sharpy only one period after the shock—in contrast to HD where it increases on impact. In other words, there is more risk-sharing for the aggregate risk when money is created under HD (albeit only on impact): we assess below the welfare cost generated by this difference. A second, related difference concerns inflation, which under OM first falls before increasing sharply (whereas it is much smoother under HD).[51] With OM, there is deflation today to provide insurance and increase c^N – because the transfer cannot be used, in other words, the Pigou effect needs to bear the adjustment. But there is also expected inflation, which increases c^P through the interest rate channel, i.e. intertemporal substitution. The planner ought to do this too, since otherwise P are hurt by deflation today, and the planner's ultimate objective is insurance. This generates volatile inflation under OM compared to HD.[52]

[51] A "production"-based explanation is as follows. Monetary policy under OM does not increase the consumption of N on impact, so aggregate demand under OM is lower than under HD (see total consumption c), and so is aggregate production: in fact, labor (not pictured) falls on impact under OM, while it increases under HD. Due to this labor market response, the real wage falls under OM and increases slightly under HD. Through the Phillips curve, this translates on impact into deflation under OM, and almost constant inflation under HD.

[52] The immediate switch from deflation to inflation under OM occurs because we assume one-period bonds and the

In a nutshell, aggregate demand is (optimally) more cyclical under HD, because money is transferred to non-participating households who have a high marginal propensity to consume. This translates into a more stable inflation under HD than under OM. Monetary policy through OM tries to make up for the lack of an instant transfer by exploiting the Pigou effect through inflation, which has perverse effects and can have large welfare costs. It is intuitive that, since monetary policy has a redistributive role in our model, HD is a preferable way to issue money because it has better insurance properties: it provides a direct transfer within the period, and it does not imply (inefficiently) relying upon the Pigou effect to do the job. The question we ask is: how large is this? What is the welfare gain of switching, in response to the same shock, from OM to HD? In all economies we studied, this welfare cost is "large"—especially compared to the small (˜0.001%) benefit of eliminating volatility reported above. As Table 5 shows, it ranges from 0.01% under the baseline model to above 0.05% in an economy with endogenous N labor and a distorted steady state (compared to the cost of business cycle volatility measured in these economies, these numbers are 10 and respectively 50 times larger—one to two orders of magnitude).

			Std. dev. (%)				Welfare (%) Δ^W
Open-Market Ramsey		π^{SS} (%)	$\hat{c}^P$	$\hat{c}^N$	$\hat{q}$	π	$OM \to HD$
Baseline							
	efficient SS	-0.79	2.6	2.7	1.1	0.05	0.01
	distorted SS	-1.54	2.2	2.5	1.0	0.07	0.026
NK cross							
	efficient SS	-0.37	3.1	3.7	1.6	0.12	0.01
	distorted SS	-0.41	2.7	3.1	1.5	0.09	0.052

Table 5. Ramsey OM: Standard deviations and welfare losses

Using our previous computations, we can also assess the welfare costs of price stability instead of following the Ramsey-optimal policy, but conducted through OM. These follow directly by subtracting from the numbers in the last column of Table 5 the relevant number from Tables 2, 3 and 4. We thus obtain, for the same four cases as in Table 5: 0.07, $0.45, 0.01$, and 0.11 respectively.

7 Conclusions

In monetary policy analysis, a new synthesis looms: the integration of sticky-price, New Keynesian models and models of heterogeneous households, incomplete markets and limited participation. This

transfer of cental bank profits takes place within one period. Changing these assumptions would deliver a smoother path, but the intuition would be the same: first deflate, then inflate (because what matters for for P agents is the "long" rate).

very active research area (reviewed in the Introduction) started in the early 2000s and is going at full speed. We hope to have contributed to these convergence efforts a fully-fledged, NK-style *optimal* monetary policy analysis, in a tractable framework that captures key mechanisms of heterogeneous-agent incomplete-markets models and includes a deep reason for money as a self-insurance device. Liquidity and insurance (a limited notion of inequality) are intimately related in our model: today's liquidity (lower interest rates) implies tomorrow's insurance (and lower inequality). Aggregate demand in this model depends on: money, or liquidity, which relaxes the constraint of non-participating households and provides a saving vehicle for participating households; interest, because of intertemporal substitution by participating households; and prices, or inflation, because a Pigou effect operates for non-participating households, and (expected) inflation is the relevant return for holding liquidity for participants.

The link between liquidity and insurance is the keystone for optimal monetary policy in this model: a novel trade-off arises between providing liquidity for insurance purposes, and stabilization of inflation and real activity. We first illustrate this trade-off analytically by means of a second-order approximation to the aggregate welfare function, in the New Keynesian tradition pioneered by Woodford (2003): there is a first-order benefit to providing insurance, insofar as the long-run equilibrium is characterized by a lack thereof (that is, by long-run consumption inequality). This first-order benefit of liquidity provision implies that the standard objective of eliminating inflation volatility take a back seat.

A quantitative assessment of this trade-off shows that in an economy with a long-run equilibrium characterized by imperfect insurance (consumption inequality), deviations from price stability are optimal. This holds, first, in the long run: the optimal inflation target should be between zero and the Friedman rule; this is no surprise—it is true in most monetary models. But in our framework, unlike in others, it is *also true in the short run.* Optimal policy implies inflation volatility in response to (productivity) shocks that otherwise create no trade-off.

What is more, this volatility matters for welfare. A policy of stabilizing prices (albeit around the optimal inflation target) incurs a large welfare loss. This happens because short-run volatility has a first-order effect on constrained households: optimal policy requires giving *less weight* to inflation stabilization—which de facto implies giving more weight to constrained households.

This cannot be emphasized enough: the optimal policy prescription is *not* that the central bank should or needs to do anything radically different from what central banks are currently doing. Instead, it describes how, within that existing policy framework, this novel liquidity-insurance motive can be reinterpreted as a quantitative modification of the central banks' policy objectives—namely, more tolerance to inflation volatility when this is a side-effect of liquidity provision.

While we view our study as a step in the direction of the new synthesis that we mention at the outset of these concluding comments, we think such efforts should continue, for much remains to be done. Our tractable framework allows the calculation of optimal policy, but it inherently leaves out several other, surely important redistributive aspects of monetary policy acknowledged in Section 2.1. Incorporating some of these other "HANK" channels (more realistic wealth distributions; endogenous portfolio shares; nominal debt; endogenous unemployment risk; etc.) is paramount in order to attain a thorough understanding of how monetary policy works and how it should be conducted in a world where household heterogeneity matters.

8 References

Adao, B., I. Correia, and P. Teles, 2003. "Gaps and Triangles," *Review of Economic Studies*, 70(4), 699-713

Adam, K. and J. Zhu, 2014 "Price Level Changes and the Redistribution of Nominal Wealth Across the Euro Area" *Journal of the European Economic Association*

Adam, K. and P. Tzamourani, 2016, "Distributional Consequences of Asset Price Inflation in the Euro Area", *European Economic Review*, 89, 172–192.

Algan, Y. and X. Ragot. 2010., "Monetary Policy with Heterogeneous Agents and Borrowing Constraints." *Review of Economic Dynamics*, 13, 295-316.

Algan, Y., E. Challe and X. Ragot, 2010. "Incomplete Markets and the Output-Inflation Trade-off", *Economic Theory*, 46(1), 2011, 55-84

Alvarez F., A. Atkeson and C. Edmond. 2009. "Sluggish Responses of Prices and Inflation to Monetary Shocks in an Inventory Model of Money Demand." *The Quarterly Journal of Economics*, MIT Press, vol. 124(3), pages 911-967, August.

Alvarez, F., A. Atkeson and P. J. Kehoe. 2002. "Money, Interest Rates, and Exchange Rates with Endogenously Segmented Markets," *Journal of Political Economy*, University of Chicago Press, vol. 110(1), pages 73-112, February.

Alvarez, F. and A. Atkeson, 1997. "Money and exchange rates in the Grossman-Weiss-Rotemberg model." *Journal of Monetary Economics* 40(3), pages 619-640, December.

Alvarez, F. and F. Lippi. 2009. "Financial Innovation and the Transactions Demand for Cash." *Econometrica* 77 (2): 363–402.

Alvarez, F. and F. Lippi. 2014. "The demand of liquid assets with uncertain lumpy expenditures." *Journal of Monetary Economics*, forthcoming.

Alvarez, F. and F. Lippi. 2014. "Persistent Liquidity Effects and Long-Run Money Demand." *American Economic Journal: Macroeconomics* 6(2), 71-107, April.

Andrés, J., O. Arce and C. Thomas, 2013, "Banking competition, collateral constraints, and optimal monetary policy," *Journal of Money, Credit and Banking*, 45(s2)

Auclert, A. 2016, "Monetary Policy and the Redistribution Channel" Mimeo Stanford

Bailey, M., 1956, "The Welfare Cost of Inflationary Finance", *Journal of Political Economy*, 64(2), 93-110

Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden, 2017, "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk" Mimeo Bonn GSE

Benigno, P., and M. Woodford. 2005. "Inflation Stabilization and Welfare: The Case of a Distorted Steady State." *Journal of the European Economic Association*, 3(6): 1185–1236.

Benigno, P., and M. Woodford. 2012. "Linear-Quadratic Approximation of Optimal Policy Problems." *Journal of Economic Theory*, 147(1): 1–42.

Bernanke, B. S. 2007, "The Level and Distribution of Economic Well-Being," speech delivered at the Greater Omaha Chamber of Commerce, Omaha, Nebraska

Bernanke, B. S. 2015, "Monetary Policy and Inequality", Brookings.edu

Bewley, T. F. 1977. "The Permanent Income Hypothesis: A Theoretical Formulation". *Journal of Economic Theory* 16: 252–92

Bewley, T. F. 1983 "A Difficulty with the Optimum Quantity of Money" *Econometrica*, 1983, 51(5), 1485-504

Bhandari, A., D. Evans, M. Golosov, and T. Sargent, 2017, "Optimal Fiscal-Monetary Policy with Redistribution", Mimeo

Bilbiie F. O. 2004. "Limited Asset Market Participation, Monetary Policy, and (Inverted) Keynesian Logic", Working Paper Nuffield College, Oxford University

Bilbiie F. O. 2008. "Limited Asset Market Participation, Monetary Policy, and (Inverted) Aggregate Demand Logic." *Journal of Economic Theory* 140, 162-196.

Bilbiie, F.O., 2017, "The New Keynesian Cross: Understanding Monetary Policy with Hand-to-Mouth Households", Mimeo PSE.

Bilbiie, F. O., I. Fujiwara, and F. Ghironi, 2014, "Optimal Monetary Policy with Endogenous Entry and Product Variety." *Journal of Monetary Economics*, 64, 1-20

Bilbiie, F.O., and R. Straub, 2012, "Changes in the Output Euler Equation and Asset Markets Participation" J*ournal of Economic Dynamics and Control* 36(11), 1659-1672

Bilbiie F. O. and R. Straub, 2013, "Asset Market Participation, Monetary Policy Rules and the Great Inflation", *Review of Economics and Statistics*, 95(2), pp 377-392

Bilbiie F. O., T. Monacelli and R. Perotti 2013, "Public Debt and Redistribution with Borrowing Constraints", *The Economic Journal* 2013, F64-F98, 02

Bilbiie, F.O. and X. Ragot, 2017, "Liquidity, Aggregate Demand, and Interest Rate Rules", Mimeo

Braun, A. and T. Nakajima, 2012. "Uninsured Countercyclical Risk: An Aggregation Result And Application To Optimal Monetary Policy," Journal of the European Economic Association, 10(6), 1450-1474.

Bricker, J., L. Dettling, A. Henriques, J. Hsu, K. Moore, J. Sabelhaus, J. Thompson, and R. Windle.

2014. “Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances,” Federal Reserve Bulletin, 100(4).

Brunnermeier, M. K., and Y. Sannikov. 2016. “On the Optimal Inflation Rate.” American Economic Review Papers and Proceedings

Cao, S, C. A. Meh, J.V. Rıos-Rull and Y. Terajima. 2012. "Inflation, Demand for Liquidity and Welfare", Working Paper.

Campbell, J. Y. and Mankiw, N. G. 1989. ”Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence,” in Blanchard, O. and Fisher, S., eds, NBER Macroeconomics Annual 185-216, MIT Press.

Challe, E., 2017, "Uninsured unemployment risk and optimal monetary policy," Mimeo Ecole Polytechnique

Challe, E. and X. Ragot, 2014 “Precautionary Saving Over the Business Cycle,” *The Economic Journal*.

Challe, E; Ragot, X. Matheron, J. and J. Rubio-Ramirez, (2016), "Precautionary saving and aggregate demand", *Quantitative Economics*, Forthcoming

Chari,V.V., L.J. Christiano, and P.J. Kehoe, 1991, Optimal fiscal and monetary policy: some recent results, *Journal of Money, Credit, and Banking* 23 519–539.

Christiano, L. J., 1991 "Modelling the Liquidity Effect of a Monetary Shock" Quarterly Review Minneapolis Fed

Christiano, L. J. and M. Eichenbaum 1992. "Liquidity effects and the monetary transmission mechanism." *American Economic Review* PP

Christiano, L. J. and M. Eichenbaum 1995. "Liquidity effects, monetary policy and the business cycle." *Journal of Money, Credit and Banking* 27(4,1): 1113-1136.

Christiano, L. J., M. Eichenbaum and C. Evans, 2005, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1): 1-45

Chodorow-Reich, G. and L. Karabarbounis 2014, “The Cyclicality of the Opportunity Cost of Employment,” Mimeo Harvard.

Clarida, R., J. Galí and M. Gertler, 1999, "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37 1661-1707

Cloyne, J., C. Ferreira, and P. Surico 2015 "Monetary Policy when Households have Debt: New Evidence on the Transmission Mechanism" CEPR Discussion Paper

Cœuré, B. 2012, “What can monetary policy do about inequality?” Speech at Extreme Poverty and Human Rights, Fourth World Committee, European Parliament, Brussels

Coibion, O., Y. Gorodnichenko, L. Kueng and J. Silvia. 2012. "Innocent Bystanders? Monetary Policy and Inequality in the U.S", Forthcoming, *Journal of Monetary Economics*.

Correia, I. and P. Teles 1999, "The optimal inflation tax", *Review of Economic Dynamics*.

Cui, W. and S. Radde, 2016, "Search-Based Endogenous Asset Liquidity and the Macroeconomy", Mimeo UCL and ECB

Curdia, V. and M. Woodford, 2009, "Credit Frictions and Optimal Monetary Policy", Forthcoming, *Journal of Monetary Economics.*

Den Haan,W., Rendahl, P. and M. Riegler 2016, "Unemployment (Fears) and Deflationary Spirals", working paper.

Deutsche Bundesbank Monthly Report (2016), "Distributional Effects of Monetary Policy", September 2016.

Doepke, M. and M. Schneider 2006, "Inflation and the Redistribution of Nominal Wealth", *Journal of Political Economy* 114 (6), 1069-97

Draghi, M., 2016 "Stability, Equity and Monetary Policy" 2nd DIW Europe Lecture, DIW Berlin

Eggertsson G. and P. Krugman 2012, "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach, *Quarterly Journal of Economics*, 127(3): 1469-1513.

Erosa, A., Ventura,G. 2002. "On Inflation as a Regressive Consumption Tax." *Journal of Monetary Economics* 49, 761-795.

Feldman, M and Gilles, C. 1985. "An Expository note on Individual risk without aggregate uncertainty." *Journal of Economic Theory* 35, 26-32.

Friedman, M. 1969, The Optimum Quantity of Money, Macmillan

Fuerst, T. 1992. "Liquidity, loanable funds and economic activity." *Journal of Monetary Economy* 29(1),3-24.

Furceri, D., P. Loungani,and A. Zdzienicka, 2016 "The Effects of Monetary Policy Shocks on Inequality" International Monetary Fund Working Paper

Galí, J. 2008, "Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications", Princeton University Press

Galí, J., Lopez-Salido D. and J. Valles. 2004. "Rule-of-Thumb Consumers and the Design of Interest Rate Rules." *Journal of Money, Credit, and Banking* 36(4), 739-764.

Galí, J., D. López-Salido, and J. Vallés, 2007. "Understanding the Effects of Government Spending on Consumption". *Journal of the European Economic Association*, March, vol. 5 (1), 227-270.

Gorea, D. and V. Midrigan, 2017, "Liquidity Constraints in the U.S. Housing Market." Mimeo NYU

Gornemann, N., K. Kuester, and M. Nakajima, 2012 "Monetary Policy with Heterogeneous Agents," Working Papers 12-21, Federal Reserve Bank of Philadelphia

Gottlieb, C. 2015, "On the Distributive Effects of Inflation" Mimeo EUI and Oxford

Green, E. 1994. "Individual-level randomness in a nonatomic population", working paper.

Grossman S.J and Weiss L. 1983. "A transactions-based model of the monetary transmission mechanism." *American Economic Review.* 73 (1983). 871-880.

Iacoviello, M., 2005 House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle, *American Economic Review*

Imrohoroglu, A., 1992. The welfare cost of inflation under imperfect insurance. *Journal of Economic Dynamics and Control* 16, 79-91.

Ireland, P. 1996. "The role of countercyclical monetary policy." *Journal of Political Economy* 104(4), 704-723.

Jappelli, T and Pistaferri, L, 2014, "Fiscal Policy and MPC Heterogeneity", *American Economic Journal: Macroeconomics*, 6(4), 107-36

Kaplan, G. and G. Violante, 2014 "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*

Kaplan, G., G. Violante, and J. Weidner, 2014 "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity.*

Kaplan, G., B. Moll and G. Violante 2015, "Monetary Policy According to HANK", mimeo Princeton and NYU

Kehoe, T., D. Levine, and M. Woodford, 1992, "The optimum quantity of money revisited," in Partha Dasgupta, Douglas Gale, Oliver Hart, and Eric Maskin, editors, The Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn, MIT Press, 1992, 501–26.

Khan, A., R. G. King, and A. L. Wolman 2003. "Optimal Monetary Policy." *Review of Economic Studies*, 70(4), 825-860.

Khan, A. and J. Thomas 2011. "Inflation and Interest Rates with Endogenous Market Segmentation", working paper.

King, R. G., and A. L. Wolman, 1999 "What Should the Monetary Authority Do When Prices are Sticky?," in John B. Taylor, editor, Monetary Policy Rules, The University of Chicago Press, 1999, 349-398.

Kiyotaki, N. and J. Moore, 1997 "Credit Cycles," Journal of Political Economy, 105 (1997), 211-248

Kiyotaki, N and M. Gertler, 2010. "Financial Intermediation and Credit Policy in Business Cycle Analysis," in *Handbook of Monetary Economics*, Volume 3A, eds B. Friedman and M. Woodford, Amsterdam: Elsevier

Kiyotaki, N. and J. Moore, 2012, "Liquidity, Business Cycles, and Monetary Policy," mimeo Princeton

Krueger, D., K. Mitman and F. Perri, 2017, "Macroeconomics and Household Heterogeneity, Including Inequality", *Handbook of Macroeconomics*, Vol 2B, chapter 11

Krusell, P. and A. Smith. 1998. "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy* 106(5), 867-896.

Leeper, E. M., 1991, "Equilibria under active and passive monetary and fiscal policies", *Journal of Monetary Economics*, 129-147

Leeper, E. M. and C. Leith, 2016, "Understanding Inflation as a Joint Monetary-Fiscal Phenomenon", *Handbook of Macroeconomics*, vol 2, Eds. J. Taylor and H. Uhlig

Le Grand, F. and Ragot, X. 2016 "Optimal fiscal policy with heterogeneous agents and aggregate shocks", mimeo, Sciencespo.

Lippi, F., S. Ragni and N. Trachter 2015. "Optimal monetary policy with heterogeneous money holdings." *Journal of Economic Theory* 339-368.

Lucas, R. E. Jr. 1987. “Models of Business Cycles.” Oxford: Basil Blackwell.

Lucas R. E. Jr., 1990 "Liquidity and Interest Rates." *Journal of Economic Theory* 50, 237 - 264.

Lucas, R.E. Jr., 2000. "Inflation and Welfare". *Econometrica* 68, 247-274

Lucas, R. E., Jr. 2003 "Macroeconomic Priorities". *American Economic Review* 93 (1): 1–14.

Lucas, R. E., Jr., and N. L. Stokey, 1983, “Optimal Fiscal and Monetary Policy in an Economy without Capital,” *Journal of Monetary Economics* 12, 55-93.

McKay A. and R. Reis, "The Role of Automatic Stabilizers in the U.S. Business Cycle" , *Econometrica*, Vol. 84, No. 1, January 2016.

McKay A., Nakamura E. and J. Steinsson 2016, "The power of forward guidance revisited", *American Economic Review*.

Mehrling, P., 1995, “A Note on the Optimum Quantity of Money.” *Journal of Mathematical Economics* 24(3): 249–58

Mersch, Y., 2014, "Monetary policy and economic inequality", Keynote Speech at Corporate Credit Conference, Zurich

Misra, K. and P. Surico, 2014, "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs"*American Economic Journal: Macroeconomics*, 6(4),

Nistico, S., 2015, "Optimal monetary policy and financial stability in a non-Ricardian economy." Forthcoming, *Journal of the European Economic Association*

Nuno, G. and C. Thomas, 2017, "Optimal Monetary Policy with Heterogeneous Agents" Mimeo Bank of Spain

Patinkin, D. 1956, "Money, Interest and Prices: An Integration of Monetary and Value Theory", Second edition (1965), MIT Press

Pigou, A.C., 1943. "The Classical Stationary State". *Economic Journal* 53 (212): 343–351

Ragot, X. 2014. "The Case for a Financial Approach to Money Demand." *Journal of Monetary Economics*, in press.

Ragot, X. 2016. "Money and Capital Accumulation over the Business Cycle." *Mimeo*.

Ravn, M. and V. Sterk, 2013 “Job Uncertainty and Deep Recessions,” forthcoming, *Journal of Monetary Economics*.

Rocheteau, G., and P. O. Weill 2011: “Liquidity in Frictional Asset Markets,” *Journal of Money, Credit and Banking*, 43, 261–282.

Rocheteau, G., P. O. Weill and T. Wong, 2015, "A Tractable Model of Monetary Exchange with Ex-Post Heterogeneity" Mimeo UC Irvine, UCLA and Bank of Canada

Rocheteau, G., P. O. Weill and T. Wong, 2016, "Working through the Distribution: Money in the Short and Long Run" Mimeo UC Irvine, UCLA and Bank of Canada

Rotemberg, J. J. 1982, “Monopolistic Price Adjustment and Aggregate Output”. *Review of Economic Studies*, 49(4), 517-531

Samuelson, P. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy*, 66, 467.

Scheinkman, J.A. and Weiss, L., 1986, "Borrowing constraints and aggregate economic activity", *Econometrica*, 54(1), 23-45.

Schmitt-Grohe, S. and M. Uribe, 2004, "Optimal Fiscal and Monetary Policy under Sticky Prices." *Journal of Economic Theory*, February 2004, 114(2): 198-230

Schmitt-Grohe, S. and M. Uribe, 2007, "Optimal Simple and Implementable Monetary and Fiscal Rules." *Journal of Monetary Economics* 54, 6

Sheedy, K. D. 2014, "Debt and Incomplete Financial Markets: A Case for Nominal GDP Targeting," *Brookings Papers on Economic Activity*, 301-373

Shimer, R. 2005. "The cyclical behaviour of equilibrium employment and vacancies." *American Economic Review* 95(1), pp. 25–49.

Sterk, V. and S. Tenreyro, 2016, "The Transmission of Monetary Policy through Redistributions and Durable Purchases" Mimeo UCL and LSE

Surico, P. and R. Trezzi, 2016, "Consumer Spending and Fiscal Consolidation: Evidence from a Housing Tax Experiment", Mimeo LBS and FRB

Vissing-Jorgensen, A. 2002. "Towards an explanation of the households portfolio choice heterogeneity: Non financial income and participation cost structure" NBER Working Paper 8884.

Vissing-Jorgensen, A. 2002. "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution" *Journal of Political Economy.*

Wallace, N. 1981. A Modigliani-Miller Theorem for Open-Market Operations, *The American Economic Review*, Vol. 71, No. 3 (Jun., 1981), pp. 267-274

Wallace, N. 2014. "Optimal money-creation in pure-currency economies: a conjecture." *Quarterly Journal of Economics* 129(1), 259-274.

Weill, P. O. 2007 “Leaning Against the Wind,” *Review of Economic Studies*, 74(4), 1329–1354.

Werning, I., 2015, "Incomplete Markets and Aggregate Demand", Mimeo MIT

Woodford, M. 1990. "Public Debt as Private Liquidity" *American Economic Review* 80(2), pp. 382-88

Woodford, M. 1996. "Controlling the Public Debt: A Requirement for Price Stability?" *NBER Working Paper*

Woodford, M. 1998. "Doing Without Money: Controlling Inflation in a Post-Monetary World," *Review of Economic Dynamics*, 1: 173-219

Woodford, M., 2003, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press

Woodford, M. 2012. "Methods of Policy Accommodation at the Interest-Rate Lower Bound”, Jackson Hole Symposium.

Yellen, J. L. 2014, "Perspectives on Inequality and Opportunity from the Survey of Consumer Finances," speech delivered at the Conference on Economic Opportunity and Inequality, Federal Reserve Bank of Boston, Boston, Mass.

A Model Summary

The equations describing our model in the general case are:

$$
\begin{aligned}
u'\left(c_t^P\right) &= \beta E \frac{1+i_t}{1+\pi_{t+1}} u'\left(c_{t+1}^P\right) \\
u'\left(c_t^P\right) &= \beta E\left[\alpha u'\left(c_{t+1}^P\right) + (1-\alpha) u'\left(c_{t+1}^N\right)\right] \frac{1}{1+\pi_{t+1}} \\
u'\left(c_t^N\right) &\geq \beta E\left[(1-\rho) u'\left(c_{t+1}^P\right) + \rho u'\left(c_{t+1}^N\right)\right] \frac{1}{1+\pi_{t+1}} \\
c_t^P + \tilde{m}_{t+1}^P &= w_t l_t^P - \tau_t^P + \frac{1}{n} d_t + \frac{\alpha}{1+\pi_t} \tilde{m}_t^P + \frac{1-\alpha}{1+\pi_t} \tilde{m}_t^N \\
\tilde{m}_{t+1}^N + c_t^N &= \delta - \tau_t^N + \frac{1-\rho}{1+\pi_t} \tilde{m}_t^P + \frac{\rho}{1+\pi_t} \tilde{m}_t^N \\
\pi_t\left(1+\pi_t\right) &= \beta E_t\left[\left(\frac{c_t^P}{c_{t+1}^P}\right)^\gamma \frac{Y_{t+1}}{Y_t} \pi_{t+1}\left(1+\pi_{t+1}\right)\right] + \frac{\varepsilon-1}{\nu}\left[\frac{\varepsilon}{\varepsilon-1} \frac{w_t}{A_t} - (1+\sigma)\right] \\
Y_t &= n A_t l_t^P \\
w_t &= \chi\left(l_t^P\right)^\varphi \left(c_t^P\right)^\gamma \\
d_t &= \left(1+\sigma - \frac{w_t}{A_t} - \frac{\nu}{2}\pi_t^2\right) Y_t \\
n \tilde{m}_{t+1}^P + (1-n) \tilde{m}_{t+1}^N &= \frac{n \tilde{m}_t^P + (1-n) \tilde{m}_t^N}{1+\pi_t} + x_t \\
\tau_t &= -x_t \\
\tau_t^P &= \frac{\omega}{n} \tau_t ; \;\; \tau_t^N = \frac{1-\omega}{1-n} \tau_t \\
\tilde{m}_{t+1}^P, \tilde{m}_{t+1}^N &\geq 0
\end{aligned}
$$

where m_{t+1}^{CB} is a period t monetary shock (new money created) and A_t exogenous labor productivity. The economy resource constraint follows by Walras' law:

$$
n c_t^P + (1-n) c_t^N = \left(1 - \frac{\nu}{2}\pi_t^2\right) Y_t + (1-n)\delta
$$

and can replace (for instance) the P agents' budget constraint in the system above.

An equilibrium of the economy is a sequence $\{c_t^P, c_t^N, c_t\, \pi_t, x_t, \tilde{m}_{t+1}^P, \tilde{m}_{t+1}^N, w_t, l_t^P, \tau_t^P, \tau_t^N, \tau_t, d_t, Y_t\}$ satisfying the previous conditions. Assuming that nominal bonds are in zero net supply, we guess-and-verify the structure of the equilibrium with $\tilde{m}_t^N = 0$, i.e. non-participating households never hold money at the end of the period. The conditions for households in the N island not to hold money, which we check holds in the equilibrium we consider, is:

$$
u'\left(c_t^N\right) > \beta E\left[(1-\rho) u'\left(c_{t+1}^P\right) + \rho u'\left(c_{t+1}^N\right)\right] \frac{1}{1+\pi_{t+1}}.
$$

Consider the conditions for a monetary equilibrium to exist.[53] In a monetary steady-state, N agents do not hold money when $1+\pi > \beta$, while P agents save in money comparing the gain to self-insure and the opportunity cost (deflation). Thus, we have $c^P > c^N$ and $q > 1$. Since it is costly for P agents to save (the return on money is lower than the discount factor), they rationally choose not to perfectly self-insure. Using this inequality in the condition (7), we have $u'\left(c^N\right) > \left[(1-\rho)\,u'\left(c^P\right) + \rho u'\left(c^N\right)\right]\frac{\beta}{1+\pi}$: N agents do not hold money at the end of each period, and $\tilde{m}^N = 0$. In steady state, positive money demand requires the restriction that the outside option not be too good:

$$\delta_c < \bar{\delta} = h = \left[1 + n\left(\left(\frac{1+\pi-\alpha\beta}{\beta\,(1-\alpha)}\right)^{\frac{1}{\gamma}} - 1\right)\right]^{-1} \tag{28}$$

Under a Taylor rule, the steady-state inflation rate π is determined by the central bank's target and the above condition is parametric. Under Ramsey policy, π is endogenous (it depends, among other things, on δ) and the above condition defines a threshold implicitly.

B Derivations and Proofs

B.1 New Keynesian Phillips curve

The intermediate goods producers solve:

$$\max_{P_t(z)} E_0 \sum_{t=0}^{\infty} Q^s_{0,t}\left[(1+\sigma)\,P_t(z)Y_t(z) - W_t l_t(z) - \frac{\nu}{2}\left(\frac{P_t(z)}{P_{t-1}(z)} - 1\right)^2 P_t Y_t\right],$$

where $Q^s_{0,t} \equiv \beta^t_s\left(P_0 c^P_0 / P_t c^P_t\right)^{\gamma}$ is the marginal rate of intertemporal substitution of participants between times 0 and t, and σ is a sales subsidy. Firms face demand for their products from two sources: consumers and firms themselves (in order to pay for the adjustment cost); the demand function for the output of firms z is $Y_t(z) = \left(P_t(z)/P_t\right)^{-\varepsilon} Y_t$. Substituting this into the profit function,

[53] Monetary variables are generally not uniquely determined. There always exists an equilibrium of our model where money has no value. If agents anticipate that money is not traded in the future, they do not accept money today and the price of money is 0. The reason for the existence of a non-monetary equilibrium is the same as in the monetary overlapping-generations model of Samuelson (1958). In such a cashless equilibrium, the consumption of N agents is $c^N = \delta_t$ in each period. The consumption of P agents is easily determined; this is akin to the standard cashless New Keynesian model studied in Woodford (2003).

the first-order condition is, after simplifying:

$$
\begin{aligned}
0 = Q_{0,t}\left(\frac{P_t(z)}{P_t}\right)^{-\varepsilon} Y_t &\left[(1+\sigma)(1-\varepsilon)+\varepsilon\frac{W_t}{P_t}\left(\frac{P_t(z)}{P_t}\right)^{-1}\right] \\
&- Q_{0,t}\nu P_t Y_t\left(\frac{P_t(z)}{P_{t-1}(z)}-1\right)\frac{1}{P_{t-1}(z)}+ \\
&+ E_t\left\{Q_{0,t+1}\left[\nu P_{t+1}Y_{t+1}\left(\frac{P_{t+1}(z)}{P_t(z)}-1\right)\frac{P_{t+1}(z)}{P_t(z)^2}\right]\right\}
\end{aligned} \tag{29}
$$

In a symmetric equilibrium all producers make identical choices (including $P_t(z)=P_t$); defining net inflation $\pi_t \equiv (P_t/P_{t-1})-1$, and noticing that $Q_{0,t+1}=Q_{0,t}\beta\left(c_t^P/c_{t+1}^P\right)^{\gamma}(1+\pi_{t+1})^{-1}$, equation (29) becomes:

$$
\begin{aligned}
\pi_t(1+\pi_t) = \beta E_t&\left[\left(\frac{c_t^P}{c_{t+1}^P}\right)^{\gamma}\frac{Y_{t+1}}{Y_t}\pi_{t+1}(1+\pi_{t+1})\right]+ \\
&+\frac{\varepsilon-1}{\nu}\left[\frac{\varepsilon}{\varepsilon-1}\frac{w_t}{A_t}-(1+\sigma)\right]
\end{aligned} \tag{30}
$$

B.2 Loglinearized equilibrium conditions

Table 1 outlines the equilibrium conditions loglinearized around an arbitrary steady state, denoting $\hat{\pi}_t=\ln\frac{1+\pi_t}{1+\pi}$ and $\hat{\imath}_t=\ln\frac{1+i_t}{1+i}$.

Table 1: Summary of loglinearized equilibrium conditions	
Euler P	$\hat{c}_t^P = E_t\hat{c}_{t+1}^P - \gamma^{-1}(\hat{\imath}_t - E_t\hat{\pi}_{t+1})$
Self-insurance/money demand	$\hat{c}_t^P = \frac{\alpha\beta}{1+\pi}E_t\hat{c}_{t+1}^P + \left(1-\frac{\alpha\beta}{1+\pi}\right)E_t\hat{c}_{t+1}^N + \gamma^{-1}E_t\hat{\pi}_{t+1}$
Budget constraint N	$\hat{c}_t^N = \frac{1-\rho}{n(1+\pi)}\frac{\mu}{h}\left(\hat{m}_t^{tot}-\hat{\pi}_t\right) - \frac{\mu}{h}\frac{1-\omega}{1-n}\hat{\tau}_t + \frac{\delta_c}{h}\hat{\delta}_t$
Ec res cons (replace P's BC)	$\hat{c}_t = (1-(1-n)\delta_c)\left(\hat{a}_t+\hat{l}_t-\frac{\nu\pi(1+\pi)}{1-\frac{\nu}{2}\pi^2}\hat{\pi}_t\right)+(1-n)\delta_c\hat{\delta}_t$
Aggregate labor	$\hat{l}_t = \hat{l}_t^P$
Labor supply	$\varphi\hat{l}_t^P = \hat{w}_t - \gamma\hat{c}_t^P$
Aggregate C	$\hat{c}_t^P = \frac{1}{np}\hat{c}_t - \frac{(1-n)h}{np}\hat{c}_t^N$
Transfer	$-\hat{\tau}_t = \hat{x}_t$
Money growth	$\hat{x}_t = \hat{m}_{t+1}^{tot} - \hat{m}_t^{tot} + \hat{\pi}_t$
Phillips curve	$\hat{\pi}_t = \beta E_t\hat{\pi}_{t+1} + \frac{1+\pi}{1+2\pi}\frac{\varepsilon-1}{\nu}(\hat{w}_t-\hat{a}_t)$

B.3 Friedman Rule with flexible prices

First we show that, as in other monetary economies, the price level is indeterminate at the Friedman rule. For $i=0$, the steady state implies $1+\pi=\beta$; $c^P=c^N=c$ and $m^{CB}=\left(1-\beta^{-1}\right)\left(n\tilde{m}^P+(1-n)\tilde{m}^N\right)$. The real allocation is determined by $1=\chi\left(l^P\right)^{\varphi}\left(c^P\right)^{\gamma}$, $c=nl^p+(1-n)\delta$ and

$$
c=\delta+\frac{1}{\beta}\left(1-\rho-(1-(1-\beta)\phi)\left(\frac{1}{\beta}-1\right)n\right)\tilde{m}^P+\frac{1}{\beta}\left(\rho-\beta-(1-(1-\beta)\phi)\left(\frac{1}{\beta}-1\right)(1-n)\right)\tilde{m}^N
$$

There is indeterminacy, even though the real variables c and l^p are uniquely determined as the steady-state first-best values: the monetary variables $\tilde{m}^P$ and $\tilde{m}^N$ must satisfy only one equation, so the real quantity of money is indeterminate.

Second, we show (in the nonlinear model) convergence to the first-best allocation (when $\nu = 0$) if $2 - \alpha - \rho > \beta^{-1} - 1$; the steady-state allocation converges to the first best when $i \longrightarrow 0^+$. In this case, $1 + \pi \longrightarrow \beta^+$. For $0 < k < 1$, define $\hat{l}_t(k)$ as the unique solution to the equation:

$$(n + (1-n)k)\left(\frac{A_t}{\chi}\right)^{\frac{1}{\gamma}}\left(\hat{l}_t^P(k)\right)^{-\frac{\varphi}{\gamma}} = A_t \hat{l}_t^p(k) + (1-n)\delta_t$$

As the left hand side is decreasing and the right hand side increasing in $\hat{l}_t^p$, there always exists a positive solution to the previous equation, whatever $A_t, \delta_t > 0$. Define $\hat{c}_t^P(k)$ as:

$$\hat{c}_t^P(k) \equiv \frac{nA_t l_t^p + (1-n)\delta_t}{n + (1-n)k}$$

For any $k < 1$, we show that this can reach allocations where $c_t^P = \hat{c}_t^P(k)$, $c_t^N = kc_t^p(k)$ and $l_t = \hat{l}_t^P(k)$. When k equals 1, the allocation is exactly the first-best allocation. When k approaches 1, the allocation can be made arbitrarily close to the first-best allocation and the nominal interest rate i_t tends toward 0^+. Take now the model equations from Appendix A, for the case of flexible prices $\nu = 0$ and using the money market equilibrium to substitute for $\tilde{m}_{t+1}^P$. We proceed by guess and verify. At any period, the variables m_t^{CB} and $\tilde{m}_t^P$ are predetermined. As a consequence, assume that the period t money creation m_{t+1}^{CB} is determined by the following law:

$$m_{t+1}^{CB} = k\hat{c}_t^P - \delta - \frac{1}{\beta}\frac{u'\left(\hat{c}_{t-1}^P\right)}{u'\left(\hat{c}_t^P\right)}\frac{1-\rho}{\alpha + (1-\alpha)k^{-\gamma}}\tilde{m}_t^P$$

It is easy to show that the allocation $c_t^P = \hat{c}_t^P$; $c_t^N = k\hat{c}_t^P$; $i_t = \alpha + (1-\alpha)k^{-\gamma}$; $\tau_t = -m_{t+1}^{CB}$ and

$$1 + \pi_t = \beta\left[\alpha + (1-\alpha)k^{-\gamma}\right]\frac{u'\left(\hat{c}_t^P\right)}{u'\left(\hat{c}_{t-1}^P\right)}$$

is an equilibrium of the model, because it satisfies all equations. The equilibrium is locally unique, which we show by standard perturbation methods in a more general case in our companion paper.

B.4 Optimal Ramsey Policy

The constraints of the Ramsey planner are (these are the model equations, with relevant substitutions and using the economy resource constraint instead of the P budget constraint):

$$\begin{aligned}
u'\left(c_t^P\right) &= \beta E_t\left[\frac{\alpha u'\left(c_{t+1}^P\right)+(1-\alpha)\,u'\left(c_{t+1}^N\right)}{1+\pi_{t+1}}\right] \\
nc_t^P+(1-n)\,c_t^N &= \left(1-\frac{\nu}{2}\pi_t^2\right)nA_t l_t^P+(1-n)\,\delta \\
\pi_t\left(1+\pi_t\right) &= \beta E_t\left[\left(\frac{c_t^P}{c_{t+1}^P}\right)^{\gamma}\frac{A_{t+1}l_{t+1}^P}{A_t l_t^P}\pi_{t+1}\left(1+\pi_{t+1}\right)\right]+\frac{\varepsilon-1}{\nu}\left[\frac{\varepsilon}{\varepsilon-1}\frac{\chi\left(l_t^P\right)^{\varphi}\left(c_t^P\right)^{\gamma}}{A_t}-(1+\sigma)\right] \\
c_t^N &= \delta_t+\left(m_{t+1}^{tot}-\frac{m_t^{tot}}{1+\pi_t}\right)+\frac{1}{1+\pi_t}\frac{1-\alpha}{1-n}m_t^{tot} \\
u'\left(c_t^P\right) &= \beta E_t\frac{1+i_t}{1+\pi_{t+1}}u'\left(c_{t+1}^P\right)
\end{aligned}$$

When money is created through helicopter-drop, within-period transfers, only the first three equations above are constraints for the Ramsey planner.[54] Indeed, once c^P and π are known, i follows from the Euler equation for bonds (which hence will not bind as a constraint). Similarly, once the allocation of the consumption of N and inflation have been chosen, the quantity of money delivering it can be recovered through the following equation:

$$c_t^N=\delta_t+m_{t+1}^{tot}-\frac{\alpha-n}{1-n}\frac{m_t^{tot}}{1+\pi_t},$$

where, implicitly, we concentrate only on equilibria where money is used.

The central bank chooses c^P, c^N, l^P, π to maximize the objective defined in the text, subject to the above system of 3 constraints which we denoted in text by Γ_t and write here explicitly for reference:

$$\max_{\{c_t^P,c_t^N,l_t^P,\pi_t\}} E_0\sum_{t=0}^{\infty}\beta^t\{n\left[u\left(c_t^P\right)-\chi\frac{\left(l_t^P\right)^{1+\varphi}}{1+\varphi}\right]+(1-n)\left[u\left(c_t^N\right)-\chi\frac{\delta^{1+\varphi}}{1+\varphi}\right] \tag{31}$$

$$+\omega_{1t}\left[\left(1+\pi_{t+1}\right)\left(c_t^P\right)^{-\gamma}-\beta\alpha\left(c_{t+1}^P\right)^{-\gamma}-\beta\left(1-\alpha\right)\left(c_{t+1}^N\right)^{-\gamma}\right] \tag{32}$$

$$+\omega_{2t}\left[nc_t^P+(1-n)\,c_t^N-\left(1-\frac{\nu}{2}\pi_t^2\right)nA_t l_t^P-(1-n)\,\delta\right]$$

$$+\omega_{3t}\left[\pi_t\left(1+\pi_t\right)-\beta E_t\left[\left(\frac{c_t^P}{c_{t+1}^P}\right)^{\gamma}\frac{Y_{t+1}}{Y_t}\pi_{t+1}\left(1+\pi_{t+1}\right)\right]-\frac{\varepsilon}{\nu}\left(\frac{\chi\left(l_t^P\right)^{\varphi}\left(c_t^P\right)^{\gamma}}{A_t}+\Phi-1\right)\right]\}$$

The solution is a system of 4 first-order conditions and 3 constraints, for 4 variables and 3 co-states (the Lagrange multipliers on the constraints). The first-order conditions of the Ramsey problem are for each variable respectively:

[54] Whereas with open-market operations, *all* of the above equations are constraints; we analyze this case and provide a welfare comparison in the companion paper.

$$
\begin{aligned}
0 &= n\left(c_t^P\right)^{-\gamma} - \gamma\omega_{1t}\left(1+\pi_{t+1}\right)\left(c_t^P\right)^{-\gamma-1} + \gamma\omega_{1t-1}\alpha\left(c_t^P\right)^{-\gamma-1} \\
&\quad + \omega_{2t}n - \frac{\varepsilon}{\nu}\gamma\omega_{3t}\frac{\chi\left(l_t^P\right)^{\varphi}\left(c_t^P\right)^{\gamma-1}}{A_t} \\
0 &= (1-n)\left(c_t^N\right)^{-\gamma} + \gamma\omega_{1t-1}(1-\alpha)\left(c_t^N\right)^{-\gamma-1} + \omega_{2t}(1-n) \\
0 &= -n\chi\left(l_t^P\right)^{\varphi} - \omega_{2t}\left(1-\frac{\nu}{2}\pi_t^2\right)nA_t - \omega_{3t}\varphi\frac{\varepsilon}{\nu}\frac{\chi\left(l_t^P\right)^{\varphi-1}\left(c_t^P\right)^{\gamma}}{A_t} \\
0 &= \omega_{1t-1}\beta^{-1}\left(c_{t-1}^P\right)^{-\gamma} + \omega_{2t}\nu\pi_t n A_t l_t^P + \left(\omega_{3t}-\omega_{3t-1}\right)\left(1+2\pi_t\right)
\end{aligned}
$$

plus the three constraints with complementary slackness.

A steady-state of the Ramsey problem is defined by:

$$
\begin{aligned}
\omega_1 &= 0 \text{ or } \left(c^P\right)^{-\gamma} = \frac{1-\alpha}{\frac{1+\pi}{\beta}-\alpha}\left(c^N\right)^{-\gamma} \\
\omega_2 &= 0 \text{ or } nc_t^P + (1-n)c_t^N = \left(1-\frac{\nu}{2}\pi^2\right)nl^P + (1-n)\delta \\
\omega_3 &= 0 \text{ or } \pi(1+\pi) = \frac{\varepsilon}{\nu(1-\beta)}\left[\chi\left(l^P\right)^{\varphi}\left(c^P\right)^{\gamma} - (1-\Phi)\right]
\end{aligned}
$$

$$
\begin{aligned}
0 &= n\left(c^P\right)^{-\gamma} - \gamma\omega_1(1+\pi-\alpha)\left(c^P\right)^{-\gamma-1} + \omega_2 n - \frac{\varepsilon}{\nu}\gamma\omega_3\chi\left(l^P\right)^{\varphi}\left(c^P\right)^{\gamma-1} \\
0 &= (1-n)\left(c^N\right)^{-\gamma} + \gamma\omega_1(1-\alpha)\left(c^N\right)^{-\gamma-1} + \omega_2(1-n) \\
0 &= -n\chi\left(l^P\right)^{\varphi} - \omega_2\left(1-\frac{\nu}{2}\pi^2\right)n - \omega_3\varphi\frac{\varepsilon}{\nu}\chi\left(l^P\right)^{\varphi-1}\left(c^P\right)^{\gamma} \\
0 &= \omega_1\beta^{-1}\left(c^P\right)^{-\gamma} + \omega_2\nu\pi n l^P
\end{aligned}
$$

The Proof of the proposition pertaining to optimal long-run inflation is now immediate. With flexible prices $\nu = 0$ and optimal subsidy, the only solution to the above system of equations is perfect insurance through the Friedman Rule:

$$
\frac{1+\pi}{\beta} = 1 \rightarrow c^P = c^N = c
$$

With sticky prices and inelastic labor $\varphi \rightarrow \infty$, the intratemporal optimality condition disappears from the set of constraints, labor is fixed, and it can be easily shown that inflation tends to zero ($\pi = 0$ solves the above system).

Computing the welfare cost To calculate the welfare cost of inflation, we proceed in the standard way pioneered by Lucas (1987). Denote with an upper-script SS the allocation for the inflation rate π^{SS} and no shock. We denote the welfare of an economy where inflation is, say π^*, as V^*. We then compute the proportional decrease in consumption for all households in the economy with inflation rate π^* to equalize the two welfare measures. Formally we compute Δ^W to have:

$$E_0 V^* = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ n \left[u\left(\left(1-\Delta^W\right) c_t^{PSS} \right) - \chi \frac{\left(l_t^{PSS}\right)^{1+\varphi}}{1+\varphi} \right] + (1-n) \left[u\left(\left(1-\Delta^W\right) c_t^{NSS} \right) - \chi \frac{\delta^{1+\varphi}}{1+\varphi} \right] \right\}$$

B.5 Open Market Operations and Ramsey Problem

The general form of money issuance under OM is captured by a "fiscal rule" that insures that the intertemporal government equation holds as a constraint for any price level, for example

$$\tau_t = \phi_b \frac{1+i_{t-1}}{1+\pi_t} b_t - \phi_x x_t. \tag{33}$$

This nests the two extreme cases considered in text HD with $\phi_x = 1$ and OM with one-period repayment. More generally, the coefficient ϕ_b captures how fast money is issued in the economy once the OM operation took place. We checked numerically that the value of ϕ_b delivering the highest welfare is 1, i.e. one-period repayment, the intuition being the one provided in text. Rigorously, the Ramsey problem becomes with OM one of choosing the allocation $\{c_t^P, c_t^N, l_t^P, \pi_t, m_{t+1}^{tot}, i_t\}$ to maximize a Lagrangian similar to the one in (23) but :

$$E_0 \sum_{t=0}^{\infty} \beta^t \left\{ n \left[u\left(c_t^P\right) - \chi \frac{\left(l_t^P\right)^{1+\varphi}}{1+\varphi} \right] + (1-n) \left[u\left(c_t^N\right) - \chi \frac{\delta^{1+\varphi}}{1+\varphi} \right] + \omega_t \Gamma_t \right\} \tag{34}$$

where ω_t is the vector of costate, Lagrange multipliers, one for each constraint in Γ_t—which include, other than the constraints with HD, the additional:

$$\begin{aligned}
c_t^N &= \delta - \tau_t + \frac{1-\alpha}{1-n} \frac{m_t^{tot}}{1+\pi_t} \\
b_{t+1} + \tau_t + x_t &= \frac{1+i_{t-1}}{1+\pi_t} b_t \\
x_t &= m_{t+1} - \frac{1}{1+\pi_t} m_t \\
\tau_t &= \phi_b \frac{1+i_{t-1}}{1+\pi_t} b_t - \phi_x x_t \\
u'\left(c_t^P\right) &= \beta E_t \frac{1+i_t}{1+\pi_{t+1}} u'\left(c_{t+1}^P\right).
\end{aligned}$$

B.6 Welfare function and second-order approximation

The second-order approximation technique used is described in detail in Woodford (2003, Chapter 6), Benigno and Woodford (2005, 2012), and Bilbiie (2008) for the case of two agents. A second-order approximation of P agents' utility delivers:

$$U_t^P - U^P = U_c^P c^P \left(\hat{c}_t^P + \frac{1-\gamma}{2}\left(\hat{c}_t^P\right)^2\right) + U_L^P l^P \left(\hat{l}_t^P + \frac{1+\varphi}{2}\left(\hat{l}_t^P\right)^2\right)$$
$$= \left(c^P\right)^{1-\gamma}\left(\hat{c}_t^P + \frac{1-\gamma}{2}\left(\hat{c}_t^P\right)^2 - \frac{l^P}{c^P}\left(\hat{l}_t^P + \frac{1+\varphi}{2}\left(\hat{l}_t^P\right)^2\right)\right)$$

where the second equality used SS under subsidy $w = 1$, $U_L^P = -U_C^P$

For N agents:

$$U_t^N - U^N = \left(c^N\right)^{1-\gamma}\left(\hat{c}_t^N + \frac{1-\gamma}{2}\left(\hat{c}_t^N\right)^2\right)$$

Aggregating:

$$U_t - U = \left(c^P\right)^{-\gamma}\left(nc^P\hat{c}_t^P - nl^P\hat{l}_t^P + \left(\frac{c^N}{c^P}\right)^{-\gamma}(1-n)\,c^N\hat{c}_t^N\right)$$
$$+ n\left(c^P\right)^{1-\gamma}\left(\frac{1-\gamma}{2}\left(\hat{c}_t^P\right)^2 - \frac{l^P}{c^P}\frac{1+\varphi}{2}\left(\hat{l}_t^P\right)^2\right) + (1-n)\left(c^N\right)^{1-\gamma}\frac{1-\gamma}{2}\left(\hat{c}_t^N\right)^2$$

Take **the linear term** first:

$$\left(c^P\right)^{-\gamma}\left(c\hat{c}_t - l\hat{l}_t + \left(\left(\frac{c^N}{c^P}\right)^{-\gamma} - 1\right)(1-n)\,c^N\hat{c}_t^N\right)$$

The economy resource constraint to second order is (denote $\Delta_t = 1 - \frac{\nu}{2}\pi_t^2$):

$$\hat{c}_t = (1-(1-n)\,\delta_c)\left(a_t + \hat{l}_t + \hat{\Delta}_t\right) \text{ where}$$
$$\hat{\Delta}_t = -\frac{\nu\pi}{1-\frac{\nu}{2}\pi^2}\pi_t - \frac{1}{2}\frac{\nu}{1-\frac{\nu}{2}\pi^2}\pi_t^2$$

Note that under zero inflation the linear term disappears. The squared term captures the welfare cost of inflation.

The linear term becomes hence:

$$\left(c^P\right)^{-\gamma} c\left(\hat{\Delta}_t + a_t + (q^\gamma - 1)(1-n)\,h\hat{c}_t^N\right)$$

where we recall $q^\gamma = 1 + \frac{(1+\pi)\beta^{-1}-1}{1-\alpha}$; at the Friedman rule this is unity, and the linear term drops out. Otherwise, it is larger than 1 and the linear term has a positive coefficient – increasing the consumption of N closes the inequality gap, providing a first-order benefit.

The quadratic term is (ignoring price dispersion because in quadratic terms it becomes third or fourth order):

$$\frac{\left(c^P\right)^{-\gamma}c}{2}\left((1-\gamma)\left(np\left(\hat{c}_t^P\right)^2 + (1-n)\,hq^\gamma\left(\hat{c}_t^N\right)^2\right) - \frac{1+\varphi}{1-(1-n)\,\delta_c}\hat{c}_t^2\right)$$

Thus the loss function becomes, rearranging and ignoring terms independent of policy:

$$
\begin{aligned}
\mathcal{L} &= \left(c^P\right)^{-\gamma} c \begin{pmatrix} -\hat{\Delta}_t - \left(q^\gamma - 1\right)(1-n) h \hat{c}_t^N \\ +\frac{\gamma-1}{2}\left(np\left(\hat{c}_t^P\right)^2 + (1-n) h q^\gamma \left(\hat{c}_t^N\right)^2\right) + \frac{1+\varphi}{2}\frac{1}{1-(1-n)\delta_c}\hat{c}_t^2 \end{pmatrix} \\
&= \left(c^P\right)^{-\gamma} c \begin{pmatrix} -\hat{\Delta}_t - \left(q^\gamma - 1\right)(1-n) h \left(\hat{c}_t^N + \frac{1-\gamma}{2}\left(\hat{c}_t^N\right)^2\right) \\ +\frac{\gamma-1}{2} np (1-n) h \hat{q}_t^2 + \left(\frac{1+\varphi}{2}\frac{1}{1-(1-n)\delta_c} + \frac{\gamma-1}{2}\right)\hat{c}_t^2 \end{pmatrix}
\end{aligned}
$$

Adding and subtracting the steady-state inflation constant and ignoring all terms independent of policy, we obtain the loss function

$$
\frac{1}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \Lambda_\pi \tilde{\pi}_t^2 + \Lambda_c \hat{c}_t^2 + \Lambda_q \hat{q}_t^2 - \Lambda \left[\hat{c}_t^N + \frac{1-\gamma}{2}\left(\hat{c}_t^N\right)^2\right] \right\}
$$

where the optimal relative weights are:

$$
\begin{gathered}
\Lambda_\pi = \frac{\nu}{1-\frac{\nu}{2}\pi^2};\ \Lambda_c = \gamma - 1 + \frac{1+\varphi}{1-(1-n)\delta_c} \\
\Lambda = 2(1-n) h \left(q^\gamma - 1\right);\ \Lambda_q = (\gamma-1) np (1-n) h.
\end{gathered}
$$

This nests the $\gamma = 1$ case presented in text. The new relevant terms are $\Lambda_q \hat{q}_t^2$, which captures the welfare cost of the *volatility of inequality* (naturally, this drops out if all agents are identical, $n = 0$ or $n = 1$); and the last term that is just a risk correction.

B.7 "New Keynesian cross" model with endogenous N income

We refer the interested reader to Bilbiie (2008, 2017) for more details; here, we outline the equations that change relative to our baseline model with exogenous income of N. In this model, N agents work and receive the same market real wage as P agents. For simplicity and to isolate our channel of insurance through liquidity (rather than different hours worked), we further assume that hours worked are pooled by a union and labor is demand-determined—thus following Galí et al, 2007 (see also Ascari et al, 2016 for a setup with a union and sticky wages). The equilibrium implication is that hours worked by each agent are identical (and hence equal to the aggregate, since total mass is 1):

$$
l_t^N = l_t^P = l_t,
$$

and determined by an aggregate hours schedule that we specify as:

$$
w_t = \chi \left(l_t\right)^\varphi \left(c_t\right)^\gamma .
$$

Since l_t now replaces δ in the expression for utility of N, aggregate welfare will be $nu\left(c_t^P\right) + (1-n) u\left(c_t^N\right) - \frac{\chi}{1+\varphi} l_t^{1+\varphi}$. The budget constraint for N agents is thus (having imposed asset market clearing and equilibrium, in particular for money)

$$c_t^N = w_t l_t - \tau_t + \frac{1-\alpha}{1-n}\frac{m_t^{tot}}{1+\pi_t} - T_t^N$$

where τ_t are as before monetary transfers and T_t^N are taxes used to pay for the sales subsidy given to firms. As n the benchmark, firms' profits are

$$d_t = \left(1+\sigma - \frac{w_t}{A_t} - \frac{\nu}{2}\pi_t^2\right) Y_t$$

where total production is now $Y_t = A_t l_t$. Assuming that the entire subsidy is financed via taxes levied on consumers every period, we specify a process for the distribution of taxes as:

$$T_t^N = \frac{\theta}{1-n}\sigma Y_t,$$

where θ is thus the share of the subsidy to firms levied on N households. The economy resource constraint is $c_t = \left(1 - \frac{\nu}{2}\pi_t^2\right) Y_t$.

Assuming and optimal subsidy $\sigma = (\varepsilon - 1)^{-1}$ and $\theta = 0$ (our implicit assumption in the baseline model with exogenous N income) is no longer possible, because the assumption of endogenous labor would then imply perfect insurance in steady-state ($w = 1 \rightarrow c^N = l = c = c^P$). This, in turn, implies a cashless equilibrium as there is no reason to hold money. We therefore assume for this part that there is no implicit fiscal redistribution associated with the supply-side policy: $\theta = 1 - n$ implying that taxes are uniform $T_t^N = T_t^P = T_t = \sigma Y_t$. We contrast this with the "distorted SS" case where $\sigma = 0$.

As for the quadratic approximation to the aggregate welfare function, the main difference is as follows. Utility of P is unchanged, but utility of N becomes

$$U_t^N - U^N = \left(c^N\right)^{1-\gamma}\left(c_t^N + \frac{1-\gamma}{2}\left(c_t^N\right)^2 - \frac{l^N}{c^N}\left(l_t^N + \frac{1+\varphi}{2}\left(l_t^N\right)^2\right)\right)$$

Aggregating as before and imposing the symmetric labor choice in steady state it can be easily shown that the linear term boils down to

$$\left(c^P\right)^{-\gamma} c\left(-\Delta_t + \left(q^\gamma - 1\right)(1-n)\left(hc_t^N - l_t^N\right)\right)$$

which for our model with equal hours across agents is $\left(c^P\right)^{-\gamma} c\left(-\Delta_t + \left(q^\gamma - 1\right)(1-n)\left(hc_t^N - l_t\right)\right)$. The quadratic term is different from the previous only in that the weight on consumption/output volatility is $\Lambda_c = \varphi + \gamma$. (this part is identical to Bilbiie, 2008—but that paper focuses on a steady-state with $q = 1$). Notice that the expression for q that governs the liquidity-insurance motive (or distortion) is as before—because the self-insurance equation still holds—but in order to determine optimal long-run inflation and the consumption share of N we have now (because $wl = c$):

$$\begin{aligned} h &= \frac{c^N}{c} = \frac{wl}{c} + \left(\pi + \frac{1-\alpha}{1-n}\right)\frac{1}{1+\pi}\frac{m}{c} - \frac{T^N}{c} \\ &= \frac{(1+\sigma)\frac{\varepsilon-1}{\varepsilon} - \frac{\theta}{1-n}\sigma}{1-\frac{\nu}{2}\pi^2} + \left(\pi + \frac{1-\alpha}{1-n}\right)\frac{\mu}{1+\pi} \end{aligned}$$

Aggregate IS curve. Without loss of generality, consider a steady state with zero inflation $\pi = 0$, optimal subsidy $\sigma = (\varepsilon - 1)^{-1}$ and $\theta = 1 - n$, then we have $wl = c$ and $h = 1 - \sigma + \frac{1-\alpha}{1-n}\mu$, and a loglinear approximation around it:

$$h\hat{c}_t^N = \hat{w}_t + \hat{l}_t + \frac{1-\alpha}{1-n}\mu\left(\hat{m}_t^{tot} - \hat{\pi}_t\right) + \mu\hat{x}_t.$$

Using labor supply $\varphi\hat{l}_t = \hat{w}_t - \gamma\hat{c}_t$ and production function+resource constraint $\hat{l}_t = \hat{c}_t$ which give $\hat{w}_t = (\varphi + \gamma)\hat{c}_t$, we rewrite as:

$$h\hat{c}_t^N = (1 + \varphi + \gamma)\hat{c}_t + \frac{1-\alpha}{1-n}\mu\left(\hat{m}_t^{tot} - \hat{\pi}_t\right) + \mu\hat{x}_t.$$

We replace this and the relevant equations in the Euler equation of participants (12) to obtain:

$$\begin{aligned}\hat{c}_t &= E_t\hat{c}_{t+1} - \frac{\gamma^{-1}np}{1-(1-n)(1+\varphi+\gamma)}\left(\hat{\imath}_t - E_t\hat{\pi}_{t+1}\right) \\ &\quad + \frac{\mu(1-\alpha)}{1-(1-n)(1+\varphi+\gamma)}E_t\hat{\pi}_{t+1} + \frac{\mu(\alpha-n)}{1-(1-n)(1+\varphi+\gamma)}\hat{x}_t - \frac{\mu(1-n)}{1-(1-n)(1+\varphi+\gamma)}E_t\hat{x}_{t+1}\end{aligned}$$

The amplification of interest rate changes follows the same "New Keynesian cross" logic unveiled in Bilbiie (2008); the same mechanism also delivers amplification of liquidity changes, and of expected inflation—since they both trigger increases in aggregate demand, shifts in labor demand, increases in the wage, and thus further demand increases for N who consume this wage. One key difference here with respect to Bilbiie (2017) is that this amplification does not affect the "discounting" (the coefficient in front of expected consumption)—see McKay, Nakamura and Steinsson (2015, 2016) for the original point of discounting in the aggregate Euler equation under incomplete markets; see also Werning (2015) for a more general framework that delivers amplification under certain conditions on cyclicality of income risk and liquidity. The reason why no discounting occurs here is that bonds are illiquid—whereas in those papers and in Bilbiie, 2017 they are liquid.