A HANK Model with Monetary Search Frictions

Joel Marbet*
CEMFI

Job Market Paper

Please click here for the latest version

January 14, 2023

Abstract

I develop a framework for studying media of exchange within Heterogenous Agent New Keynesian (HANK) models. To this end, I extend an otherwise standard HANK model with search-theoretic monetary frictions as in Lagos and Wright (2005). I show that the medium of exchange role of money breaks monetary super-neutrality in HANK, meaning that changes in the central bank's inflation target affect real variables and the wealth distribution in the long run. The extent of non-neutrality can be substantial, with aggregate consumption declining by 0.74% after an increase in the central bank's inflation target from 0% to 5%. I then apply the framework to study how heterogeneity in the dependence on non-interest-bearing payment instruments shapes the welfare costs of inflation across wealth and income distributions. I show quantitatively that the welfare costs of inflation are about 8% higher for the wealth- and income-poor households in the economy. The result stems from the fact that poor households, in line with microdata, depend more on non-interest-bearing payment instruments, such as cash.

Keywords: Inflation, CBDC, Heterogeneous Agent, New Keynesian, HANK, Monetary Search Theory.
JEL Classification Codes: E31, E41, E42, E52, E58.

*CEMFI, Calle Casado del Alisal 5, 28014 Madrid, Spain. Email: joel.marbet@cemfi.edu.es
I am very grateful to my advisor Dante Amengual for his guidance and advice on this research project. I also would like to thank Samuel Bentolila, Antonio Diez de los Rios, Sebastián Fanelli, Galo Nuño, Josep Pijoan-Mas, Omar Rachedi, Morten O. Ravn, Rafael Repullo, Hernán Ruffo, Pedro Silos, Javier Suarez, and all participants of workshops at CEMFI for valuable comments and suggestions during various stages of this project.

1 Introduction

Monetary models in the New Keynesian tradition have become a model of choice for analyzing the short-term effects of monetary policy. More recent vintages of these models have also been used to highlight the importance of household heterogeneity in the monetary transmission mechanism (e.g., McKay et al., 2016; Gornemann et al., 2016; Kaplan et al., 2018; Luetticke, 2019). However, typically New Keynesian models abstract from the medium of exchange role of money, making them less well-equipped for analyzing recent innovations in payments and money, such as the rise of novel forms of private money or the potential introduction of central bank digital currencies (CBDC).[1] In contrast, monetary models in the New Monetarist tradition have strongly emphasized microfoundations for money as a medium of exchange. These models typically do not consider the short-run dynamics resulting from shocks hitting the economy. Instead, they focus on long-run steady-state comparisons.[2]

In this paper, I develop a framework that allows for the study of a wealth of questions at the intersection of inequality, innovations in payments, and monetary policy in both the short and the long run. To this end, I extend an otherwise standard Heterogenous Agent New Keynesian (HANK) model with search-theoretic monetary frictions as in Lagos and Wright (2005).[3] The resulting framework features microfoundations for exchange based on monetary search theory, the consumption-saving behavior inherent in an incomplete markets model, and a role for monetary policy due to both nominal rigidities and monetary frictions.

The model embeds two sectors: In every period, some economic activity takes place in a decentralized market (DM) and some in a centralized market (CM). In the decentralized market, a fraction of households prefers to produce, while another fraction prefers to consume. Households can engage in bilateral trade with households of the other type within randomly matched meetings. Since the meetings are anonymous, a generally accepted medium of exchange is required, which introduces a demand for money into the model. On the other hand, the centralized market does not feature such trading frictions. It resembles the setup in standard macro models meaning that households take prices as given and trade in a Walrasian market where there is no explicit need for a medium of exchange. Production in the centralized market is carried out by firms, some of which are subject to price adjustment costs, introducing nominal rigidities into the model. Furthermore,

[1]While it is feasible to analyze such issues within New Keynesian models (see, e.g., Ferrari et al., 2020; Uhlig and Xie, 2021; Gross and Schiller, 2021; Abad et al., 2022), it usually requires a more reduced form approach through a money-in-utility or cash-in-advance type of specification. This might limit the range of questions that can be answered in such a framework and is more likely to be subject to the Lucas critique.

[2]Exceptions include Aruoba (2011) and Chiu and Molico (2021) who study the short-run effects of monetary policy within a New Monetarist model.

[3]To more be precise, I draw from a large literature that followed Lagos and Wright (2005) in embedding monetary search-theory into macroeconomic models including, but not limited to, Aruoba et al. (2011), Aruoba and Schorfheide (2011), Chiu and Molico (2011), Chiu et al. (2021), and Keister and Sanches (2022).

households are subject to uninsurable idiosyncratic labor income risk realized during the centralized market. At the core, this is still a prototypical HANK model, but the addition of the decentralized market introduces a microfounded demand for a medium of exchange, which in the model can either be cash, meaning banknotes and coins, or deposits, meaning debit/credit card payments.

Since the terms of trade in the decentralized market are determined through bargaining, the model poses a major computational challenge. Household decisions depend not only on their own asset holdings and income types but also on the asset holdings and income type of the counterparty they meet in the decentralized market. These additional state variables increase the problem's dimensionality and make the model more difficult to solve using current methods. I show how to reframe the household problem to make it computationally more tractable without imposing additional assumptions (e.g., bounded rationality). The computational approach I propose allows solving for both the steady state and transitional dynamics. Furthermore, it is feasible to introduce additional liquid assets in the decentralized market with only a small increase in computational complexity, meaning that the framework provides an excellent basis for extensions interested in analyzing aggregate and distributional implications of innovations in payments and money, such as central bank digital currencies.

The model is calibrated to be consistent with several aggregate moments for the euro area, such as currency in circulation, deposit holdings of households, the aggregate capital stock, and banks' reserve holdings. Furthermore, the probability of acceptance of different payment methods in the decentralized market is calibrated to match the cash and debit/credit card acceptances rates at points of sale (POS) based on data from the study on the payment attitudes of consumers in the euro area (SPACE, European Central Bank, 2020). The model can then endogenously capture that poorer households hold a larger fraction of their liquid assets in cash consistent with microdata from the Encuesta Financiera de las Familias (EFF). Through the lens of the model, this is because households with very few resources prefer to hold cash to meet their liquidity needs – even if it means they have to forgo the interest on deposits – since cash is more likely to be accepted by sellers in decentralized market meetings and therefore, implicitly provides a higher marginal utility.

I show that the medium of exchange role of money breaks the monetary super-neutrality in the model developed in this paper, which stands in contrast to the standard HANK model nested within the framework.[4] This means that changes in the central bank's inflation target affect real variables and the wealth distribution in the long run. At the root of this non-neutrality lies that changes in the inflation target directly affect the real return on cash since its nominal return is fixed at 0%. In contrast, the real return on deposits is tied to the return on capital through the banking system and only responds to changes in the inflation target to the extent that the supply of capital and labor change in the aggregate.

[4] If the probability of decentralized market meetings is zero, the model collapses to a standard HANK model that exhibits super-neutrality.

Because of this, an increase in the inflation target increases the wedge between the real returns on cash and deposits, which then induces changes in the consumption-savings behavior of households. In particular, households economize on cash, and due to the fall in the effective return on savings, they also tend to accumulate less wealth, meaning that their deposit holdings and consumption decrease. Since the monetary super-neutrality is tied to how changes in the inflation target affect the return differential, it can be restored if the central bank replaces cash with a CBDC that bears an interest rate that adjusts one-for-one with inflation. This not only restores super-neutrality but also eliminates the welfare cost of inflation.

The impact of a change in the inflation target is not uniform across wealth and income distributions. Poor, low-income households face a more significant decline in the effective return on their savings since they hold relatively more cash. As a result, they experience a more substantial decrease in consumption and tend to accumulate less wealth. For rich, high-income households, the reduction in consumption is less pronounced since the effective return on their savings falls by a smaller amount. In equilibrium, they even tend to accumulate slightly more wealth. Overall, these differences in wealth accumulation between income-poor and income-rich households result in a reduction in wealth accumulation in the aggregate and a slightly higher dispersion in the wealth distribution.

A quantitative analysis of the model shows that the extent of the non-neutrality can be substantial. For example, an increase in the central bank's inflation target from 0% to 5% implies a reduction in aggregate consumption by 0.74%. If the inflation target would be increased to 10%, we would observe an even higher reduction in aggregate consumption of 1.36%. An increase in government revenue, mainly due to seignorage revenue from the implicit inflation tax on cash holdings under higher inflation targets, implies that aggregate output falls only 6bp at 5% inflation. Nevertheless, the model produces a long-run tradeoff between output and inflation similar to the Generalized New Keynesian model (Ascari, 2004; Ascari and Sbordone, 2014) but using a different mechanism. The non-neutrality also affects the level of real interest rates in the long run: an increase from 0% to 5% inflation, for example, increases the real interest rate on bonds by 3bp. Related to this, a recent literature on the deflationary bias in New Keynesian models finds a similar positive relationship between real interest rates and the inflation target of the central bank (e.g. Bianchi et al., 2021; Fernández-Villaverde et al., 2021). However, in those models, it is the result of an interaction between aggregate uncertainty and the zero lower bound (ZLB) on nominal interest rates, while in the model in this paper, which has neither of these features, the super-non-neutrality is a result of the medium of exchange role of money and the fact that cash has a constant nominal return of 0%.

The medium of exchange role of money also has implications for the transmission of monetary policy shocks. While the response of interest rates and inflation are quite similar, aggregate consumption increases less in response to an expansionary monetary policy shock. This results from a transactional motive for savings and an increase in idiosyncratic risk due to the liquidity shocks in the decentralized market, which both lead to a rise in savings relative to a HANK model without the decentralized market. Since households with high marginal propensities to consume (MPC) are the ones at the borrowing constraint, as in standard heterogeneous agent models, the overall increase in savings by households leads to a shift in the distribution to the right and a reduction in the average MPC, leading to a more subdued reaction of consumption in response to monetary policy shocks.

I then apply this framework to study how heterogeneity in the dependence on non-interest-bearing payment instruments shapes the welfare costs of inflation across wealth and income distributions. The median household would be willing to give up about 0.5% of consumption every period to avoid having to live in an economy with 5% inflation. Furthermore, the welfare costs of 10% inflation are around 1% in consumption equivalent terms for the median household. While these welfare costs are roughly in line with previous estimates in the literature (e.g., Lucas, 2000), I find that they are more tilted towards wealth- and income-poor households. In particular, the welfare costs of inflation are about 8% higher for low-income households at the 10th wealth percentile than for high-income households at the 90th wealth percentile. The result stems from the fact that poor households, in line with microdata, depend more on non-interest-bearing payment instruments, such as cash.

This paper is related to several strands of the literature on monetary economics and heterogeneous agent models. In particular, I build on the heterogenous agent models in the spirit of Bewley (1983), İmrohoroğlu (1992), Huggett (1993), and Aiyagari (1994), and a long line of models thereafter, which introduce nominal rigidities into this class of models (McKay et al., 2016; Gornemann et al., 2016; Kaplan et al., 2018; Luetticke, 2019). In contrast to this literature, I explicitly introduce a medium of exchange role of money in an incomplete markets model with nominal rigidities.

Furthermore, the paper is related to a growing literature on New Monetarist models that use search theory to model monetary exchange following Kiyotaki and Wright (1989). Lagos and Wright (2005) (LW) paved the way for incorporating monetary search theory into more traditional macroeconomic models. Aruoba and Schorfheide (2011) showed that the LW approach can be extended with nominal rigidities in a representative agent context and that both monetary frictions and nominal rigidities are equally important quantitatively. I take advantage of these contributions but extend them to incorporate uninsurable idiosyncratic labor income risk in line with the heterogeneous agent literature described above. While the New Monetarist literature typically focuses on analytical solutions with degenerate distributions, more recently Molico (2006), Chiu and Molico (2010, 2011, 2021), Rocheteau et al. (2021), and Bethune and Rocheteau (2022), among others, have studied household heterogeneity within New Monetarist models. However, these papers focus on money injections through transfers to households. In contrast, monetary transmission in the model developed in this paper works through the interest rate policy of the central bank following the New Keynesian literature.

Finally, this paper is related to a large literature on the welfare cost of inflation which is one of the classical issues in monetary economics. Lucas (2000) provides a survey of this literature. To analyze the welfare costs, several papers have employed heterogeneous agent models such as İmrohoroğlu (1992) who finds that if money is used to smooth out uninsurable income risk the welfare costs of inflation can be much higher. Erosa and Ventura (2002) show that inflation acts as a regressive consumption tax in a model with a reduced-form technology for credit services. In their setup, the welfare costs can be substantially higher for low-income individuals if households can evade inflation by paying for credit services that exhibit economies of scale. Similarly, in my model, the welfare costs for the poor are higher than for the rich along both the income and the wealth distribution due to a stronger dependence on consumption purchases with cash. However, the stronger dependence on cash is the result of differences in the consumption baskets of rich and poor households and not due to differences in transaction technology.

The rest of the paper is structured as follows: Section 2 discusses the model setup. Section 3 describes the calibration of the model. Section 4 discusses the computational challenge of solving a HANK model with search-theoretic monetary frictions. Section 5 analyzes the quantitative properties of the model and discusses the aggregates and distributional implications. Section 6 concludes.

2 Model

I extend an otherwise standard Heterogenous Agent New Keynesian (HANK) model with search-theoretic monetary frictions as in Lagos and Wright (2005). The monetary frictions introduce a transaction motive for holding assets which has important aggregate and distributional consequences.

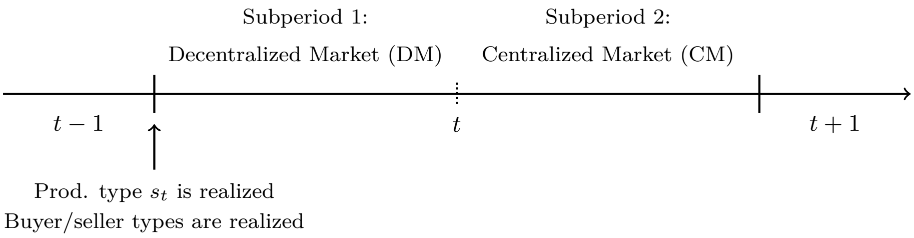
2.1 Households

There is a continuum of households with mass one. Each period is divided into two subperiods. In the first subperiod, agents meet bilaterally in a decentralized market (DM). In the second subperiod, they meet in a centralized market (CM). Figure 1 shows the timing of the model economy.[5]

Households are ex-ante identical but face two sources of uninsurable idiosyncratic risk. First, they are subject to labor earnings risk s_t, which follows a Markov chain with N states and transition matrix Ω. Labor earnings risk is a standard source of idiosyncratic risk in the heterogeneous agent literature (e.g. Huggett, 1993; Aiyagari, 1994). Second, the model features additional idiosyncratic risk because households can become either buyers

[5]Some notes on the notation: Variables determined in the decentralized market are denoted with hats, e.g. $\hat{x}_t$. Moreover, the time subscript always refers to the period in which the variables have been determined (e.g., a household decision taken in time t carries subscript t). For example, $\hat{x}_t$ has been determined in the decentralized market in period t.

Figure 1: Model Timeline



or sellers at the beginning of the decentralized market, where they are randomly matched with other households in bilateral meetings.[6]

2.1.1 Centralized Market

The household's problem in the centralized market corresponds to a standard consumption-savings decision, which is at the core of most macroeconomic models. Let $\hat{a}_t$ denote the total wealth (incl. interest rate payments) that households bring from the DM into the CM

$$\hat{a}_t = \frac{R^M_{t-1}}{\pi_t}\hat{m}_t + \frac{R^D_{t-1}}{\pi_t}\hat{d}_t$$

where $\hat{m}_t$ and $\hat{d}_t$ are cash and deposits brought in from the DM, R^M_{t-1} is the gross nominal return on cash, R^D_{t-1} is the gross nominal return on deposits, and $\pi_t = \frac{P_t}{P_{t-1}}$ is the gross inflation rate with P_t being the price level in the centralized market.

Households maximize expected discounted utility and can save in two types of assets, cash m_t and deposits d_t. Cash is dominated in return but can be used to transact in some decentralized market meetings in the subsequent period, where deposits are not accepted. The problem can be written as

$$\begin{aligned} V^{CM}_t(\hat{a}_t, s_t) &= \max_{c_t, m_t, d_t} U(c_t, h_t) + \beta \mathrm{E}_t \left[V^{DM}_{t+1}(m_t, d_t, s_{t+1})\right] \\ \text{s.t.} \quad c_t + m_t + d_t &= (1-\tau_h) w_t s_t h_t + \hat{a}_t \,, \\ m_t, d_t &\geq 0 \,, \end{aligned}$$

where $\beta \in (0,1)$ is the discount factor, c_t is consumption during the CM, h_t is labor supply during the CM, $\hat{m}_t$ and $\hat{d}_t$ are cash and deposits brought in from the DM, m_t and d_t are the choices of cash and deposits for the next DM, w_t is the real wage, and τ_h is the labor income tax rate. The labor supply decision is delegated to a labor union, which allows for the introduction of wage rigidities.

[6] Note that part of the New Monetarist literature assumes that buyer and seller types are fixed (e.g. Chiu et al., 2021). However, this assumption would imply that sellers only invest in the asset with the highest return since they don't need to hold balances for transactions in the decentralized market. The assumption that buyer/seller types are determined at the beginning of each DM ensures that all households benefit from holding transaction balances at the end of the CM in the case that they become a buyer in the DM.

Preferences are additively separable

$$U(c_t, h_t) = \Gamma\left(U(c_t) - \Upsilon(h_t)\right) = \Gamma\left(\frac{c_t^{1-\sigma}}{1-\sigma} - \chi\frac{h_t^{1+\nu}}{1+\nu}\right),$$

where Γ is a parameter governing the relative importance of centralized and decentralized market utility, $\sigma > 0$ is the risk aversion coefficient of the centralized market good, $\nu > 0$ is the inverse of the Frisch labor supply elasticity, and χ determines the preference for labor.

Wage rigidities are introduced as in Erceg et al. (2000), particularly following the extension of Hagedorn et al. (2019) for a heterogeneous agent context. Households, indexed by $i \in [0,1]$, provide differentiated labor services which are transformed by a representative, competitive labor recruiting firm into an aggregate effective labor input

$$H_t = \left(\int_0^1 s_{it}\left(h_{it}\right)^{\frac{\varepsilon_w-1}{\varepsilon_w}} di\right)^{\frac{\varepsilon_w}{\varepsilon_w-1}},$$

where ε_w is the elasticity of substitution across differentiated labor.

The labor recruiting firm buys labor services not directly from households but from labor unions. Labor unions sell labor services $s_{it}h_{it}$ at the nominal wage W_{it} to recruiters. The problem of the recruiter is then to minimize costs $\int_0^1 W_{it}s_{it}h_{it}di$ to produce a given amount of aggregate labor H_t. This implies the following demand for labor services of household i

$$h_{it} = h(W_{it}; W_t, H_t) = \left(\frac{W_{it}}{W_t}\right)^{-\varepsilon_w} H_t\,,$$

where the nominal wage W_t follows from the same minimization problem

$$W_t = \left(\int_0^1 s_{it}\left(W_{it}\right)^{1-\varepsilon_w} di\right)^{\frac{1}{1-\varepsilon_w}}.$$

Labor unions set the nominal wage $W_{it} = \hat{W}_t$ per effective unit of labor, i.e., they charge the same to recruiters irrespective of which household i supplies the effective units of labor $s_{it}h_{it}$. They maximize profits subject to wage adjustment costs analogous to Rotemberg (1982)

$$\Theta_t^w\left(\frac{W_{it}}{W_{it-1}}\right) = s_{it}\frac{\theta_w}{2}\left[\log\left(\frac{W_{it}}{W_{it-1}\tilde{\pi}^w}\right)\right]^2 H_t\,.$$

The union's wage-setting problem is

$$\mathrm{E}_t \sum_{k=t}^{\infty} \beta^{k-t} \int_0^1 \left[\tilde{\Pi}_k^w(\hat{W}_{ik}) - \Theta_k^w \left(\frac{\hat{W}_{ik}}{\hat{W}_{ik-1}} \right) \right] di \,,$$

where

$$\tilde{\Pi}_k^w(\hat{W}_{ik}) = \frac{s_{it}(1-\tau_h)\hat{W}_t}{P_t} h(\hat{W}_t; W_t, H_t) - \frac{\Upsilon(h(\hat{W}_t; W_t, H_t))}{U'(C_t)} \,,$$

and $C_t = \int_0^1 c_{it} di$ is aggregate consumption.

The solution to this problem yields the New Keynesian wage inflation equation

$$\log\left(\frac{\pi_t^w}{\tilde{\pi}^w}\right) = \beta \mathrm{E}_t \left[\log\left(\frac{\pi_{t+1}^w}{\tilde{\pi}^w}\right) \frac{H_{t+1}}{H_t} \right] + \frac{\varepsilon_w}{\theta_w} \left(\frac{\Upsilon'(H_t)}{U'(C_t)} - (1-\tau_h) \frac{\varepsilon_w - 1}{\varepsilon_w} w_t \right) \,,$$

where $\pi_t^w = \frac{W_t}{W_{t-1}} = \pi_t \frac{w_t}{w_{t-1}}$ is wage inflation and $\tilde{\pi}^w = \tilde{\pi}$ is wage inflation in the steady state, which is equal to the inflation target of the central bank.

2.1.2 Decentralized Market

At the beginning of the decentralized market, the idiosyncratic productivity for each household is realized. Furthermore, a preference shock determines whether a household becomes a buyer or a seller (each with probability $\gamma \in (0, 0.5]$). Buyers and sellers are then matched with each other in anonymous bilateral meetings, where sellers would like to produce and sell a special good, while buyers would like to buy said good. The DM value function before the realization of the preference shock is

$$\begin{aligned} V_t^{DM}(m_{t-1}, d_{t-1}, s_t) = {} & \gamma V_t^b(m_{t-1}, d_{t-1}, s_t) + \gamma V_t^s(m_{t-1}, d_{t-1}, s_t) \\ & + (1-2\gamma) V_t^{CM}(a_{t-1}, s_t) \,, \end{aligned}$$

where $a_{t-1} = \frac{R_{t-1}^M}{\pi_t} m_{t-1} + \frac{R_{t-1}^D}{\pi_t} d_{t-1}$ is the total wealth including interest payments which accrue in the centralized market. Note that with probability $1 - 2\gamma$, the household does not meet anyone in the decentralized market and directly continues to the centralized one.

Exchange in the decentralized market has to be quid pro quo since meetings are anonymous and credit arrangements are impossible. Following Chiu et al. (2021) there are three kinds of decentralized market meetings: First, with probability ϑ_1, the seller only accepts cash (e.g., he does not have a payment terminal to handle debit card transactions). This can be thought of as a small mom-and-pop store at the corner of a street, where the small number of transactions does not make investing in a POS terminal worthwhile. Another interpretation could be that at least some of these transactions are related to the informal economy, where cash is the preferred payment method. Second, with probability ϑ_2, the seller accepts only deposits. This can be thought of as an online store where payment with cash is risky or prohibitively costly due to the physical distance. Finally, with probability $\vartheta_3 = 1 - \vartheta_1 - \vartheta_2$, both payment forms are accepted, corresponding to buying goods at department stores and the like.

The value of being buyer can be written as follows

$$V_t^b(m_{t-1}, d_{t-1}, s_t) = \sum_{i=1}^{3} \left\{ \vartheta_i \int \left[u(\hat{c}_t^i(\xi_t, \tilde{\xi})) + V_t^{CM}\left(a_{t-1} - \hat{x}_t^i(\xi_t, \tilde{\xi}), s_t\right) \right] dF_t(\tilde{\xi}) \right\} ,$$

where $\xi_t = (m_{t-1}, d_{t-1}, s_t)'$ represent all state variables, $F(\cdot)$ is the distribution over asset holdings and productivity of households, $\hat{c}_t^i(\xi_t, \tilde{\xi})$ is the amount of goods produced/consumed, and $\hat{x}_t^i(\xi_t, \tilde{\xi})$ is the payment made by the buyer to the seller. Note that Lagos and Wright (2005) assume linearity of the CM utility in labor supply, which results in a degenerate distribution $F_t(\cdot)$ and allows for analytical results. Since the purpose of this paper is to devise a framework that allows for the analysis of differences along this distribution, I need to resort to numerical methods to keep track of the distribution.

Similarly, we have for the value function of the seller

$$V_t^s(m_{t-1}, d_{t-1}, s_t) = \sum_{i=1}^{3} \left\{ \vartheta_i \int \left[-v(\hat{c}_t^i(\tilde{\xi}, \xi_t)) + V_t^{CM}\left(a_{t-1} + \hat{x}_t^i(\tilde{\xi}, \xi_t), s_t\right) \right] dF_t(\tilde{\xi}) \right\} .$$

In the bargaining problem that is discussed below, the participation constraint of the seller will always be binding in equilibrium. Therefore, we have that

$$V_t^s(m_{t-1}, d_{t-1}, s_t) = V_t^{CM}(a_{t-1}, s_t) \, .$$

Bargaining. Following Keister and Sanches (2022) and Chiu et al. (2021), I assume that buyers make take-it-or-leave-it offers to sellers. In a meeting of type i, buyers solve

$$\begin{aligned} \max_{\hat{c}_t^i, \hat{x}_t^i} \quad & u(\hat{c}_t^i) + V_t^{CM}(a_{t-1} - \hat{x}_t^i, s_t) \\ \text{s.t.} \quad & -v(\hat{c}_t^i, \tilde{s}_t) + V_t^{CM}(\tilde{a}_{t-1} + \hat{x}_t^i, \tilde{s}_t) \geq V_t^{CM}(\tilde{a}_{t-1}, \tilde{s}_t) && \text{(Seller Participation Constraint)} \\ & f_i(m_{t-1}, d_{t-1}) \geq \hat{x}_t^i \geq 0 \, , && \text{(Liquidity Constraint)} \end{aligned}$$

where sellers state variables have been denoted with tildes, $f_i(\cdot)$ maps the assets of the household into feasible payment options for the meeting of type i. Note that the participation constraint must be binding, otherwise, the buyer could increase consumption $\hat{c}_t^i$

and, in turn, his utility for a given $\hat{x}_t^i$. The $f_i(\cdot)$'s are given by

$$
\begin{aligned}
f_1(m_{t-1}, d_{t-1}) &= \frac{R^M_{t-1}}{\pi_t} m_{t-1}\,, \\
f_2(m_{t-1}, d_{t-1}) &= \frac{R^D_{t-1}}{\pi_t} d_{t-1}\,, \\
f_3(m_{t-1}, d_{t-1}) &= \frac{R^M_{t-1}}{\pi_t} m_{t-1} + \frac{R^D_{t-1}}{\pi_t} d_{t-1}
\end{aligned}
$$

A key complication of solving the model proposed in this paper in comparison to standard HANK models is that due to the search frictions, the policies in the decentralized market not only depend on the households state variables but also the state variables of the counterparty a household meets, i.e. $\hat{c}_t^i = \hat{c}_i(\xi_t, \tilde{\xi})$ and $\hat{x}_t^i = \hat{x}_i(\tilde{\xi}, \xi_t)$. This implies that households, in contrast to the standard HANK models, need to take the whole distribution into account when determining their policies.

Following Chiu and Molico (2011), preferences in the decentralized market are given by

$$
u(\hat{c}) = \frac{(\hat{c} + b)^{1-\varsigma} - b^{1-\varsigma}}{1-\varsigma}\,,
$$

and

$$
v(\hat{h}) = \hat{h}\,,
$$

where $\varsigma > 0$ is the risk aversion coefficient of the decentralized market good, and $b \approx 0$ is needed to have a well-defined outside option of "no trade" in the decentralized market bargaining. It is assumed that producing one unit of consumption in the DM requires $1/s_t$ units of labor, i.e. $\hat{c}_t = s_t \hat{h}_t$. This implies that we can redefine $v(\cdot)$ as a function of $\hat{c}_t$ and s_t

$$
v(\hat{c}, s) = \frac{\hat{c}}{s}\,.
$$

2.2 Firms

2.2.1 Final Good Producer

A perfectly competitive final good producer uses a continuum of differentiated retail goods, indexed by $i \in [0, 1]$, to produce Y_t using a CES production function

$$
Y_t = \left(\int_0^1 Y_{it}^{\frac{\varepsilon_y - 1}{\varepsilon_y}} \, di \right)^{\frac{\varepsilon_y}{\varepsilon_y - 1}}
$$

where $\varepsilon_y > 1$ is the elasticity of substitution across retail goods.

The demand for intermediate good i can be derived from the profit maximization problem of the final good producer. We have that

$$Y_{it} = \left(\frac{P_{it}}{P_t}\right)^{-\varepsilon_y} Y_t \,.$$

Substituting this into the production function, one can derive an expression for the price level in the economy in terms of intermediate good prices

$$P_t = \left(\int_0^1 P_{it}^{1-\varepsilon_y} di\right)^{\frac{1}{1-\varepsilon_y}} .$$

2.2.2 Retailers

A continuum of monopolistically competitive retailers re-package intermediate goods and sell the resulting retail goods to final good producers. Retailers are subject to price adjustment costs as in Rotemberg (1982), which will introduce price stickiness into the model. I follow Fernández-Villaverde et al. (2021) in the specification of the adjustment costs and allow for indexation to the inflation target of the central bank

$$\Theta_t^y\left(\frac{P_{it}}{P_{it-1}}\right) = \frac{\theta_y}{2}\left[\log\left(\frac{P_{it}}{P_{it-1}\tilde{\pi}}\right)\right]^2 Y_t$$

where $\theta_y > 0$ determines the degree of price stickiness and $\tilde{\pi}$ is the inflation target of the central bank.

The problem of retailers is to choose a sequence of prices $\{P_{it}\}_{t\geq 0}$ to maximize

$$\mathrm{E}_t \sum_{k=t}^{\infty} \beta^{k-t}\left[\tilde{\Pi}_k^y(P_{ik}) - \Theta_k^y\left(\frac{P_{ik}}{P_{ik-1}}\right)\right]$$

where

$$\tilde{\Pi}_k^y(P_{ik}) = \left(\frac{P_{ik}}{P_k} - mc_k\right)\left(\frac{P_{ik}}{P_k}\right)^{-\varepsilon_y} Y_k$$

are profits net of adjustment cost, $mc_t = \frac{P_t^m}{P_t}$ are marginal costs and P_t^m is the price of intermediate goods. As in Hagedorn et al. (2019) and Fernández-Villaverde et al. (2021), I assume that price adjustment cost are virtual and do not imply real resource cost. They only affect decisions of retailers.

Profits of retailers Π_t^y are then given by

$$\Pi_t^y = \tilde{\Pi}_t^y(P_t) = (1 - mc_t)\, Y_t\,.$$

Solving the problem of retailers yields the New Keynesian Philips curve

$$\log\left(\frac{\pi_t}{\tilde{\pi}}\right) = \beta \mathrm{E}_t \left[\log\left(\frac{\pi_{t+1}}{\tilde{\pi}}\right) \frac{Y_{t+1}}{Y_t}\right] + \frac{\varepsilon_y}{\theta_y}\left(mc_t - \frac{\varepsilon_y - 1}{\varepsilon_y}\right),$$

where $\pi_t = \frac{P_t}{P_{t-1}}$ is gross inflation and $\tilde{\pi}$ is the gross inflation target of the central bank.

2.2.3 Intermediate Good Producers

Competitive intermediate-good firms produce intermediate goods and sell the goods to retailers. An intermediate good producer acquires capital K_t at a unit price Q_t at the end of period t to be used in production at $t+1$. Funding for the purchase of capital is obtained from banks in the form of a loan L_t. In equilibrium, we have that

$$Q_t K_t = L_t\,.$$

Intermediate goods are produced using capital K_{t-1} and labor H_t

$$Y_t^m = A_t K_{t-1}^{\alpha} H_t^{1-\alpha}\,.$$

At t, the amount of capital available for production K_{t-1} is predetermined. Therefore, the intermediate good producer solves

$$\max_{H_t} \quad P_t^m Y_t^m - P_t w_t H_t\,.$$

The first order condition of this problem yields

$$w_t = (1-\alpha) A_t mc_t \left(\frac{K_{t-1}}{H_t}\right)^{\alpha},$$

where marginal cost are defined as $mc_t = \frac{P_t^m}{P_t}$.

The cash flow from production is paid out to banks. We have that

$$P_t^m Y_t^m - P_t w_t H_t = P_t r_t^k K_{t-1}\,.$$

Rearranging yields

$$r_t^k = \alpha A_t mc_t \left(\frac{K_{t-1}}{H_t}\right)^{\alpha-1} .$$

Depreciated capital is sold back to capital good producers at price Q_t after production. The total return on loan L_t is, thus,

$$\frac{R_t^L}{\pi_t} = \frac{r_t^k + (1-\delta)Q_t}{Q_{t-1}} .$$

Competitive capital good producers buy capital from intermediate good producers after production, refurbish it and produce new capital, which they then sell to intermediate good producers for production in the subsequent period. Since there are no capital adjustment costs, we have that $Q_t = 1$ for all t.

2.3 Banks

A continuum of competitive banks, each endowed with a fixed amount of equity $\bar{E}$, invest in bonds/reserves B_t, grant loans L_t to firms and raise deposits D_t from households. The problem of a bank is

$$\begin{aligned} \max_{L_t B_t, D_t} \quad & \mathrm{E}_t \left[\frac{R_{t+1}^L}{\pi_{t+1}} L_t + \frac{R_t^B}{\pi_{t+1}} B_t - \frac{R_t^D}{\pi_{t+1}} D_t - \bar{E}\right] \\ \text{s.t.} \quad & L_t + B_t = D_t + \bar{E} . \end{aligned}$$

Using the balance sheet identity, the problem can be rewritten as

$$\max_{L_t, D_t} \mathrm{E}_t \left[\left(\frac{R_{t+1}^L}{\pi_{t+1}} - \frac{R_t^B}{\pi_{t+1}}\right) L_t + \left(\frac{R_t^B}{\pi_{t+1}} - \frac{R_t^D}{\pi_{t+1}}\right) D_t + \left(\frac{R_t^B}{\pi_{t+1}} - 1\right) \bar{E}\right] .$$

At the $t+1$ when firms pay back loans and the bank has to pay deposit rates to households, we have that the cash flow (or profits) of the bank are

$$\Pi_{t+1} = \left(\frac{R_{t+1}^L}{\pi_{t+1}} - \frac{R_t^B}{\pi_{t+1}}\right) L_t + \left(\frac{R_t^B}{\pi_{t+1}} - 1\right) \bar{E} ,$$

where we have used the fact that in equilibrium $R_t^D = R_t^B$.

Note that along a perfect foresight transition path we also have that $R_{t+1}^L = R_t^B$. However, on impact of an MIT shock, the return on bonds/reserves is predetermined, while R_{t+1}^L adjusts in response to the shock.

2.4 Consolidated Government

The consolidated government budget constraint is

$$B_t + M_t + \mathcal{T}_t = \frac{R^M_{t-1}}{\pi_t} M_{t-1} + \frac{R^B_{t-1}}{\pi_t} B_{t-1} + G_t \,.$$

where M_t refers to the supply of cash, B_t is consolidated government debt[7], $\mathcal{T}_t$ is tax revenue and G_t refers to government spending. Note that M_t is supplied elastically to satisfy household demand.

2.4.1 Monetary Policy

The central bank follows a standard Taylor rule

$$R^B_t = \tilde{R}^B \left(\frac{R^B_{t-1}}{\tilde{R}^B} \right)^{\rho_R} \left[\left(\frac{\pi_t}{\tilde{\pi}} \right)^{\phi_\pi} \left(\frac{Y_t}{\tilde{Y}} \right)^{\phi_y} \right]^{(1-\rho_R)} \,.$$

where $\tilde{\pi}$ is the gross inflation target of the central bank, $\tilde{R}^B$ is the gross nominal rate on bonds/reserves.[8] The policy rule allows for interest rate smoothing through $\rho_R \in [0, 1)$.

2.4.2 Fiscal Policy

The government raises proportional taxes on labor income, fully taxes profits, and raises a lump-sum tax T_t. Total tax revenue is given by

$$\mathcal{T}_t = \tau_h w_t H_t + \Pi_t + T_t \,.$$

Government debt is assumed to be constant

$$B_t = \tilde{B} \,,$$

which implies that in equilibrium, G_t adjusts to clear the government budget constraint.

[7] This includes central bank reserves and debt in the form of government bonds.

[8] Note the current specification only considers inflation and output in the centralized market in the Taylor rule to reduce the computational complexity. Aruoba and Schorfheide (2011) specify a Taylor rule where the central bank reacts to Fischer indices across DM and CM for output and inflation.

2.5 Market Clearing & Aggregation

In equilibrium all markets clear. In the centralized market, we have

$$
\begin{aligned}
Y_t = A_t K_{t-1}^{\alpha} H_t^{1-\alpha} &= \underbrace{\int_0^1 c_{t,i} di}_{=C_t} + \underbrace{K_t - (1-\delta) K_{t-1}}_{=I_t} + G_t \,, && \text{(Goods Market)} \\
H_t &= \int_0^1 s_{t,i} h_{t,i} di \,, && \text{(Labor Market)} \\
L_t &= Q_t K_t \,, && \text{(Loan Market)} \\
D_t &= \int_0^1 d_{t,i} di \,, && \text{(Deposit Market)} \\
M_t &= \int_0^1 m_{t,i} di \,, && \text{(Cash Market)} \\
L_t + B_t &= D_t + \bar{E} \,. && \text{(Bank Balance Sheet)}
\end{aligned}
$$

In the decentralized market, market clearing is implicit in the bargaining problem where it is assumed that the production of the seller is consumed by the buyer.

Following Aruoba and Schorfheide (2011), aggregate output in terms of the centralized market good is given by

$$
\mathcal{Y}_t = \gamma \sum_{i=1}^{3} \vartheta_i \int \int \hat{p}_t^i(\xi, \tilde{\xi}) \hat{c}_t^i(\xi, \tilde{\xi}) dF_t(\tilde{\xi}) dF_t(\xi) + Y_t \,,
$$

where $\hat{p}_t^i(\xi, \tilde{\xi}) = \hat{x}_t^i(\xi, \tilde{\xi}) / \hat{c}_t^i(\xi, \tilde{\xi})$ is the price of the decentralized market good in terms of the centralized market good.

Analogously, I define the aggregate consumption expenditures in terms of the centralized market good as

$$
\mathcal{C}_t = \gamma \sum_{i=1}^{3} \vartheta_i \int \int \hat{p}_t^i(\xi, \tilde{\xi}) \hat{c}_t^i(\xi, \tilde{\xi}) dF_t(\tilde{\xi}) dF_t(\xi) + C_t \,.
$$

Finally, aggregate (gross) inflation is defined using Fisher's ideal price index

$$
\begin{aligned}
\pi_t^{agg} = \pi_t \Bigg[& \frac{\gamma \sum_{i=1}^{3} \vartheta_i \int \int \hat{p}_t^i(\xi, \tilde{\xi}) \hat{c}_t^i(\xi, \tilde{\xi}) dF_t(\tilde{\xi}) dF_t(\xi) + Y_t}{\gamma \sum_{i=1}^{3} \vartheta_i \int \int \hat{p}_{t-1}^i(\xi, \tilde{\xi}) \hat{c}_t^i(\xi, \tilde{\xi}) dF_t(\tilde{\xi}) dF_t(\xi) + Y_t} \\
& \times \frac{\gamma \sum_{i=1}^{3} \vartheta_i \int \int \hat{p}_t^i(\xi, \tilde{\xi}) \hat{c}_{t-1}^i(\xi, \tilde{\xi}) dF_{t-1}(\tilde{\xi}) dF_{t-1}(\xi) + Y_{t-1}}{\gamma \sum_{i=1}^{3} \vartheta_i \int \int \hat{p}_{t-1}^i(\xi, \tilde{\xi}) \hat{c}_{t-1}^i(\xi, \tilde{\xi}) dF_{t-1}(\tilde{\xi}) dF_{t-1}(\xi) + Y_{t-1}} \Bigg]^{0.5} \,.
\end{aligned}
$$

Note that in the steady state, we have that $\tilde{\pi}^{agg} = \tilde{\pi}$.

3 Calibration

Table 1 shows the calibration of the model and the chosen targets. The model is calibrated to a quarterly frequency using euro area aggregate data. To match micromoments of the data to those in the model, I use Spanish microdata from the Encuesta Financiera de las Familias (EFF) since the equivalent Euro area survey does not have questions regarding the cash holdings of households.

Table 1: Baseline Calibration

Parameter		Value	Target/Source
Household			
β	Discount factor	0.99	$\tilde{K}/\tilde{Y} = 9.7$
Γ	CM utility weight	1.72	$\tilde{M}/\tilde{Y} = 0.19$, $\tilde{D}/\tilde{Y} = 3.86$
σ	Risk aversion (CM)	1	Standard
$1/\nu$	Frisch elasticity of labor supply	0.5	Standard
χ	Disutility of labor	23.72	$\tilde{H} = 0.33$
ς	Risk aversion (DM)	2	Cash-to-consumption ratio
b	Utility term (DM)	10^{-4}	Standard
γ	Probability of being a seller/buyer	0.4	$\tilde{M}/\tilde{Y} = 0.19$, $\tilde{D}/\tilde{Y} = 3.86$
ϑ_1	Probability of type 1 meetings	0.19	European Central Bank (2020)
ϑ_2	Probability of type 2 meetings	0.09	European Central Bank (2020)
ϑ_3	Probability of type 3 meetings	0.72	European Central Bank (2020)
ε_w	Labor substitution elasticity	11	10% wage markup
θ_w	Wage adjustment cost	90	NK Philips curve slope
ρ_s	AR coefficient of process for s_t	0.98	Bayer et al. (2021)
ω_s	Standard deviation of prod. shock	0.12	Bayer et al. (2021)
Firms			
α	Capital share	0.33	Standard
δ	Depreciation rate	0.025	Standard
ε_y	Good substitution elasticity	11	10% price markup
θ_y	Price adjustment cost	90	NK wage inflation slope
Banks			
$\bar{E}$	Fixed bank equity	7.07	$\tilde{K}/\tilde{Y} = 9.7$
Consolidated Government			
$\tilde{\pi}$	Inflation target	$1.02^{1/4}$	2% annual inflation target
ϕ_π	Taylor rule coeff. on inflation	1.5	Standard
ϕ_y	Taylor rule coeff. on output	0.1	Standard
$\tilde{B}$	Steady state government debt	1.17	$\tilde{B}/\tilde{Y} = 1.16$

I start with the description of externally calibrated parameters. I set the annual inflation target for the calibration to 2%, i.e., $\tilde{\pi} = 1.02^{1/4}$. In the exercises of the following sections, however, this value will be varied to compare different inflation levels in the steady state. The Taylor rule coefficients are set to satisfy the Taylor principle and chosen in accordance with the literature. In particular, I set the coefficient on deviations of inflation from the target to $\phi_\pi = 1.5$ and the coefficient on deviations of output from the steady state to $\phi_y = 0.1$. The capital share is set to a value of $\alpha = 0.33$ with the depreciation rate being set to around 10% per annum, i.e., $\delta = 0.025$, in line with the values commonly found in the literature. Regarding the parameters governing the New Keynesian Philips curve and the wage inflation equation, I choose the substitution elasticities to be $\varepsilon_y = 11$ and $\varepsilon_w = 11$ which imply a 10% markup in the steady state for prices and wages, respectively. Furthermore, I set the price and wage adjustment cost parameters to $\theta_y = \theta_w = 90$ such that the slope of the New Keynesian Philips curve and the wage inflation equation are equal $\varepsilon_y/\theta_y = \varepsilon_w/\theta_w = 0.12$, which is slightly above the estimates in Bayer et al. (2022).

The probabilities of the different meetings in the decentralized market are determined following the approach of Chiu et al. (2021) applied to data from the study on the payment attitudes of consumers in the euro area (SPACE, European Central Bank, 2020). I assume that all online payments (7% of transactions) accept only deposits as payment. In 21% of POS transactions, cards are not accepted. Thus, the probability of a meeting where only cash is accepted is set to $\vartheta_1 = (1 - 0.07)0.21 = 0.1953$. Using the fact that cash is accepted in 98% of POS transactions in the data, we can compute that the probability of a meeting where only deposits are accepted is $\vartheta_2 = 0.07 + (1 - 0.07)(1 - 0.98) = 0.0886$. Finally, the probability of a meeting where both forms of payments are accepted is then given by the residual $\vartheta_3 = 1 - \vartheta_2 - \vartheta_3 = 0.7161$. I set the risk aversion coefficient in the centralized market to $\sigma = 1$, implying log utility and the Frisch labor supply elasticity to $1/\nu = 0.5$. Both values are commonly chosen in the literature. The idiosyncratic risk is calibrated based on an AR(1) process with coefficient $\rho_s = 0.98$, and standard deviation $\omega_s = 0.12$ for the shock as in Bayer et al. (2021) and is then discretized into a two-state Markov chain using the Rouwenhorst method.

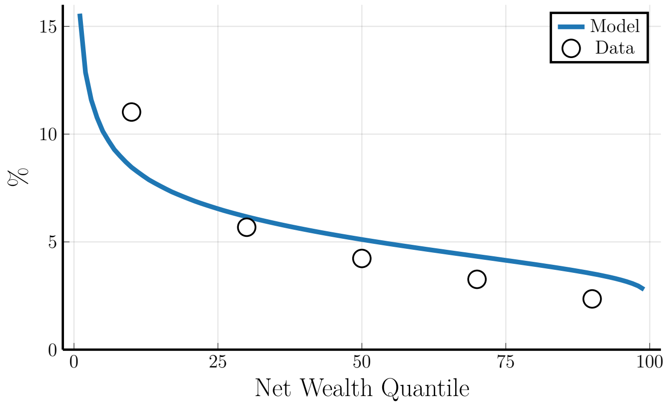
Table 2: Calibration Targets

Target	Data	Model
Currency to GDP Ratio (M_t/Y_t)	0.19	0.19
Deposit to GDP Ratio (D_t/Y_t)	3.86	3.86
Reserves/Bonds (held by Banks) to GDP Ratio (B_t/Y_t)	1.16	1.16
Capital to GDP Ratio (K_{t-1}/Y_t)	9.7	9.7
Hours worked (H_t)	0.33	0.33
ΔCash-to-Consumption Ratio across Income	0.01	0.01

The remaining parameters are calibrated jointly to match several moments in the data as shown in Table 2. The labor disutility parameter $\chi = 23.72$ is set such that aggregate labor supply $\tilde{H} = 0.33$ in the steady state. The CM utility weight $\Gamma = 1.72$ and the probability of being a buyer/seller $\gamma = 0.4$ influence the relevance of the decentralized market and are set to match the currency in circulation (excluding the 48% of currency held by foreigners as estimated in Zamora-Pérez, 2021) relative to quarterly GDP $\tilde{M}/\tilde{Y} = 0.19$ and deposits (M3 - currency in circulation) relative to deposits $\tilde{D}/\tilde{Y} = 3.86$. To match the aggregate capital stock relative to quarterly GDP $\tilde{K}/\tilde{Y} = 9.7$, as determined in de Ferra et al.

(2020), I set the discount factor to $\beta = 0.98998$ and the fixed equity of banks to $\bar{E} = 7.07$, which allows accounting for funding sources of banks other than bank deposits. Finally, I calibrate $\varsigma = 2$ to match the difference in the cash-to-consumption ratio across the bottom 50% and the top 50% of households in terms of income in the EFF data to the difference in the cash-to-consumption ratio of low-income and high-income types in the model.

Figure 2: Untargeted Moment: Cash as a Share of Liquid Assets



Notes: The data shows the median cash holdings as a share of liquid assets for different quantiles of the net wealth based on an average of all waves of the EFF from 2002 to 2017. The model equivalent shows $m_t/(m_t + d_t)$.

Figure 2 shows the model fit based on cash as a share of liquid assets (defined as the sum of cash and deposit accounts that can be used for payments). It is important to note that this was not a calibration target. Nevertheless, the model can capture that poorer households hold a larger fraction of their liquid assets in cash quite well. This is because for very poor households the marginal utility of holding cash is larger than the marginal utility of holding deposits even though deposits yield a positive nominal return. Cash can be used to buy consumption goods in over 90% of the potential meetings in the decentralized market, while deposits can only be used in around 80% of the meetings. Due to this high acceptance rate, which is consistent with the data (European Central Bank, 2020), households with very few resources prefer to hold cash to meet their liquidity needs, even if it means they forgo interest.

The calibration implies that the decentralized market output is around 7.5% of total output $\mathcal{Y}_t$. The DM share of output is roughly in line with other New Monetarist models with a centralized market setup similar to the model developed in this paper. For example, the DM share is a bit larger than the 3% DM share in Aruoba et al. (2011), but smaller than the 20% DM share in Aruoba and Schorfheide (2011).

4 Numerical Solution

The solution of the proposed model poses a computational challenge. Because of the search setup in the decentralized market, households' policy functions directly depend on the whole distribution of asset holdings and household productivities. This stands in contrast to the household's problem in standard HANK models, which can be solved independently of the asset distributions since prices (e.g., interest rates and wages) capture all the relevant information from a household's perspective.

The key contribution of this paper in terms of the algorithm lies not in the techniques themselves but in a particular reframing of the centralized and decentralized market problems, which makes them tractable in the face of the so-called "curse of dimensionality". The "curse of dimensionality" stems from the fact that in order to solve the household problem one needs to discretize the state space on which the policy functions are defined. The computational complexity then strongly depends on the number of state variables the household problem has and the number of grid points used for the discretization. It is particularly pronounced in the model proposed in this paper because households consider the whole distribution when forming expectations about potential meetings in the decentralized market. Therefore, the "curse" depends on not only the household's own states but also the states of the counterparty.

The algorithm exploits two properties of the model: First, the centralized market problem only depends on two state variables $(\hat{a}_t, s_t)$ even though the decentralized market problem, and, therefore, the dynamic programming problem as a whole when viewed from one period to the next, depends on three state variables (m_{t-1}, d_{t-1}, s_t).[9] Second, the decentralized market bargaining problem can be restated in terms of assets that are accepted in the exchange and assets that are not accepted in the exchange, making the solution of this subproblem independent of the type of meeting but only dependent on the amount of "liquid and illiquid" assets that buyers and sellers bring to a meeting (in addition to their respective productivity types).

In Appendix B, I discuss the algorithms to solve for the model's steady state and the transitional dynamics after MIT shocks, respectively.

5 Quantitative Results

This section discusses the quantitative model properties in the steady state, the dynamics after an MIT shock, and the welfare implications of inflation in the context of the model. Throughout this section, I will refer to the model proposed in this paper as HANK-NM, where NM stands for New Monetarist, while I refer to the standard HANK model with $\gamma = 0$ simply as HANK.[10]

[9] From the point of extensibility, this is crucial. The proposed algorithm allows for additional liquid assets that can be used for payments in the decentralized market with only a small increase in computational complexity since the relevant state space is given by the centralized market state variables $(\hat{a}_t, s_t)$. That the state space in the decentralized market grows with each added liquid asset only matters to the extent that it implies an additional asset choice in the centralized market but it does not affect the dimensionality of the policy functions.

[10] β, $\bar{E}$, χ, B, and Γ in HANK have been recalibrated to match the same aggregate moments (except for aggregate cash holdings) as HANK-NM. Appendix C.3 shows the main results for HANK with the same

5.1 Breaking the Monetary Super-Neutrality in HANK

Money is not super-neutral in the model developed in this paper, in contrast to a large range of Heterogeneous Agent New Keynesian (HANK) models, where changes in the inflation target only change nominal variables in the long run but have no effect on real variables.[11] As a result, changes in the inflation target $\tilde{\pi}$ by the central bank and the resulting change in the long-run level of inflation do have lasting effects on aggregates and the wealth distribution.

The non-neutrality stems from the transactional motive for cash, which is introduced through the decentralized market setup borrowed from Lagos and Wright (2005).[12] Since the nominal return on cash R^M_{t-1} is fixed at zero percent, an increase in the inflation target of the central bank reduces the real return on cash $\frac{R^M_{t-1}}{\pi_t}$ and, therefore, affects the saving decision of households. This, in turn, has several aggregate and distributional implications, which are discussed in detail below. It is important to note, however, that since the New Keynesian block of the model still exhibits the features that lead to the super-neutrality of money in HANK models, e.g., Rotemberg adjustment costs are indexed to inflation, the super-neutrality can be restored by setting the probability of decentralized market meetings to zero ($\gamma = 0$). In such a case, households would have no incentive to hold cash since it is dominated in return by deposits, and changes in the central bank's inflation target would not affect households' savings decisions.

To understand the channels at work, it is important to keep in mind that the real return on cash $\frac{R^M_{t-1}}{\pi_t}$ always falls proportionally to the inflation rate since the nominal return is fixed at 0%, while the real return on deposits $\frac{R^D_{t-1}}{\pi_t}$ is always tied to the real return on capital through the banking system and, therefore, only changes to the extent that aggregate capital and labor supply change. Thus, there is a direct effect of a change in the inflation target on the real return on cash but only an indirect, general equilibrium effect on the return on deposits.

Figure 3 summarizes the savings behavior of households. Panel (a) shows a partial equilibrium setting where all prices and aggregates are fixed, but the real return on cash falls proportionally to the inflation rate. In response to the lower real return on cash, households reduce their cash holdings and substitute them to some extent by increasing their deposit holdings.[13] However, since the effective return on savings

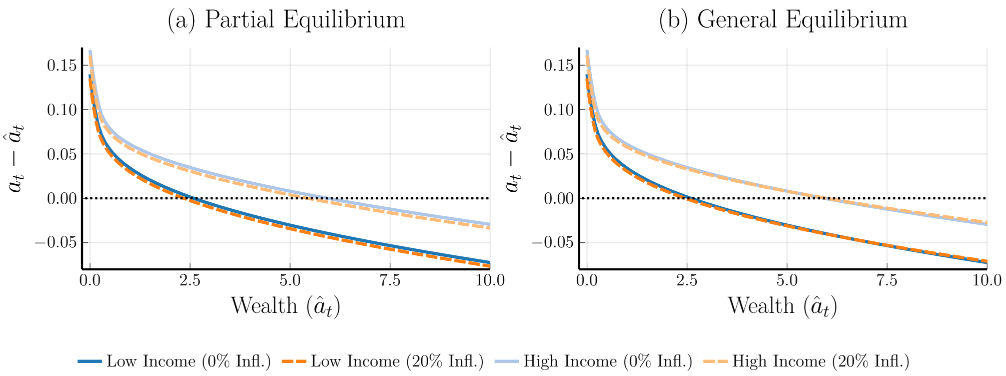
parametrization as HANK-NM for reference.

[11] HANK models featuring Rotemberg price/wage adjustment costs that are indexed to inflation such as Fernández-Villaverde et al. (2021) feature super-neutrality of money in the (deterministic) steady state. However, nominal rigidities à la Calvo or Rotemberg adjustment costs that are not indexed to inflation break monetary super-neutrality as well. Appendix C.1 shows the main results if monetary super-neutrality fails due to both nominal rigidities and monetary search frictions.

[12] The monetary super-non-neutrality due to the medium of exchange role can also be obtained in a representative agent context as in Lagos and Wright (2005). However, as shown in Appendix C.4, the preference restrictions required for the representative agent setup imply that centralized market consumption does not respond to changes in the inflation target. Thus, HANK-NM quantitatively implies a much stronger total consumption response than the representative agent model.

[13] This can be seen in Figures 15, 16 and 18 in Appendix D for different levels of wealth.

Figure 3: Change in Wealth Accumulation



Notes: In partial equilibrium, only the real return on cash R^M_{t-1}/π_t is assumed to be affected by inflation, while all other prices (including the real return on deposits R^D_{t-1}/π_t) and aggregates are held constant. In general equilibrium, all prices and aggregates adjust such that markets clear.

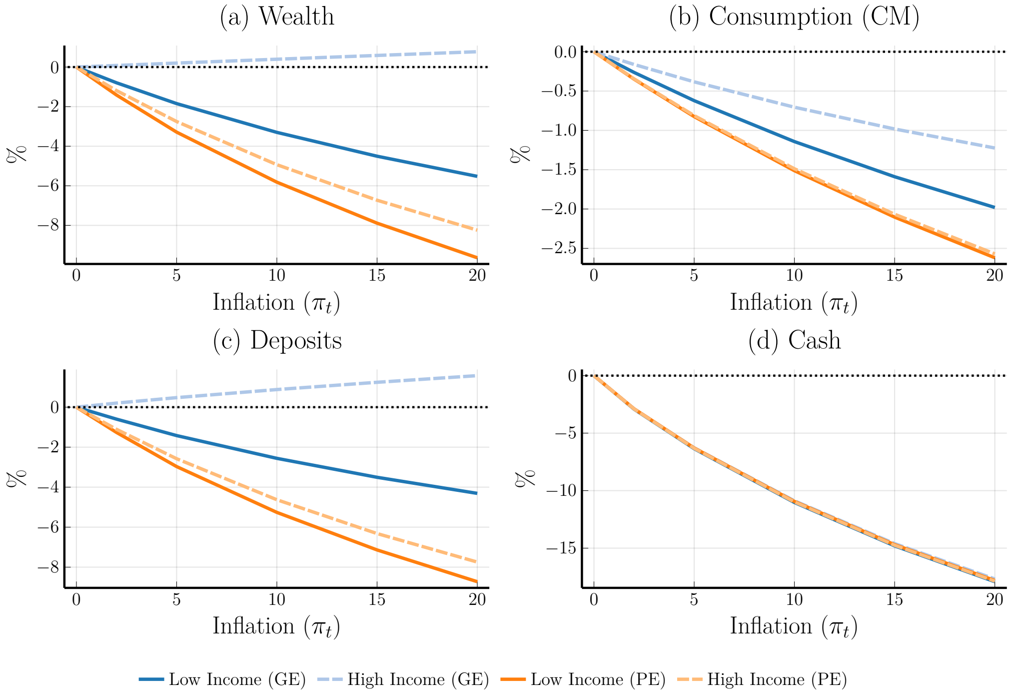
$\frac{R^E_{t-1}}{\pi_t} = \frac{R^M_{t-1}}{\pi_t}\frac{m_{t-1}}{m_{t-1}+d_{t-1}} + \frac{R^D_{t-1}}{\pi_t}\frac{d_{t-1}}{m_{t-1}+d_{t-1}}$, a portfolio-weighted average of the returns on cash and deposits is lower for all households, both low- and high-income households tend to accumulate less wealth than under zero inflation.[14] Figure 4 shows how the reduction in wealth accumulation then maps into lower consumption, deposit, and cash choices for both low- and high-income households.

Two channels dampen the partial equilibrium effects on wealth accumulation in general equilibrium (see Panel (b) of Figure 3). First, since households in partial equilibrium tend to accumulate less wealth, they also tend to reduce their deposit holdings. This means that less capital is available to firms for production, leading to an increase in the return on capital and, in turn, to a rise in the return on deposits. This mainly induces wealthy households, who hold more deposits in relative and absolute terms, to increase their savings. Second, labor income increases for all households because of an increase in the labor supply negotiated by labor unions (more on this below). Interestingly, these general equilibrium increases in the real return on deposits and labor income are strong enough for high-income households to overturn the partial equilibrium result, meaning that they tend to accumulate slightly more wealth as inflation increases, as seen in Figure 4. However, low-income households still accumulate substantially less wealth under higher inflation rates. Furthermore, wealth-poor households tend to accumulate wealth more slowly under higher inflation rates, while wealth-rich households deaccumulate more slowly which can be seen in Figure 3. Since the effect on poorer households is stronger, the economy as a whole still accumulates less wealth under higher inflation but the wealth distribution is slightly more dispersed because of the differential effects discussed above.

Table 3 shows the resulting aggregate effects. The aggregate demand for cash M_t falls

[14]Since poorer households depend more on cash the effective return on savings that they face (after portfolio rebalancing) falls more strongly than the one for richer households, as can be seen in Figure 14 in Appendix D.

Figure 4: Change in Individual Stochastic Steady State (iSSS)



Notes: The individual stochastic steady state (iSSS) is the amount of wealth $\hat{a}_t$ a household would accumulate (and the policy choices it would make with that amount of wealth) if it would not experience any change in its productivity/income type or not receive a liquidity shock (i.e. be a non-participant in the DM) for a long time until it converges to a point where its wealth and policies remain constant. PE refers to partial equilibrium, while GE refers to general equilibrium. The iSSS captures both the policy change and the fact that households tend to accumulate less wealth, which would not be the case if policies were compared for given points in the state space.

by more than 6% at 5% inflation and almost 18% at 20% inflation. Aggregate deposit holdings D_t fall by 0.42% at 5% inflation due to the lower wealth accumulation discussed above, even though households substitute some of their cash holdings with deposits. As a result, aggregate wealth $\hat{A}_t = \int_0^1 \hat{a}_t di$ falls by 0.75% at 5% inflation. Furthermore, the real return on deposits $\frac{R^D_{t-1}}{\pi_t}$ increases by around 3bp at 5% inflation, meaning that the model features a positive relationship between the inflation target and long-run real interest rates.

The changes in savings behavior also lead to changes in the consumption of households. In the aggregate, consumption in the centralized market falls by 0.51% at 5% inflation. From an individual's perspective, this is the result of several channels. First, the fall in the effective return on savings discussed above leads to an increase in consumption through a substitution effect that is overturned by an income effect for wealthier households. Second, since cash and deposits are not only used for shifting resources across periods but also for purchases in the decentralized market, there is an additional force preventing a further reduction in cash and deposits at the cost of lower consumption in the centralized market.[15]

[15]See Appendix A.1.2 for details on the liquidity premia present in the Euler equations.

Table 3: Comparative Statics

	Inflation				
	2%	5%	10%	15%	20%
Centralized Market					
Output (Y_t)	0.03	0.08	0.15	0.21	0.27
Consumption (C_t)	−0.22	−0.51	−0.94	−1.31	−1.63
Government Spending (G_t)	0.31	0.73	1.35	1.89	2.37
Taxes ($\mathcal{T}_t$)	0.08	0.20	0.36	0.51	0.64
Wealth ($\hat{A}_t$)	−0.34	−0.75	−1.32	−1.77	−2.13
Cash (M_t)	−2.90	−6.32	−10.97	−14.72	−17.82
Deposits (D_t)	−0.18	−0.42	−0.74	−0.98	−1.17
Capital (K_t)	−0.07	−0.17	−0.29	−0.39	−0.47
Labor Supply (H_t)	0.08	0.20	0.36	0.51	0.64
Wage (w_t)	−0.05	−0.12	−0.22	−0.30	−0.36
Real Deposit Rate (R^D_{t-1}/π_t)	0.01	0.03	0.06	0.08	0.09
Decentralized Market					
Consumption ($\hat{C}_t$)	−0.52	−1.23	−2.22	−3.04	−3.73
Payment ($\hat{X}_t$)	−0.75	−1.75	−3.17	−4.34	−5.34
Relative Price ($\hat{P}_t$)	−0.22	−0.53	−0.97	−1.35	−1.68
Cross-Market Aggregates					
Output ($\mathcal{Y}_t$)	−0.03	−0.06	−0.10	−0.13	−0.15
Consumption ($\mathcal{C}_t$)	−0.32	−0.74	−1.36	−1.88	−2.32

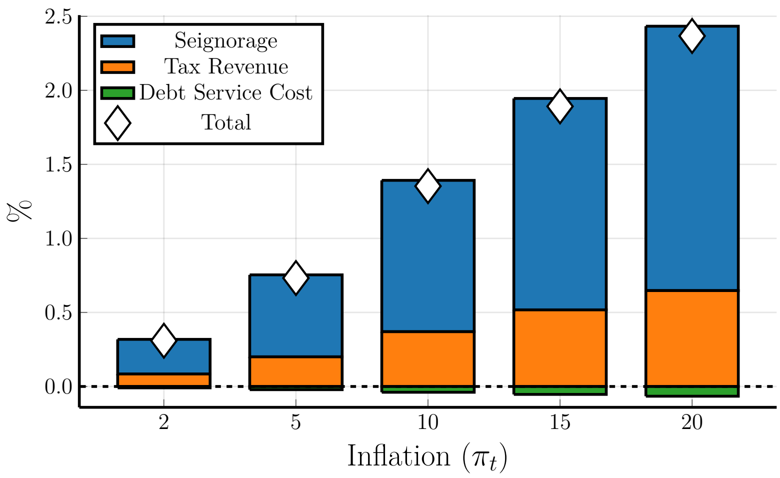
Notes: Change relative to zero-inflation steady state. All changes are expressed in percent except for interest rates, which are expressed in percentage points.

In other words, households reduce consumption in the centralized market to hold enough cash and deposits to purchase goods in the decentralized market. Finally, since households tend to accumulate less wealth as inflation increases, they also have fewer resources to spend on centralized market consumption. Together these channels reduce consumption as the inflation target increases. As shown in Figure 4, low-income households tend to experience a stronger decline in consumption than high-income households.

On the bank/firm side in the centralized market, we have that the reduction in aggregate deposit holdings of households leads to a fall in loans to intermediate-good producers and, thus, lower capital. However, since part of the loans to firms are funded using the fixed equity of banks, the drop in capital is not as pronounced as the drop in deposits as inflation increases. Furthermore, labor unions negotiate a higher labor supply because of the fall in aggregate consumption. This is similar to the wealth effect on labor supply typically found in macro-models with additively separable preferences. This increase is sufficiently large for the household to compensate for the fall in wages resulting from the smaller capital stock, meaning that the labor income of households rises. The decline in capital and the rise in labor supply lead to an overall increase in output in the centralized market. Since both consumption and investment in the centralized market fall, this higher production is mainly used for higher government spending. The increase in output also implies an increase in retailers' profits $\Pi_t^y = (1 - mc_t)Y_t$ since marginal costs mc_t are independent of inflation.

For the government, the increase in inflation leads to an increase in revenue through different channels. First, seignorage revenue increases since inflation is an implicit tax on cash holdings of households. Second, tax revenues from labor income and profit taxes increase due to the increase in labor income and profits discussed above. Finally, the debt servicing costs rise slightly because of the increase in the real interest rate on government debt.[16] Overall, this leads to an increase in government spending. Figure 5 shows that the main driver for this increase is the increase in seignorage revenue for the government.

Figure 5: Decomposition of the Change in Government Spending



Notes: The three components of the government budget constraint are defined as follows: seignorage $M_t - \frac{R_{t-1}^M}{\pi_t} M_{t-1}$, tax revenue $\mathcal{T}_t$, and debt servicing cost $B_t - \frac{R_{t-1}^B}{\pi_t} B_{t-1}$.

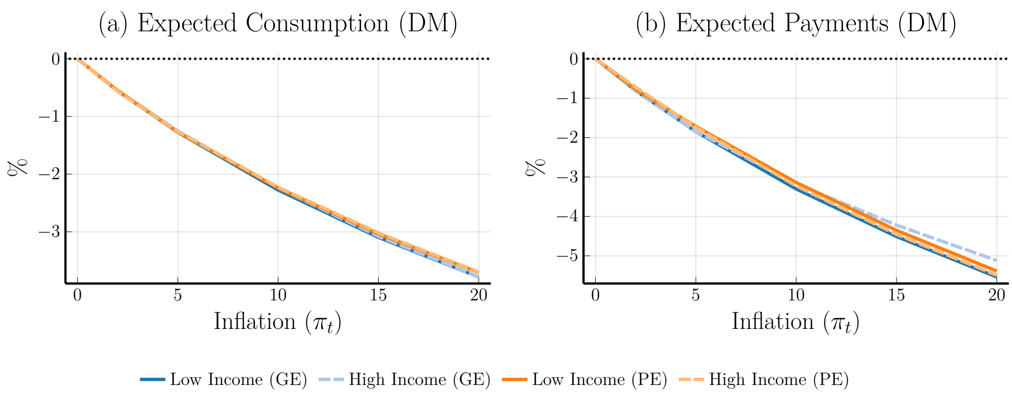
In the decentralized market, aggregate consumption $\hat{C}_t$ which is an aggregate of all meetings taking place in the DM falls by 1.23% at 5% inflation, which is more than twice as large as the reduction in centralized market consumption. The main driver for this is that households economize on cash, severely limiting the possible exchange in decentralized market meetings where sellers only accept cash as a form of payment. In fact, in cash-only meetings, consumption drops by 6.48% at 5% inflation, while in the other types of meetings it remains essentially constant. Aggregate payments $\hat{X}_t$ that are being made in the decentralized market fall even more strongly than consumption. Since households bring less wealth into the DM and, ultimately, into the CM, sellers are willing to produce more of the consumption good $\hat{c}_t$ in the decentralized market for a given payment $\hat{x}_t$. Thus, the price of DM consumption in terms of centralized market consumption falls $\hat{p}_t = \hat{x}_t / \hat{c}_t$, which can be seen in the aggregate measure of the relative price $\hat{P}_t$. As shown in Fig-

[16] In equilibrium, the real return on deposits equals the real return on government debt, i.e. $\frac{R_{t-1}^D}{\pi_t} = \frac{R_{t-1}^B}{\pi_t}$.

ure 6, there is little difference between low-income and high-income households in how decentralized market variables respond to changes in inflation in the individual stochastic steady state (iSSS). This reflects that both types of households reduce their cash holdings similarly.

Aggregating output and consumption across both sectors of the economy, i.e., across both the centralized and the decentralized market, we have that both output and consumption fall. The fall in decentralized market output, equal to decentralized market consumption, is sufficiently strong to overturn the increase in centralized market output. Thus, there is a negative relationship between the central bank's inflation target and aggregate output in the long run, even though the relationship does not hold for all sectors of the economy.

Figure 6: Change in Individual Stochastic Steady State (iSSS) in the Decentralized Market



Notes: The individual stochastic steady state (iSSS) is the amount of wealth $\hat{a}_t$ a household would accumulate (and the policy choices it would make with that amount of wealth) if it would not experience any change in its productivity/income type or not receive a liquidity shock (i.e. be a non-participant in the DM) for a long time until it converges to a point where its wealth and policies remain constant. PE refers to partial equilibrium, while GE refers to general equilibrium. The iSSS captures both the policy change and the fact that households tend to accumulate less wealth, which would not be the case if policies were compared for given points in the state space.

To summarize, since households would like to hold cash to carry out transactions in the decentralized market, money is not super-neutral in the model proposed in this paper. As a result, higher inflation leads to a reduction in output, consumption, and aggregate wealth. Poorer households (both in terms of income and wealth) tend to be more strongly affected by the negative consequences of inflation. They tend to accumulate less wealth and, as a result, suffer from a more substantial decline in consumption. Furthermore, due to the implicit inflation tax on cash and the associated increase in seignorage revenue, the consolidated government profits from higher inflation rates and increases its spending.

5.2 Restoring the Monetary Super-Neutrality through Interest on Cash

The previous section discussed how the introduction of the decentralized market into the model breaks the super-neutrality of money and how changes in the inflation target can affect real variables in the long run. The core issue leading to non-neutrality is that the nominal interest rate on cash is fixed at 0%, while the nominal return on deposits can freely adjust. As inflation increases, this leads to distortions[17] in the savings decisions of households, while in the nested HANK model with $\gamma = 0$ the nominal deposit rate adjusts to keep the household problem exactly the same in real terms.

In this subsection, I argue that the super-neutrality result can be recovered if the central bank implements the following rule for the nominal return on cash

$$R_{t-1}^{M} = \pi_t \,. \tag{1}$$

The argument for why this restores the super-neutrality is as follows: Suppose we are in the steady state with zero inflation, and the central bank follows through on the cash interest rate policy in Equation (1). As the inflation target (and therefore the inflation rate in the steady state) increases, the real return on cash $\frac{R_{t-1}^{M}}{\pi_t}$ remains constant at 0%. Suppose now that the real deposit return, wages, and asset distributions were to remain the same as well. This would mean that households face the same problem as in the zero-inflation case and would optimally make the same choices. In turn, aggregate cash and deposit holdings and aggregate labor supply would remain the same. Given the firms' problem, this is consistent with the real deposit return and wages that persisted in the zero-inflation case. The only thing left to check is the government budget constraint that happens to be unaffected by the change in inflation since all nominal interest rates adjust to keep real interest rates the same. Thus, changes in the inflation target by the central bank become irrelevant for the determination of real variables in the long run.

From a practical point of view, this rule could be understood as the central bank replacing cash with a central bank digital currency (CBDC) that pays a nominal interest rate different from zero. As we will see in Section 5.4, paying an interest rate on CBDC, which is tied to the inflation rate, can be justified through the lens of this model because it will eliminate the welfare costs of inflation.

5.3 Transitional Dynamics After an MIT Shock

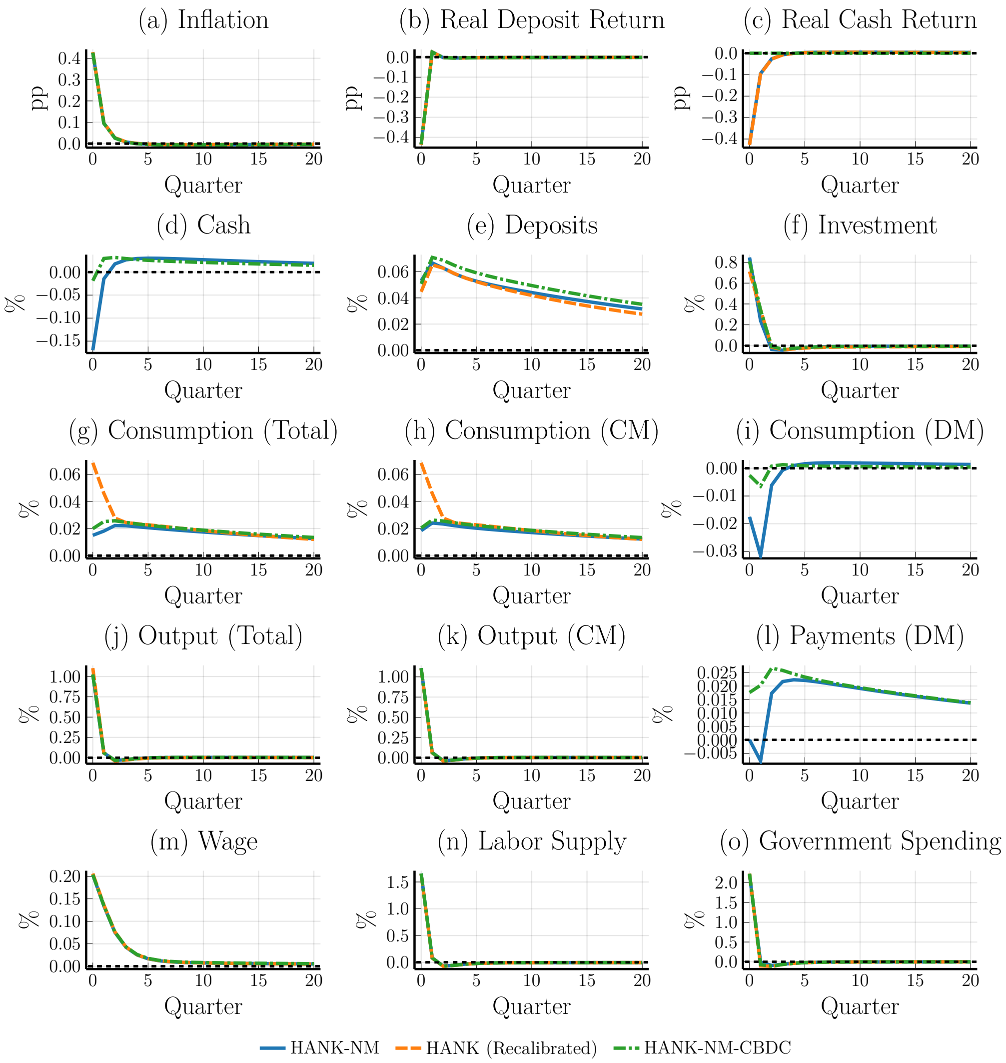
Figure 7 shows the impulse responses to an expansionary monetary policy shock in the zero-inflation steady state for HANK, HANK-NM, and HANK-NM-CBDC.[18] The responses of interest rates and inflation are quite similar in all three models. Panel (a) shows that inflation increases in response to the shock over the whole horizon of the IRF.

[17]Note that the households economize on cash even if inflation is zero, meaning that zero inflation is not necessarily the optimal inflation rate. This is related to the Friedman rule (Friedman, 1969) that can be shown to be optimal in many representative agent monetary models (e.g. Lagos and Wright, 2005). However, it is known at least since Bewley (1983) that in a heterogeneous agent model, the Friedman rule, which would require $\pi = \beta$, cannot be an equilibrium.

[18]The monetary policy shock follows an AR(1) process $\eta_t = \rho_m \eta_{t-1} + \sigma_m \upsilon_t$ with $\rho_m = 0.15$ and $\sigma_m = 0.0024$ are chosen to match the estimates in Smets and Wouters (2007). The parametrizations for HANK-NM and HANK-NM-CBDC are the same. HANK has been recalibrated to match the same aggregate moments as HANK-NM. The difference between HANK-NM and HANK-NM-CBDC is that the central bank follows a rule for the return on cash as described in Equation (1) in HANK-NM-CBDC.

Since the nominal return on cash is constant in HANK-NM, the real return on cash falls mechanically with inflation. While the effect on the real cash return is large and persistent, the real deposit return only falls on impact of the shock by a similar order of magnitude. After the impact period, the real deposit return increases slightly, as shown in Panel (b).

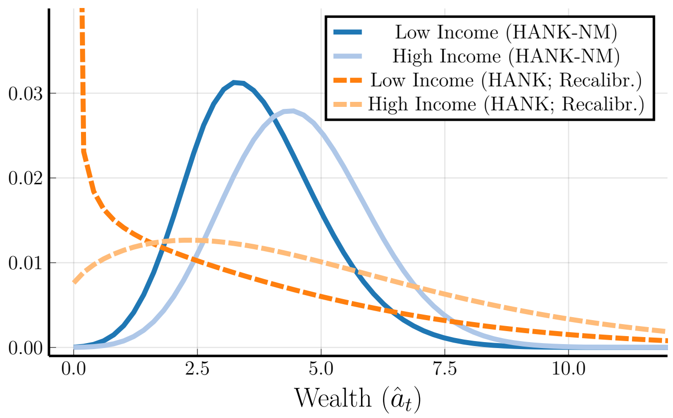
Figure 7: Impulse Responses: Expansionary Monetary Policy Shock



In response to these changes in inflation and interest rates, households economize on their cash holdings which can be seen in the decline in aggregate cash demand in Panel (d). Since the real return on cash is constant in HANK-NM-CBDC, households do not need to economize on cash due to inflation in that case. As a result, cash demand is almost unchanged on impact of the shock. The line for HANK has been omitted because there is no demand for cash in HANK. Since an expansionary monetary policy stimulates aggregate demand and increases output, real wages, and labor supply, there is an increase in savings. Initially, households mainly increase their deposit holdings, but after a few periods, cash holdings increase as well.

The most striking difference between the three models lies in the consumption response. Consumption in the centralized market in Panel (h) increases the most for HANK, while the increase in HANK-NM and HANK-NM-CBDC is more subdued. The reason for this difference lies in the transactional motive for savings that exists in both HANK-NM and HANK-NM-CBDC, which results in lower marginal propensities to consume. The transactional savings motive arises because cash and deposits can be used for transactions in the decentralized market. In HANK, households with a long spell of low productivity shocks optimally run down their savings until they hit the borrowing constraint, while households in HANK-NM or HANK-NM-CBDC never wholly use up their savings since they are required for purchases in the decentralized market.[19] This implies that compared to the model without a decentralized market, the whole distribution of households shifts to the right and away from the borrowing constraint, as seen in Figure 8. Moreover, since the liquidity shocks in the decentralized market introduce additional idiosyncratic risk into the model, households increase their precautionary savings, resulting in an additional distribution shift to the right.

Figure 8: Wealth Distribution



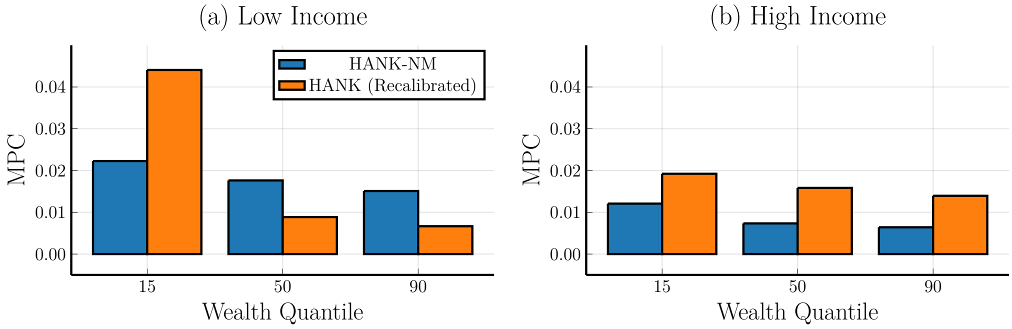
Notes: The y-axis limit has been constrained: 13.4% of the households in HANK are at the borrowing constraint.

Ultimately, this transactional motive implies that MPCs in the centralized market are lower in the HANK-NM than in HANK, as seen in Figure 9. The stark differences arise because only households with zero or close to zero wealth have very high MPCs in either model but HANK-NM features a much lower mass of households near the borrowing constraint. MPCs in the decentralized market are equal to one if the liquidity constraint becomes binding for buyers, and thus, there are cases where households can exhibit high MPCs in the model. However, the MPCs in HANK-NM tend to be lower than typically found in the empirical literature (Johnson et al., 2006; Parker et al., 2013; Broda and Parker, 2014), meaning that the model likely underestimates the effect of monetary policy on aggregates.

Returning to the impulse responses in Figure 7: Panel (i) shows that decentralized market

[19]For HANK-NM, this can be seen in Figure 3.

Figure 9: Marginal Propensities to Consume (MPCs) in the Centralized Market



Notes: Marginal propensities to consume (in the centralized market) out of an unexpected transfer $\Delta \hat{a}_t$. Aggregate MPCs are 0.017 in HANK-NM and 0.029 in HANK.

consumption falls in response to an expansionary monetary policy shock. The reason for this is twofold: First, the value of households' wealth a_{t-1} gets reduced on impact of the shock because of the unexpected reduction in the real return on cash and deposits. This slightly reduces the possibility for exchange in the decentralized market and leads to a slight fall in aggregate DM consumption. Second, after the impact of the shock, households adjust the composition of their portfolio and avoid holding cash to some extent to avoid the inflation tax. This further limits the exchange in the decentralized market, particularly in meetings where only cash is accepted by sellers, resulting in lower consumption in subsequent periods. However, decentralized market consumption increases as cash holdings are replenished and even exceed the steady state value after a few periods. Similarly, payments in the decentralized market, as shown in Panel (l), fall initially but increase in the medium term.

Looking at aggregates across centralized and decentralized markets, total output in Panel (j) increases slightly less in HANK-NM than in HANK because of a fall in decentralized market output.[20] Total consumption increases in all three models, but due to the more subdued response of centralized market consumption, total consumption does not increase as much in the HANK-NM and HANK-NM-CBDC. Furthermore, in response to an increase in wages, unions negotiate a higher labor supply, but all models' responses are similar. Finally, government spending strongly increases on impact of the shock because of a significant increase in tax revenue from labor taxes, as well as an increase in the seignorage from the implicit inflation tax. These revenue increases are enough to compensate for the decline in profit taxes due to the fall in profits in response to an expansionary monetary policy shock.

To summarize, the impulse responses to an expansionary monetary policy shock in HANK, HANK-NM, and HANK-NM-CBDC are quite similar in terms of the aggregate response of interest rates and inflation. However, because of the new transactional motive of sav-

[20] Note that DM consumption and DM output are equal to each other. Thus, Panel (h) in Figure 7 also shows the response of DM output.

ings in HANK-NM and HANK-NM-CBDC, households have lower marginal propensities to consume, which reduces the impact of an expansionary monetary policy shock on consumption. Furthermore, the same mechanisms in the decentralized market that lead to the super-non-neutrality of money in the long run also lead to changes in savings behavior in response to a monetary policy shock, which affects consumption in the decentralized market. Overall, the responses are broadly consistent with the VAR estimates in Christiano et al. (2005), meaning that an expansionary monetary policy shock increases output, consumption, investment, real wages, and inflation.

5.4 The Unequal Welfare Costs of Inflation

The previous analysis focused on the model economy's steady-state and dynamic behavior. Table 4 shows the welfare costs of 5% and 10% inflation relative to the zero inflation case for different households in the economy. In particular, the table shows how much consumption a household is willing to give up in every period to avoid having to live in an economy with 5% and 10% of inflation, respectively.

Table 4: Consumption Equivalent Welfare (CEW)

	Wealth Quantile						
	1	10	25	50	75	90	99
CEW (%) at 5% Inflation							
Low Income	−0.547	−0.540	−0.536	−0.528	−0.521	−0.513	−0.500
High Income	−0.533	−0.527	−0.522	−0.515	−0.507	−0.499	−0.485
CEW (%) at 10% Inflation							
Low Income	−1.039	−1.027	−1.019	−1.006	−0.992	−0.977	−0.952
High Income	−1.012	−1.000	−0.992	−0.978	−0.964	−0.950	−0.925

The median household is willing to give up around 0.5% of consumption every period to avoid having to live in an economy with 5% inflation. As the inflation rate is doubled to 10%, the welfare costs of inflation for median households also roughly double to around 1% in consumption equivalent terms.

Importantly, the welfare costs of inflation are more tilted towards wealth- and income-poor households in the economy. For example, while a high-income household at the 90th wealth percentile is willing to give up 0.499% of consumption every period, a low-income household at the 10th percentile is willing to give up 0.540%. This 4.1bp increase might not seem much but it implies that welfare costs are about 8% higher for an income-poor household at the 10th wealth percentile than for an income-rich household at the 90th wealth percentile. The gap in the welfare costs between the poor and the rich widens from 4.1bp to 7.7bp as inflation increases to 10%, but it remains roughly proportional to the overall size of the costs meaning that the welfare costs for the poor are still around 8% higher for the poor.

The differences in the welfare costs along the income and wealth distribution stem from the fact that poorer households, in line with data from the Encuesta Financiera de las Familias (EFF), depend more on cash for their consumption expenditures. As a result, increases in inflation and the associated economizing on cash become particularly costly for them. Additionally, these differences get reinforced by general equilibrium effects as discussed in Section 5.1. For example, the real deposit return increases, benefiting wealthier households and, therefore, reducing the negative effects for them.

6 Conclusions

In this paper, I developed a framework that combines a Heterogenous Agent New Keynesian (HANK) model with monetary search frictions as in Lagos and Wright (2005). In the resulting model, some economic activity occurs in a decentralized market, characterized by bilateral trading, while some activity occurs in a centralized market, resembling the setup in New Keynesian models. It provides a microfounded demand for money and captures the three functions of money as a unit of account, a medium of exchange, and a store of value.

I show that the model exhibits super-non-neutrality, meaning that changes in the central bank's inflation target affect real variables and the wealth distribution in the long run. These non-neutralities can be substantial, with a change in the inflation target from 0% to 5% implying a drop in aggregate consumption by 0.74%. While a representative agent version of the model would also exhibit super-non-neutrality, the more general preference specification of the heterogeneous agent version implies quantitatively larger effects on consumption since consumption falls in both markets, not just the decentralized one. Furthermore, I show that the super-neutrality can be restored if the central bank replaces cash with an interest-bearing CBDC with an interest rate that adjusts one-for-one with inflation. Finally, I show that the welfare costs of inflation are unequally distributed along the income and wealth distribution. Income-poor households at the 10th percentile of the wealth distribution experience about 8% higher welfare cost than income-rich households at the 90th percentile.

In future research, the framework developed in this paper could be applied to many questions at the intersection of innovations in payments, inequality, and monetary policy. Digitalization has brought several innovations in payments and money, such as mobile payment systems (e.g., Alipay, WeChat Pay) and novel forms of private money (e.g., stablecoins), which could affect the conduct of monetary policy (see e.g., Uhlig and Xie, 2021). Since the solution algorithm can be extended to include additional assets with little increase in computational complexity, questions that require a model with heterogeneity, microfoundations for media of exchange, and require both the analysis of transitional (or short-run) dynamics and long-run effects have come within reach.

References

Abad, J., G. Nuño, and C. Thomas (2022): "Implications of central bank digital currency for the operational framework of monetary policy," Tech. rep.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 109, 659–684.

Aruoba, S. B. (2011): "Money, Search and Business Cycles," *International Economic Review*, 52, 935–959.

Aruoba, S. B. and F. Schorfheide (2011): "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-offs," *American Economic Journal: Macroeconomics*, 3, 60–90.

Aruoba, S. B., C. J. Waller, and R. Wright (2011): "Money and capital," *Journal of Monetary Economics*, 58, 98–116.

Ascari, G. (2004): "Staggered prices and trend inflation: some nuisances," *Review of Economic Dynamics*, 7, 642–667.

Ascari, G. and L. Rossi (2012): "Trend Inflation and Firms Price-Setting: Rotemberg Versus Calvo," *The Economic Journal*, 122, 1115–1141.

Ascari, G. and A. M. Sbordone (2014): "The Macroeconomics of Trend Inflation," *Journal of Economic Literature*, 52, 679–739.

Bayer, C., B. Born, and R. Luetticke (2021): "The Liquidity Channel of Fiscal Policy," CEPR Press Discussion Paper 14883, CEPR.

——— (2022): "The Liquidity Channel of Fiscal Policy," Tech. rep.

Bethune, Z. and G. Rocheteau (2022): "Unemployment and the Distribution of Liquidity," Tech. rep.

Bewley, T. (1983): "A Difficulty with the Optimum Quantity of Money," *Econometrica*, 51, 1485–1504.

Bianchi, F., L. Melosi, and M. Rottner (2021): "Hitting the elusive inflation target," *Journal of Monetary Economics*, 124, 107–122.

Boppart, T., P. Krusell, and K. Mitman (2018): "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative," *Journal of Economic Dynamics and Control*, 89, 68–92.

Broda, C. and J. A. Parker (2014): "The Economic Stimulus Payments of 2008 and the aggregate demand for consumption," *Journal of Monetary Economics*, 68, S20–S36.

CHIU, J., S. M. DAVOODALHOSSEINI, J. HUA JIANG, AND Y. ZHU (2021): "Bank Market Power and Central Bank Digital Currency: Theory and Quantitative Assessment," Working Paper 2019-20, Bank of Canada.

CHIU, J. AND M. MOLICO (2010): "Liquidity, redistribution, and the welfare cost of inflation," *Journal of Monetary Economics*, 57, 428–438.

——— (2011): "Uncertainty, Inflation, and Welfare," *Journal of Money, Credit and Banking*, 43, 487–512.

——— (2021): "Short-run dynamics in a search-theoretic model of monetary exchange," *Review of Economic Dynamics*, 42, 133–155.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

DE FERRA, S., K. MITMAN, AND F. ROMEI (2020): "Household heterogeneity and the transmission of foreign shocks," *Journal of International Economics*, 124, 103303.

ERCEG, C. J., D. W. HENDERSON, AND A. T. LEVIN (2000): "Optimal monetary policy with staggered wage and price contracts," *Journal of Monetary Economics*, 46, 281–313.

EROSA, A. AND G. VENTURA (2002): "On inflation as a regressive consumption tax," *Journal of Monetary Economics*, 49, 761–795.

EUROPEAN CENTRAL BANK (2020): "Study on the payment attitudes of consumers in the euro area (SPACE)," Tech. rep., European Central Bank.

FERNÁNDEZ-VILLAVERDE, J., J. MARBET, G. NUÑO, AND O. RACHEDI (2021): "Inequality and the Zero Lower Bound," .

FERRARI, M. M., A. MEHL, AND L. STRACCA (2020): "Central Bank Digital Currency in an Open Economy," Working Paper Series 2488, European Central Bank.

FRIEDMAN, M. (1969): *The Optimum Quantity of Money and Other Essays.*

GORNEMANN, N., K. KUESTER, AND M. NAKAJIMA (2016): "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy," *International Finance Discussion Paper*, 2016, 1–40.

GROSS, J. AND J. SCHILLER (2021): "A Model for Central Bank Digital Currencies: Implications for Bank Funding and Monetary Policy," *SSRN Electronic Journal.*

HAGEDORN, M., I. MANOVSKII, AND K. MITMAN (2019): "The Fiscal Multiplier," Tech. rep., National Bureau of Economic Research, Cambridge, MA.

HUGGETT, M. (1993): "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 17, 953–969.

İmrohoroğlu, A. (1992): "The welfare cost of inflation under imperfect insurance," *Journal of Economic Dynamics and Control*, 16, 79–91.

Johnson, D. S., J. A. Parker, and N. S. Souleles (2006): "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review*, 96, 1589–1610.

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

Keister, T. and D. Sanches (2022): "Should Central Banks Issue Digital Currency?" *Review of Economic Studies*, forthcoming.

Kiyotaki, N. and R. Wright (1989): "On Money as a Medium of Exchange," *Journal of Political Economy*, 97, 927–954.

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

Lucas, Jr., R. E. (2000): "Inflation and Welfare," *Econometrica*, 68, 247–274.

Luetticke, R. (2019): "Transmission of Monetary Policy with Heterogeneity in Household Portfolios," .

Maliar, L., S. Maliar, and F. Valli (2010): "Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm," *Journal of Economic Dynamics and Control*, 34, 42–49.

McKay, A., E. Nakamura, and J. Steinsson (2016): "The Power of Forward Guidance Revisited," *American Economic Review*, 106, 3133–3158.

Molico, M. (2006): "The Distribution of Money and Prices in Search Equilibrium," *International Economic Review*, 47, 701–722.

Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013): "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, 103, 2530–2553.

Richter, A. W., N. A. Throckmorton, and T. B. Walker (2014): "Accuracy, Speed and Robustness of Policy Function Iteration," *Computational Economics*, 44, 445–476.

Rocheteau, G., P.-O. Weill, and T.-N. Wong (2021): "An heterogeneous-agent New-Monetarist model with an application to unemployment," *Journal of Monetary Economics*, 117, 64–90.

Rotemberg, J. J. (1982): "Sticky Prices in the United States," *Journal of Political Economy*, 90, 1187–1211.

Smets, F. and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Uhlig, H. and T. Xie (2021): "Parallel Digital Currencies and Sticky Prices," Working Paper 2020-188, Becker Friedman Institute.

Young, E. R. (2010): "Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm and non-stochastic simulations," *Journal of Economic Dynamics and Control*, 34, 36–41.

Zamora-Pérez, A. (2021): "The paradox of banknotes: understanding the demand for cash beyond transactional use," *ECB Economic Bulletin*, 121–137.

A Derivations

A.1 Optimality conditions

A.1.1 Decentralized Market

The first-order conditions of the bargaining problem in the decentralized market can be written as

$$\frac{u_c(\hat{c}_t^i)}{v_c(\hat{c}_t^i, \tilde{s}_t)} V_{a,t}^{CM}(\tilde{a}_{t-1} + \hat{x}_t^i, \tilde{s}_t) - V_{a,t}^{CM}(\tilde{a}_{t-1} - \hat{x}_t^i, \tilde{s}_t) = \hat{\mu}_t^i \,, \tag{2}$$

where $u_c(\hat{c}_t^i) = (\hat{c}_t^i + b)^{-\varsigma}$ is the marginal utility of consumption in the CM, $v_c(\hat{c}_t^i, \tilde{s}_t) = \frac{1}{\tilde{s}_t}$ is the marginal disutility of production in the DM, $V_{a,t}^{CM}$ is the derivative of the centralized market value function with respect to $\hat{a}_t$, and $\hat{\mu}_t^i$ is the Kuhn-Tucker multiplier associated with the (upper bound of the) liquidity constraint. Note that if there is an interior solution where the liquidity constraint is not binding we have that $\hat{\mu}_t^i = 0$. Equation (2) together with the (binding) participation constraint of the seller determine $\hat{c}_t^i$ and $\hat{x}_t^i$.

A.1.2 Centralized Market

The first-order conditions of the problem in the centralized market are

$$\beta \mathrm{E}_t \left[\frac{V_{m,t+1}^{DM}}{U_{c,t}} \right] + \frac{\mu_t^m}{U_{c,t}} = 1 \,, \tag{Cash Euler Eq.}$$

$$\beta \mathrm{E}_t \left[\frac{V_{d,t+1}^{DM}}{U_{c,t}} \right] + \frac{\mu_t^d}{U_{c,t}} = 1 \,, \tag{Deposit Euler Eq.}$$

where μ_t^x with $x \in \{m, d\}$ are the Kuhn-Tucker multipliers associated with the non-negativity constraints on m_t and d_t, $U_{c,t} = \Gamma c_t^{-\sigma}$ is the marginal utility of consumption at time t, $V_{m,t+1}^{DM}$ is the derivative of the DM value function with respect to m_t, and $V_{d,t+1}^{DM}$ is the derivative of the DM value function with respect to d_t.

From the envelope condition for the centralized market, we have that

$$V_{a,t}^{CM} = U_{c,t} \,.$$

Using the fact that the sellers' value function is the same as the value function in the centralized market in equilibrium, we have that

$$V_{m,t}^{DM} = \gamma V_{m,t}^b + (1-\gamma) V_{a,t}^{CM} \frac{R_{t-1}^M}{\pi_t} \,,$$

$$V_{d,t}^{DM} = \gamma V_{d,t}^b + (1-\gamma) V_{a,t}^{CM} \frac{R_{t-1}^D}{\pi_t} \,,$$

or

$$V_{m,t}^{DM} = V_{a,t}^{CM} \frac{R_{t-1}^M}{\pi_t} + \gamma \left(V_{m,t}^b - V_{a,t}^{CM} \frac{R_{t-1}^M}{\pi_t} \right) ,$$
$$V_{d,t}^{DM} = V_{a,t}^{CM} \frac{R_{t-1}^D}{\pi_t} + \gamma \left(V_{d,t}^b - V_{a,t}^{CM} \frac{R_{t-1}^D}{\pi_t} \right) .$$

Using these expressions, we can rewrite the FOCs above as follows

$$\gamma\beta \mathrm{E}_t \left[\frac{V_{m,t+1}^b - V_{a,t+1}^{CM} \frac{R_t^M}{\pi_{t+1}}}{U_{c,t}} \right] + \beta \mathrm{E}_t \left[\frac{U_{c,t+1}}{U_{c,t}} \frac{R_t^M}{\pi_{t+1}} \right] + \frac{\mu_t^m}{U_{c,t}} = 1 \,, \qquad \text{(Cash Euler Eq.)}$$
$$\gamma\beta \mathrm{E}_t \left[\frac{V_{d,t+1}^b - V_{a,t+1}^{CM} \frac{R_t^D}{\pi_{t+1}}}{U_{c,t}} \right] + \beta \mathrm{E}_t \left[\frac{U_{c,t+1}}{U_{c,t}} \frac{R_t^D}{\pi_{t+1}} \right] + \frac{\mu_t^d}{U_{c,t}} = 1 \,, \qquad \text{(Deposit Euler Eq.)}$$

Further rearranging yields the following Euler equations

$$\beta \mathrm{E}_t \left[\frac{U_{c,t+1}}{U_{c,t}} \frac{(1+\ell_{t+1}^m) R_t^M}{\pi_{t+1}} \right] + \frac{\mu_t^m}{U_{c,t}} = 1 \,, \qquad \text{(Cash Euler Eq.)}$$
$$\beta \mathrm{E}_t \left[\frac{U_{c,t+1}}{U_{c,t}} \frac{(1+\ell_{t+1}^d) R_t^D}{\pi_{t+1}} \right] + \frac{\mu_t^d}{U_{c,t}} = 1 \,. \qquad \text{(Deposit Euler Eq.)}$$

where the liquidity premia on cash and deposits, ℓ_{t+1}^m and ℓ_{t+1}^d, are defined as

$$\ell_{t+1}^m = \gamma \left(\frac{V_{m,t+1}^b}{V_{a,t+1}^{CM} \frac{R_t^M}{\pi_{t+1}}} - 1 \right) , \qquad \text{(Cash Liquidity Premium)}$$
$$\ell_{t+1}^d = \gamma \left(\frac{V_{d,t+1}^b}{V_{a,t+1}^{CM} \frac{R_t^D}{\pi_{t+1}}} - 1 \right) . \qquad \text{(Deposit Liquidity Premium)}$$

These liquidity premia arise because cash and deposits can be used as a medium of exchange in the decentralized market and, therefore, implicitly yield a higher return than captured by R_{t-1}^M and R_{t-1}^D.

Recall the value function of the buyer

$$V_t^b(m_{t-1}, d_{t-1}, s_t) = \sum_{i=1}^{3} \left\{ \vartheta_i \int \left[u(\hat{c}_i(\xi_t, \tilde{\xi})) + V_t^{CM} \left(a_{t-1} - \hat{x}_i(\xi_t, \tilde{\xi}), s_t \right) \right] dF(\tilde{\xi}) \right\} .$$

Let

$$\Lambda_t^i = u(\hat{c}_t^i) + V_t^{CM} \left(a_{t-1} - \hat{x}_t^i, s_t \right)$$

where for notational simplicity the dependence on the state variables for $\hat{c}_t^i$ and $\hat{x}_t^i$ has been omitted. However, they are still to be understood as functions of the state variables. We have that

$$\frac{\partial \Lambda_t^i}{\partial y} = u_c(\hat{c}_t^i)\frac{\partial \hat{c}_t^i}{\partial y} + V_{a,t}^{CM}(a_{t-1} - \hat{x}_t^i, s_t)\left(\frac{\partial a_{t-1}}{\partial y} - \frac{\partial \hat{x}_t^i}{\partial y}\right),$$

where $y \in \{m_{t-1}, d_{t-1}\}$. From the participation constraint in the bargaining problem, we have that

$$\frac{\partial \hat{c}_t^i}{\partial y} = \frac{V_{a,t}^{CM}(\tilde{a}_{t-1} + \hat{x}_t^i, \tilde{s}_t)}{v_c(\hat{c}_t^i, \tilde{s}_t)}\frac{\partial \hat{x}_t^i}{\partial y}.$$

Substituting this back into the expression for $\frac{\partial \Lambda_t^i}{\partial y}$ yields

$$\frac{\partial \Lambda_t^i}{\partial y} = \left[\frac{u_c(\hat{c}_t^i)}{v_c(\hat{c}_t^i, \tilde{s}_t)}V_{a,t}^{CM}(\tilde{a}_{t-1} + \hat{x}_t^i, \tilde{s}_t) - V_{a,t}^{CM}(a_{t-1} - \hat{x}_t^i, s_t)\right]\frac{\partial \hat{x}_t^i}{\partial y} + V_{a,t}^{CM}(a_{t-1} - \hat{x}_t^i, s_t)\frac{\partial a_{t-1}}{\partial y}.$$

From Equation (2), we know that the term in parenthesis is zero as long as the bargaining problem has an interior solution, i.e. the liquidity constraint is not binding. Furthermore, we have that $\frac{\partial \hat{x}_t^i}{\partial m_{t-1}} = \frac{\partial a_{t-1}}{\partial m_{t-1}} = \frac{R_{t-1}^M}{\pi_t}$ and $\frac{\partial \hat{x}_t^i}{\partial d_{t-1}} = \frac{\partial a_{t-1}}{\partial d_{t-1}} = \frac{R_{t-1}^D}{\pi_t}$ if $\hat{x}_t^i$ is constrained by the upper bound in the bargaining problem, and $\frac{\partial \hat{x}_t^i}{\partial y} = 0$ if $\hat{x}_t^i$ is constrained by the lower bound. Therefore, $\frac{\partial \hat{x}_t^i}{\partial y}$ does not need to be computed at any stage to evaluate $\frac{\partial \Lambda_t^i}{\partial y}$, which greatly simplifies the implementation.

B Solution Algorithm

In the following, I discuss algorithms to solve the model's steady state and to solve for transitional dynamics after MIT shocks or between steady states. The proposed algorithms build on a policy function iteration approach (see e.g. Richter et al., 2014) combined with the method of Young (2010) to keep track of the distribution of households as a histogram. This choice allows for a very accurate solution of the household's policy functions and does not suffer from Monte Carlo noise when propagating the distribution of households.

B.1 Steady State

At the core, the algorithm used to solve for the steady state of the model is similar to solving standard HANK models. However, due to the fact that the households take the whole distribution into account when solving their dynamic programming problem, one needs to use a procedure where for given prices the distribution and policies are updated iteratively.

The general algorithm to solve the steady state of the model is described in Algorithm 1.[21] To discretize the state space, I define grids of points for assets brought into the CM $\hat{a}_t \in [0, \hat{a}_{\max}]$, and the idiosyncratic state $s_t \in \{s^n\}_{n=1}^N$. To capture the curvature of policy functions near the borrowing constraint, I use a polynomial rule similar to Maliar et al. (2010) to place more grid points for $\hat{a}_t$ near zero. The idiosyncratic states follow from the discretization of an AR(1) process to a Markov chain using the Rouwenhorst method. The household problem in the centralized market is then solved on this grid. Whenever it is necessary to evaluate policies off the grid, I use linear interpolation due to its speed and robustness.

Algorithm 1 (Steady State)

1. *Make an initial guess for the cash policy function $m'(\hat{a}_t, s_t)$, the deposit policy function $d'(\hat{a}_t, s_t)$, the CM value function $V^{CM}(\hat{a}_t, s_t)$ and the distribution $F(\hat{a}_t, s_t)$.*
2. *Guess the aggregate cash demand $M = 0$.*
3. *Guess the (nominal) return on bonds R^B.*
4. *Compute aggregate variables based on R^B and M.*
5. *For each grid point $(\hat{a}_t, s_t)$, solve the household problem in the CM which yields updated policies $m'(\hat{a}_t, s_t)$, $d'(\hat{a}_t, s_t)$, and $V^{CM}(\hat{a}_t, s_t)$.*
6. *Propagate the asset distribution $F(\hat{a}_t, s_t)$ from the current CM to the next CM using the method of Young (2010) and the updated policies.*[22]

[21] Since aggregate variables are constant in the steady state, I drop their time subscripts.

[22] Propagating the distribution from CM to CM will also require DM policies. In particular, it will require knowing $\hat{x}_t$ for each meeting a household could have. One can solve the household problem in the DM on a grid for total assets brought into the DM $a_{t-1} \in [0, a_{\max}]$. It is convenient to compute the payment policy for a bargaining problem without the liquidity constraint $\hat{x}^u(a_{t-1}, s_t, \tilde{a}_{t-1}, \tilde{s}_t)$ from which it is straightforward to compute the policies in the constrained case: if a household does not have enough

7. *Update M based on the updated $F(\hat{a}_t, s_t)$.*

8. *Iterate on steps 4 – 7 until the maximum difference between successive updates of the policy functions and distributions ($m'(\hat{a}_t, s_t)$, $d'(\hat{a}_t, s_t)$, $V^{CM}(\hat{a}_t, s_t)$, and $F_t(\hat{a}_t, s_t)$) is less than a given degree of precision.*

9. *Check whether R^B clears the bond market. If the bond market does not clear, update the guess for R^B and go to step 3.*

Some comments are in order regarding how to solve the CM problem. While there are some degrees of freedom in how this problem is solved exactly, for computational tractability it is crucial that expectations for the Euler equation are evaluated based on the distribution $F(\hat{a}_t, s_t)$, i.e. the distribution that persisted before the CM choice was made, which is lower dimensional than the distribution $F(m_t, d_t, s_t)$, i.e. the distribution that persists after the CM choice has been made. To get cash $\tilde{m}_t$ and deposits $\tilde{d}_t$ of the counterparties for each point in $(\tilde{\hat{a}}_t, \tilde{s}_t)$ one can simply evaluate the (old) CM policy functions as needed whenever expectations are computed.

The proposed approach also has a computational advantage due to the fact that the asset distribution is propagated from CM to CM directly instead of being first propagated to the DM and then in an additional step to the next CM. This is, again, due to the fact that one needs to evaluate a much lower amount of potential meetings in the DM. Put in a different way, the proposed algorithm is computationally more tractable because it economizes on evaluating DM meetings that have a very low probability of occurring and instead focuses on meetings that are more likely to occur.

B.2 Transition

Solving for the transitional dynamics after an MIT shock or the transitional dynamics between steady states is very similar to solving the model's steady state. The difference is that we are not trying to find a single value of R^B and M that equilibrate the system but a whole path of values along a perfect-foresight transition path. Furthermore, due to the fact that inflation is not necessarily constant along the transition path it turns out to be more convenient to guess transition paths for π_t, K_t, H_t, and M_t from which the transition of all other aggregate variables can be computed.

The algorithm is similar to the one described in Boppart et al. (2018) but adapted to the fact that distribution affects policy functions directly. It proceeds as described in Algorithm 2. Crucial for computational tractability is again that one only keeps track of $F_t(\hat{a}_t, s_t)$ instead of keeping track of $F_t(m_{t-1}, d_{t-1}, s_t)$.

assets that are accepted in a meeting to pay $\hat{x}^u$, it simply pays as much as it has available. The policies for consumption in the DM can be computed from the participation constraint of the seller. If this policy is not precomputed on a grid, the bargaining problem can be solved whenever $\hat{x}^u$ needs to be evaluated. While this is more costly, it is also more accurate.

Algorithm 2 (Transition)

1. *Choose a time T at which the economy is assumed to have reached the steady state.*

2. *Guess a path for inflation π_t, capital K_t, aggregate labor supply H_t, and aggregate cash holdings M_t, i.e. $\{\pi_t, K_t, H_t, M_t\}_{t=0}^{T}$. Furthermore, guess a path for the distribution $\{F_t(\hat{a}_t, s_t)\}_{t=0}^{T}$ and counterparty's policies for cash $\{\tilde{m}'_t(\hat{a}_t, s_t)\}_{t=0}^{T}$ and deposits $\{\tilde{d}'_t(\hat{a}_t, s_t)\}_{t=0}^{T}$.*

3. *Solve the value and policy functions backward from $t = T-1, \ldots, 1$ assuming that time T value and policy functions correspond to the ones in the steady state.*[23]

4. *Update the paths for the distribution $\{F_t(\hat{a}_t, s_t)\}_{t=0}^{T}$ and counterparty policies $\{\tilde{m}'_t(\hat{a}_t, s_t)\}_{t=0}^{T}$ and $\{\tilde{d}'_t(\hat{a}_t, s_t)\}_{t=0}^{T}$ by iterating forwards from $t = 1, \ldots, T$ using the updated path of policy functions from the previous step.*

5. *Back out the implied paths for inflation π_t, capital K_t, aggregate labor supply H_t, and aggregate cash holdings M_t from the updated distributions, the wage inflation equation, and the Taylor rule.*

6. *Compute the maximum difference between the implied paths for $\{\pi_t, K_t, H_t, M_t\}_{t=0}^{T}$ and their guesses. Stop the algorithm if the maximum difference is less than a given degree of precision.*

7. *Update the guess $\{\pi_t, K_t, H_t, M_t\}_{t=0}^{T}$ by taking a weighted average of the old guess and the implied paths. Go to step 3.*

[23] This part of the algorithm proceeds analogously to solving for the steady state. For accuracy, the DM bargaining problem is solved for every meeting that is evaluated as part of the expectations individually.

C Extensions

C.1 Rotemberg Adjustment Cost without Indexation

The baseline setup indexes the Rotemberg adjustment cost to inflation for two reasons: First, it allows isolating the effects of monetary super-non-neutrality due to the medium of exchange role of money. Second, the super-non-neutrality results are not dependent on the choice of nominal rigidities (i.e., Calvo vs. Rotemberg). Ascari and Rossi (2012) have shown that the long-run New Keynesian Philips curve implies a positive relationship between the inflation target of the central bank and output in the case of Rotemberg adjustment costs, while there is a negative relationship if nominal rigidities are modeled using the Calvo approach.

Table 5: Comparative Statics without Indexation

	Inflation				
	2%	5%	10%	15%	20%
Centralized Market					
Output (Y_t)	0.06	0.15	0.29	0.42	0.55
Consumption (C_t)	−0.13	−0.31	−0.54	−0.73	−0.87
Government Spending (G_t)	0.27	0.64	1.17	1.63	2.02
Taxes ($\mathcal{T}_t$)	0.05	0.11	0.18	0.25	0.29
Wealth ($\hat{A}_t$)	−0.17	−0.35	−0.54	−0.62	−0.63
Cash (M_t)	−2.82	−6.13	−10.61	−14.22	−17.19
Deposits (D_t)	−0.02	−0.01	0.07	0.20	0.37
Capital (K_t)	−0.01	0.00	0.03	0.08	0.15
Labor Supply (H_t)	0.09	0.23	0.42	0.59	0.75
Wage (w_t)	0.01	0.03	0.08	0.15	0.21
Real Deposit Rate (R^D_{t-1}/π_t)	0.01	0.03	0.06	0.08	0.10
Decentralized Market					
Consumption ($\hat{C}_t$)	−0.52	−1.23	−2.22	−3.04	−3.72
Payment ($\hat{X}_t$)	−0.66	−1.55	−2.78	−3.78	−4.60
Relative Price ($\hat{P}_t$)	−0.14	−0.32	−0.57	−0.77	−0.93
Cross-Market Aggregates					
Output ($\mathcal{Y}_t$)	0.01	0.02	0.06	0.11	0.16
Consumption ($\mathcal{C}_t$)	−0.23	−0.54	−0.96	−1.30	−1.57

Notes: Change relative to zero-inflation steady state. All changes are expressed in percent except for interest rates, which are expressed in percentage points.

I consider a specification with Rotemberg adjustment cost not indexed to inflation in this section. This means that the New Keynesian Philips curve becomes

$$\log(\pi_t) = \beta \mathrm{E}_t \left[\log(\pi_{t+1}) \frac{Y_{t+1}}{Y_t} \right] + \frac{\varepsilon_y}{\theta_y} \left(mc_t - \frac{\varepsilon_y - 1}{\varepsilon_y} \right),$$

and the wage inflation equation becomes

$$\log\left(\pi_t^w\right) = \beta \mathrm{E}_t\left[\log\left(\pi_{t+1}^w\right)\frac{H_{t+1}}{H_t}\right] + \frac{\varepsilon_w}{\theta_w}\left(\frac{\Upsilon'(H_t)}{U'(C_t)} - (1-\tau_h)\frac{\varepsilon_w - 1}{\varepsilon_w}w_t\right).$$

Table 5 shows the comparative static results if the Rotemberg adjustment costs are not indexed to inflation. The super-non-neutrality due to nominal rigidities dampens the negative effect of inflation on consumption and leads to a sufficiently large increase in CM output that total output increases even though DM output falls. CM output rises because markups fall (marginal costs mc_t increase), and the distortion due to monopolistic competition is reduced as the inflation target is increased. Considering the results of Ascari and Rossi (2012), it would likely be the case that nominal rigidities à la Calvo would reinforce the negative long-run relationship between inflation and output in the baseline setup.

Table 6: Consumption Equivalent Welfare (CEW) without Indexation

	Wealth Quantile						
	1	10	25	50	75	90	99
CEW (%) at 5% Inflation							
Low Income	−0.431	−0.428	−0.427	−0.421	−0.417	−0.411	−0.402
High Income	−0.415	−0.412	−0.410	−0.405	−0.400	−0.394	−0.384
CEW (%) at 10% Inflation							
Low Income	−0.814	−0.810	−0.807	−0.798	−0.790	−0.780	−0.762
High Income	−0.783	−0.778	−0.773	−0.765	−0.756	−0.746	−0.728

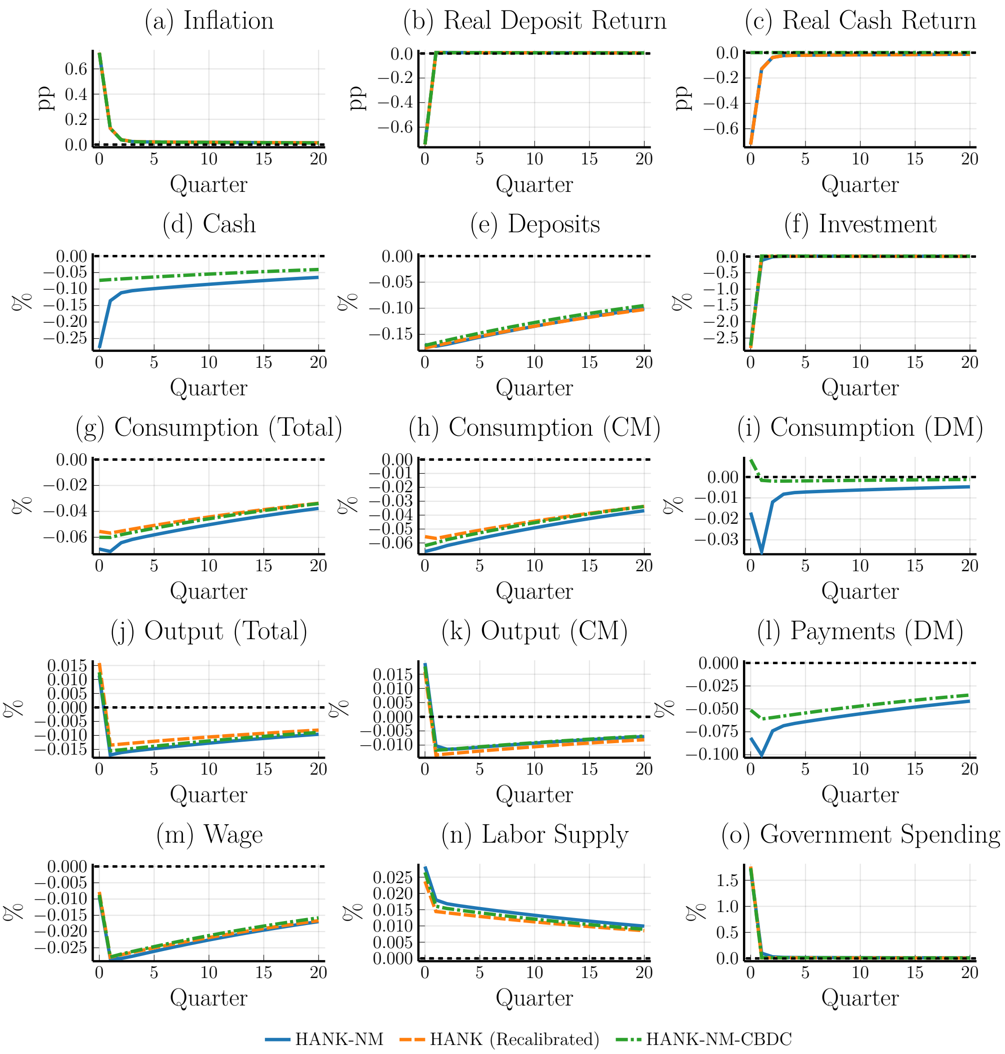
Table 6 shows the welfare cost of inflation. Relative to the baseline setup, the welfare costs of inflation are lower since higher inflation targets reduce the distortion from monopolistic competition among retailers.

C.2 No Nominal Rigidities

Figure 10 shows the impulse response to an expansionary monetary policy shock without nominal rigidities ($\theta_y = \theta_w = 0$). A monetary policy shock in this framework has a real effect even in the absence of nominal rigidities for two reasons: First, since nominal interest rates are predetermined, the increase in inflation due to the monetary policy shock reduces the (ex-post) real return on cash and deposits meaning that households have fewer resources available on the impact of the shock unless fiscal policy is designed to offset this decline. This channel drives the responses in HANK. Second, the medium of exchange role of money has similar effects in response to inflation as in the steady state. Households economize on cash which limits DM exchange and, in turn, consumption in the decentralized as inflation increases. This channel and the previous one are active in HANK-NM.

Overall, Figure 10 seems to suggest that nominal rigidities are necessary to make the

Figure 10: Impulse Responses: Expansionary Monetary Policy Shock ($\theta_y = \theta_w = 0$)



IRFs consistent with VAR estimates (see, e.g., Christiano et al., 2005). For example, consumption falls counterfactually in response to an expansionary monetary policy shock without nominal rigidities.

C.3 HANK with Baseline Parametrization

For the HANK model in the main text, β, $\bar{E}$, χ, B, and Γ have been recalibrated to match the same aggregate moments as HANK-NM. Another potential reference could be the HANK model with the same parametrization as the HANK-NM but with $\gamma = 0$. Figure 11, Figure 12, and Figure 13 show the results are qualitatively not dependent on the (re-)calibration: Without the decentralized market, households have higher MPCs, and consumption responds more strongly to the expansionary monetary policy shock.

Figure 11: Impulse Responses: Expansionary Monetary Policy Shock (incl. HANK)

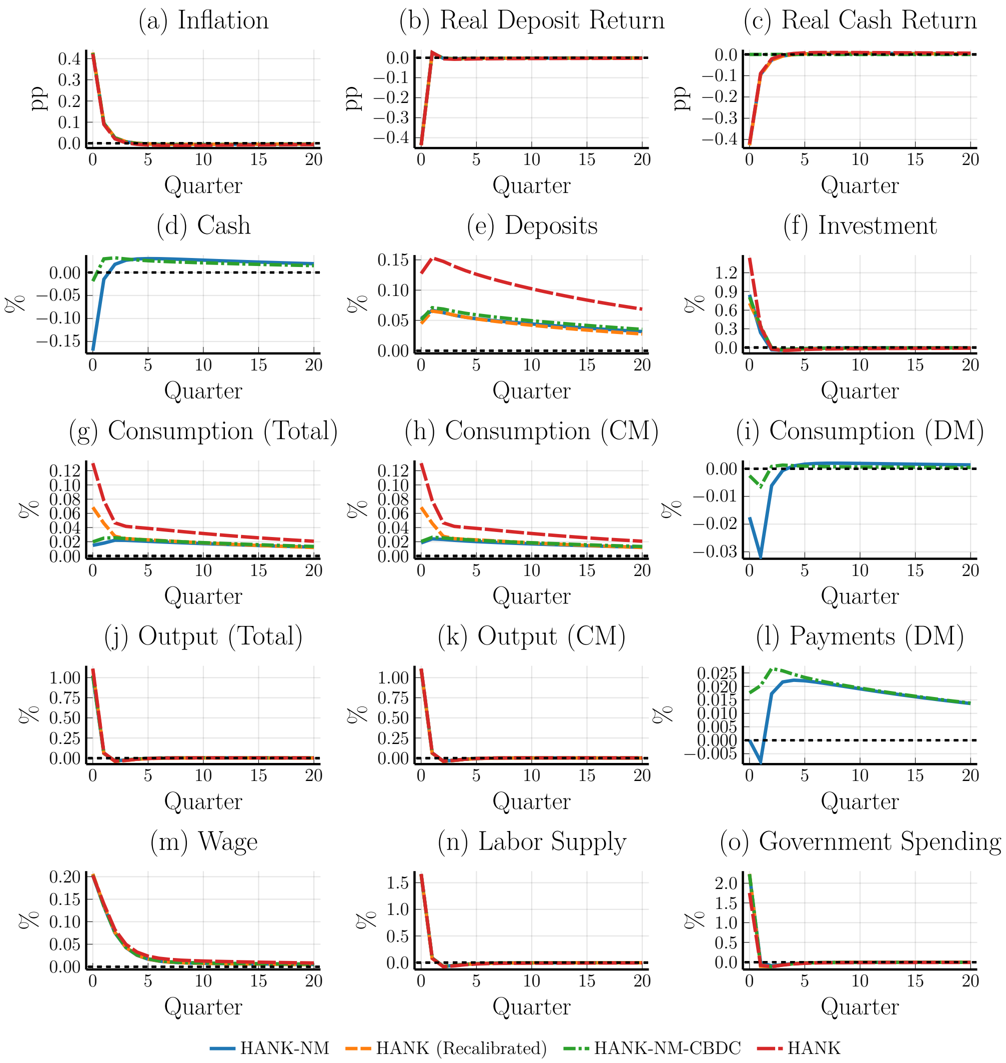
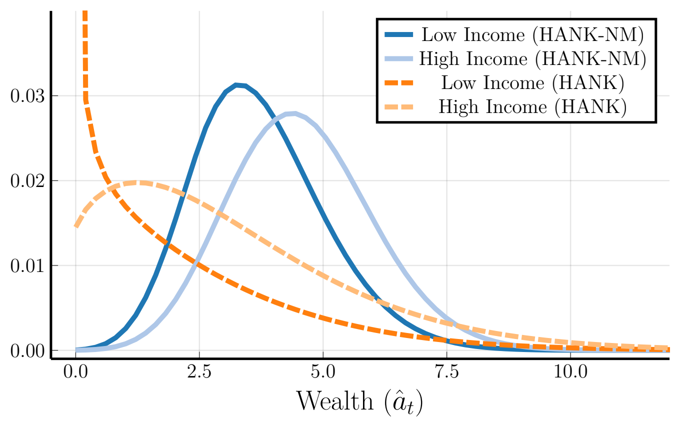
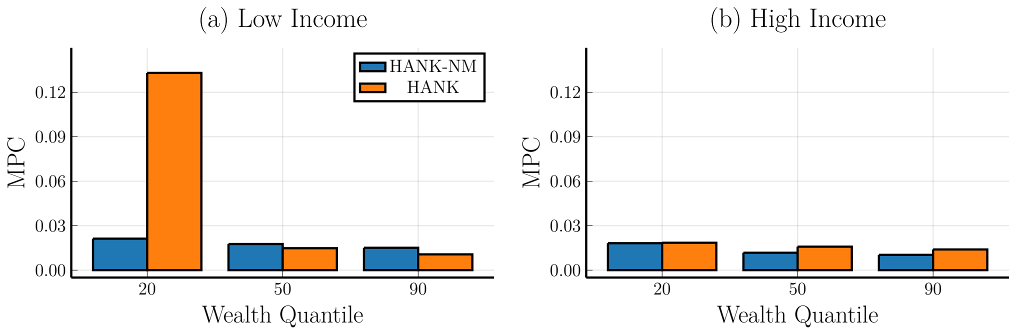


Figure 12: Wealth Distribution



Notes: The y-axis limit has been constrained: 19.8% of the households in HANK are at the borrowing constraint.

Figure 13: Marginal Propensities to Consume (MPCs) in the Centralized Market



Notes: Marginal propensities to consume (in the centralized market) out of an unexpected transfer $\Delta\hat{a}_t$. Aggregate MPCs are 0.017 in HANK-NM and 0.068 in HANK.

C.4 Representative Agent Model

The model developed in the main text does not nest its representative agent version because the decentralized market generates idiosyncratic risk even if labor income risk is reduced to zero. To get a representative agent version of the model, preferences need to be quasilinear, as in Lagos and Wright (2005). This assumption implies that all households choose to hold the same amount of cash and deposits at the end of the centralized market, i.e., the wealth distribution becomes degenerate.

The general environment and the household problem stay the same, except for the change in preferences and the fact that the labor supply decision is not delegated to a labor union, i.e., there are no wage rigidities. Preferences in the representative agent version become

$$U(c_t, h_t) = \Gamma\left(U(c_t) - \Upsilon(h_t)\right) = \Gamma\left(\frac{c_t^{1-\sigma}}{1-\sigma} - \chi h_t\right).$$

Table 7 shows the comparative statics for the model with quasilinear preferences. The model has been calibrated to the same aggregate moments and uses the same parameters for the externally calibrated parameters. Due to the quasilinearity assumption, inflation does neither affect consumption in the centralized market nor the real wage or the real deposit rate. Thus, the consumption aggregate across both markets only falls due to the decline in decentralized market consumption. While the fall in DM consumption is almost exactly the same as the one in the heterogeneous agent version of the model, aggregate consumption falls two-thirds less since CM consumption does not respond to changes in inflation.

In contrast to the model in the main text, households accumulate more wealth in response to a higher inflation target. While they reduce their holdings to a similar extent, they increase their deposit holdings substantially in the representative agent version. This leads to an increase in capital, and a stronger increase in output, allowing for a substantial increase in government spending. However, the increase in capital is matched by an increase in labor supply such that the real wage and real deposit rate stay constant.

The welfare costs of inflation are larger in the representative agent version of the model, as shown in Table 8. At 10% inflation, the consumption equivalent welfare cost of 1.296% are close to the estimates in Lagos and Wright (2005), who estimate them to be 1.4% for the case where buyers have all bargaining power.

Table 7: Comparative Statics with Representative Agent

	Inflation				
	2%	5%	10%	15%	20%
Centralized Market					
Output (Y_t)	0.24	0.58	1.06	1.48	1.84
Consumption (C_t)	0.00	0.00	0.00	0.00	0.00
Government Spending (G_t)	0.46	1.08	1.98	2.75	3.43
Taxes ($\mathcal{T}_t$)	0.22	0.52	0.95	1.33	1.65
Wealth ($\hat{A}_t$)	0.43	1.02	1.89	2.65	3.31
Cash (M_t)	−2.81	−6.52	−11.62	−15.74	−19.16
Deposits (D_t)	0.62	1.46	2.68	3.73	4.65
Capital (K_t)	0.24	0.58	1.06	1.48	1.84
Labor Supply (H_t)	0.24	0.58	1.06	1.48	1.84
Wage (w_t)	0.00	0.00	0.00	0.00	0.00
Real Deposit Rate (R^D_{t-1}/π_t)	0.00	0.00	0.00	0.00	0.00
Decentralized Market					
Consumption ($\hat{C}_t$)	−0.53	−1.24	−2.20	−2.99	−3.63
Payment ($\hat{X}_t$)	−0.53	−1.24	−2.20	−2.99	−3.63
Relative Price ($\hat{P}_t$)	0.00	0.00	0.00	0.00	0.00
Cross-Market Aggregates					
Output ($\mathcal{Y}_t$)	0.19	0.44	0.82	1.14	1.43
Consumption ($\mathcal{C}_t$)	−0.10	−0.23	−0.41	−0.56	−0.68

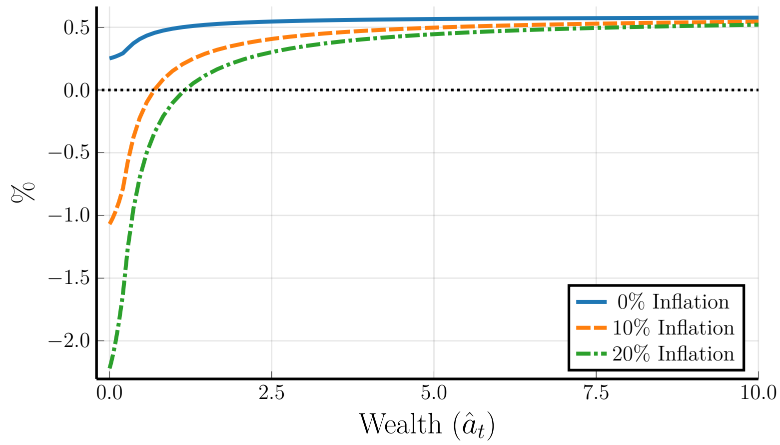
Notes: Change relative to zero-inflation steady state. All changes are expressed in percent except for interest rates, which are expressed in percentage points.

Table 8: Consumption Equivalent Welfare (CEW) with Representative Agent

	Inflation				
	2%	5%	10%	15%	20%
CEW	−0.296	−0.701	−1.296	−1.812	−2.266

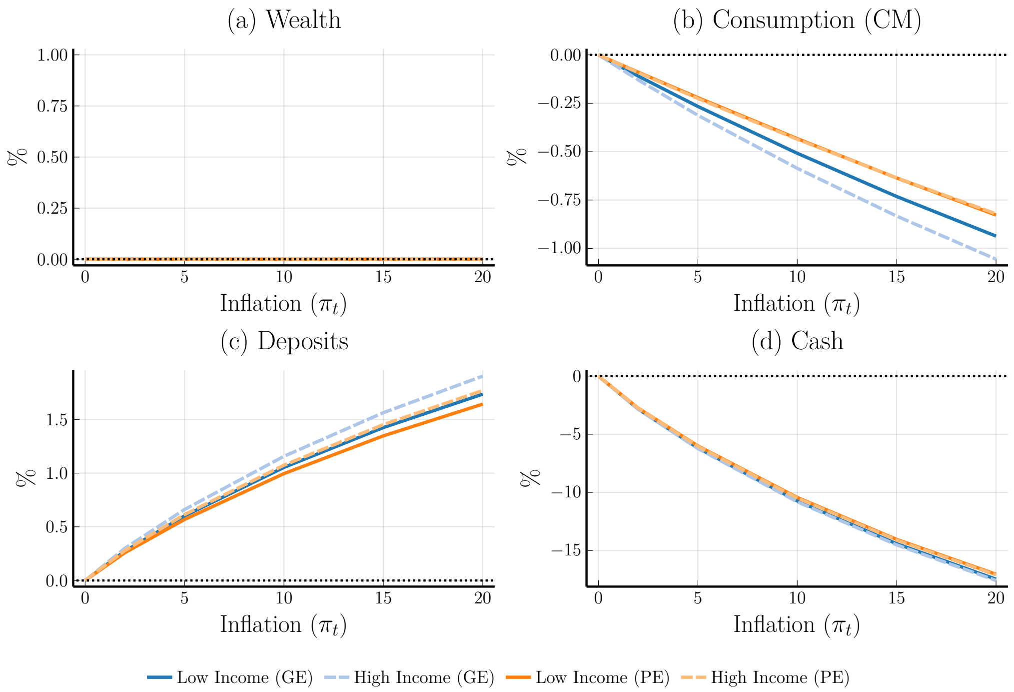
D Additonal Figures

Figure 14: Effective Returns on Savings



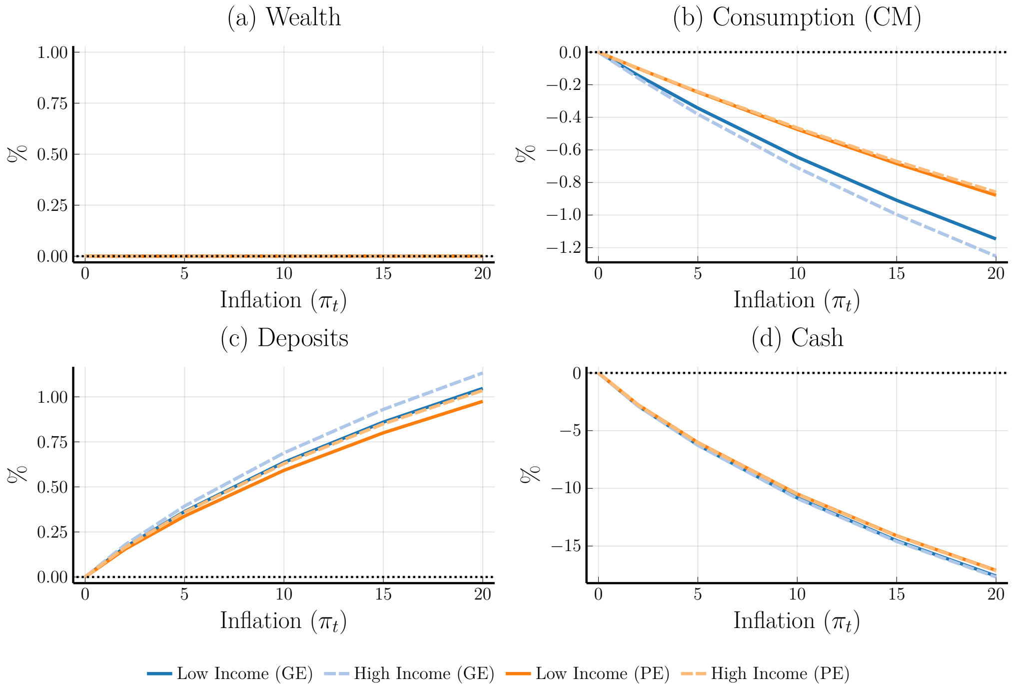
Notes: Effective returns on savings R^E_{t-1}/π_t have been computed for the portfolios of low-income households in general equilibrium.

Figure 15: Change in Consumption, Cash, and Deposits (10th Wealth Quantile)



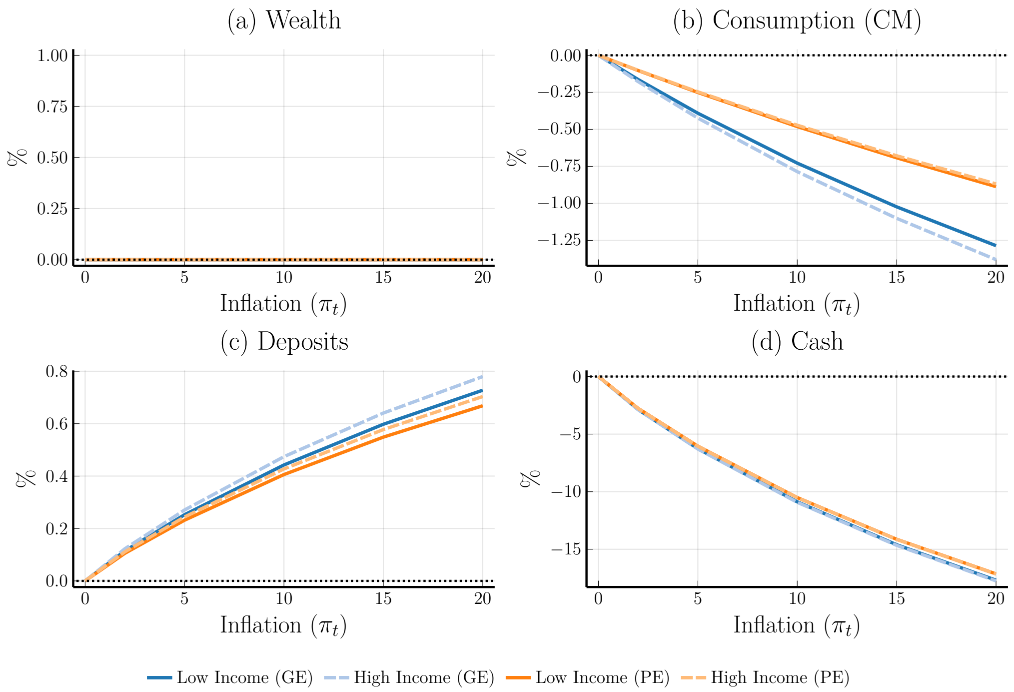
Notes: Consumption, deposit, and cash policies with wealth fixed at the 10th quantile of the wealth distribution under zero inflation. This does not take into account that households tend to accumulate less wealth under higher inflation targets.

Figure 16: Change in Consumption, Cash, and Deposits (50th Wealth Quantile)



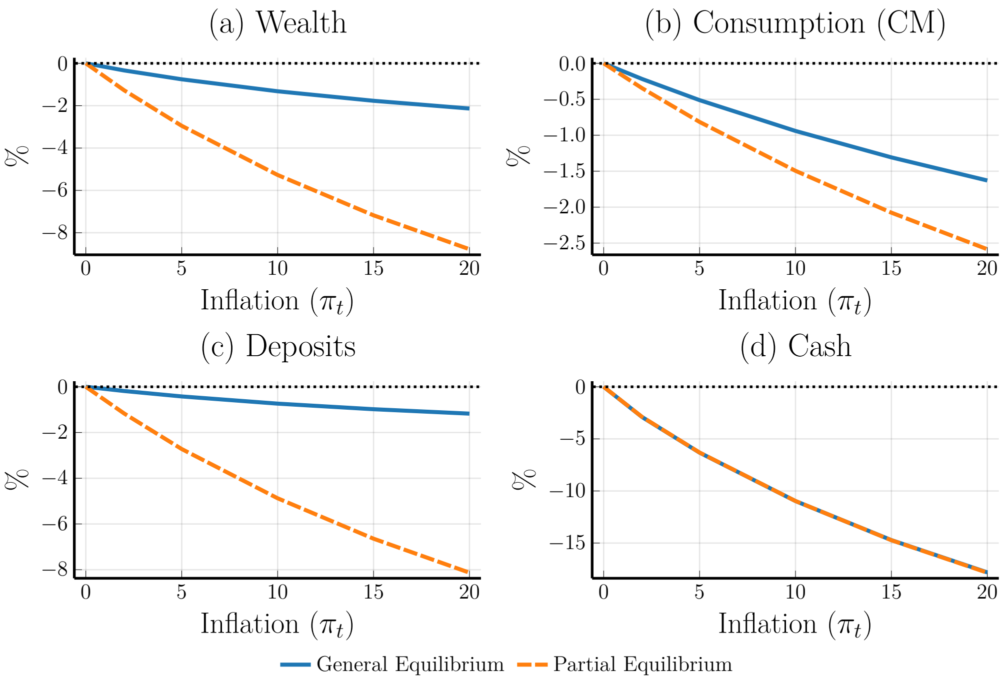
Notes: Consumption, deposit, and cash policies with wealth fixed at the 50th quantile of the wealth distribution under zero inflation. This does not take into account that households tend to accumulate less wealth under higher inflation targets.

Figure 17: Change in Consumption, Cash, and Deposits (90th Wealth Quantile)



Notes: Consumption, deposit, and cash policies with wealth fixed at the 90th quantile of the wealth distribution under zero inflation. This does not take into account that households tend to accumulate less wealth under higher inflation targets.

Figure 18: Change in Consumption, Cash, and Deposits (Aggregate)



Notes: In partial equilibrium, individual consumption, deposit, and cash policies are aggregated using the wealth distribution implied by these policies.