Welfare Costs of Inflation
Across the Money Holdings Distribution

Andres Ghini[*]

NYU

November 16, 2023

Updated regularly, please click here for the latest version

Abstract

This paper studies the welfare costs of anticipated inflation emphasizing a stark feature of the U.S. economy: around 60% of the U.S. households do not hold any interest-bearing liquid financial asset. This is particularly harmful in economic environments with high inflation rates that poses a threat to this large proportion of the U.S. households. In this paper, I explore the consequences of anticipated and persistent inflation and its unequal distribution across money holdings for this particular type of households. To do this, I develop a search-theoretic model of money where agents are ex-ante heterogeneous in their productivity level, and money serves two roles: as a medium of exchange and as an instrument for precautionary savings. The model generates a non-degenerate distribution of money in steady-state that matches the one observed in the data. This last point is crucial to assess the unequal costs of inflation across the money holdings distribution. The main result shows that welfare costs of inflation are higher than in the traditional money search models in the literature. Also, these costs are larger for higher income households whose saving capacity is particularly deteriorated, making it harder for them to hedge against income shocks. Finally, these results are sensitive to the way money is being injected into the economy: costs are higher when money is issued to finance government spending rather than via lump-sum transfers, which indicates strong distributive effects.

1   Introduction

The COVID-19 pandemic triggered a succession of economic challenges across the globe, with one of its most enduring consequences being the surge in inflation rates. The widespread disruptions in supply chains, coupled with fluctuations in demand patterns, led to a complex inflationary landscape that has persisted as a paramount issue for countries worldwide. Despite concerted efforts by governments and central banks to address various fiscal and monetary measures, the task of curbing inflation remains formidable. Consequently, the persistent inflationary pressures represent a critical concern, underscoring the enduring and far-reaching impacts of the pandemic on global economic dynamics. In light of this world-wide economic situation, this papers studies what are the welfare costs of high and persistent inflation and how are these distributed on a particularly vulnerable group of households: those who do not hold any interest-bearing liquid financial asset. In this sense, these particular households are the most exposed to the threats of high and persistent inflation, as they cannot hedge against it. Surprisingly enough, the share of this unique group of households is large for the U.S. economy: almost two-thirds of U.S. households fall into this category.

Among the literature of monetary economics, extensive effort has been made to address these relevant questions: what are the welfare cost of inflation and what are the channels through which inflation harms or benefits agents inside the economy.

Within the theoretical models used to answer these questions, we can distinguish two different approaches. On the one hand, *reduced-form* monetary models where agents face cash-in-advance constraints or directly include money in the utility function (Lucas, 2000). And, on the other hand, models that provide *micro foundations* for monetary economics based on search theory where money is essential for trading (Lagos and Wright, 2005). Despite being intrinsically different, both types of models were able to predict and quantify the welfare costs of inflation and allowed for useful monetary policy insights. However, much of this literature has focused on *aggregate* results. In this sense, little is known about the implications of having endogenous and persistent heterogeneity across agents. In this context, agents with heterogeneous portfolios and earnings might be hit differently by changes in the inflation rate, not only potentially changing the aggregate results of previous research but

also, and equally important, opening a venue for analyzing redistributive effects within the economy and the impact of inflationary processes on inequality. This paper explores the distributional consequences of anticipated inflation when agents differ in money holdings and in productivity levels.

I argue that anticipated inflation distorts households' ability to save for precautionary reasons, especially for higher income individuals, whereas for agents with lower income and money holdings, the real balance effect is more prominent, for which higher inflation leads to a decrease of the average real money balances. My argument is based on two empirical facts. First, more than one-third of U.S. households hold all their liquid assets in checking deposits, this share being particularly higher for those at the lower end of the wealth distribution. Secondly, around sixty percent of U.S. households do not hold any interest-bearing asset that would hedge them against inflation. Surprisingly, this proportion has been stable even for periods with higher inflation, including the past three years, in which inflation has been specially high. Anticipated inflation, therefore, both erodes households' real money holdings and reduces their incentive to save.

More precisely, in line with the empirical evidence, individuals with higher financial wealth tend to allocate their wealth towards high-interest assets whereas individuals in the bottom of the distribution hold their entire wealth in bank accounts that have a nominal interest rate on its deposits that is nearly zero, i.e. checking and saving accounts. Since these accounts represent poor stores of value in the presence of inflation, agents at the bottom of the financial wealth distribution are more exposed to the harms of inflation while agents with higher levels of financial wealth can hedge against it. One of the reasons why individuals hold all their financial wealth only in the most liquid accounts is because money is essential for transaction purposes. The presence of liquidity needs force all agents to hold some amount of cash -or near cash accounts- in order to conduct transactions. Yet, this constraint is more binding for those households at the bottom of the distribution, since they are closer to a hand-to-mouth situation. In contrast, for individuals closer to the top of the distribution, even when they consume more than poor agents, can buffer against liquidity shocks by holding some cash and also can mitigate the effects of inflation by holding interest-bearing assets which have a higher return. Nevertheless, I document that a large share of U.S. households do not hold such

assets. Interestingly, these very particular type of households are not poor at all . In fact, fifty-percent of households in the fourth quintile of the liquid financial wealth distribution fall into this category.

I formalize this argument using a search-theoretic model of money with heterogeneous agents that matches the money holdings distribution in the data. I use the model to quantify the costs of inflation and how these costs are distributed across such distribution. To capture the use of money holdings, agents trade sequentially in a frictionless centralized market (CM) and in a frictional decentralized market (DM) that generates the necessity of using money for transactional purposes. In this latter market, random bilateral anonymous meetings take place between agents who can trade special goods for money. Individuals are also ex-ante heterogenous in productivities corresponding to the production of the general good in the CM. I use this model to study changes in steady-state welfare due to inflation. The model predicts that a permanent increase in the rate of inflation ruins household's ability to buffer against liquidity and income shocks, distorting optimal savings-consumption decisions. In this sense, average welfare costs of inflation are higher than in previous studies, where the precautionary savings motive is absent. This result advocates for grater financial integration. Moreover, my results show that inflation hits differently to agents across the distribution of money holdings. In this sense, individuals who are closer to the top of the distribution bear a larger burden of the costs, since their saving capacity gets particularly deteriorated, making it harder for them to hedge against income shocks. Interestingly, this effect offsets the fact that high inflation distorts the consumption-savings decision of the agents, which induces agents towards higher consumption, in expense of a lower savings rate. Finally, not only the aggregate welfare costs of inflation, but also its distribution across households, crucially depends on how money is being injected into the economy. In this regard, if lump-sum transfers are used, the poorest and more liquidity-constrained agents lose the least -or even gain- welfare, as some resources get redistributed out of the wealthiest agents in the economy. Alternatively, if the fiscal-monetary authority uses the new issuance of money to finance its expenditures, average welfare costs are much larger for the bottom quintiles of the money holdings distribution.

Section 2 presents evidence on households' portfolios and deposit interest rates. Section 3

3

develops the model, Section 5 characterizes the steady state, and reveals the mechanics of the decentralized trading. Section 6 studies the cost and consequences of an increase in inflation in the long run.

Related Literature

There is a vast literature of search-theoretic models of monetary exchange that study the equilibrium properties of economies where money is essential for trading. Most influential papers of this sort include for instance Kiyotaki and Wright (1989), Trejos and Wright (1995), Shi (1995), Lagos and Wright (2005) and Rocheteau and Wright (2005). The key feature of these examples is that they either have to restrict the divisibility of assets or goods exchanged in the markets, or assume certain mechanism that eliminates the heterogeneity intrinsically generated by decentralized trading. For instance, Lagos and Wright (2005) assume that agents can participate in a Walrasian market right after having exchanged in a bilateral meeting. The preferences of these individuals in this centralized market are assumed to be quasi-linear. As a result, wealth effects are absent and every agent chooses the same amount of money for the next decentralized market. The distribution of money holdings is then degenerate at the beginning of every period. The critical reason for imposing this restrictive assumption is to solve the households' problem in a manageable manner because if the distribution of money holdings were non-degenerate, agents must take into account the whole distribution and its evolution over when solving their optimization problem. My paper departs from these assumptions and tackles the consequences of accounting for the persistent heterogeneity observed in the data.

These environments mentioned above have the advantage of being analytically tractable, yet they are not suitable for studying the implications of having heterogeneous agents in terms of money holdings in search-based monetary economies. In this regard, another generations of search-theoretic models of monetary economics such as Molico (2006), Chiu and Molico (2010), Chiu and Molico (2011) and Menzio et al. (2013) tackle this issue. The first of these papers studies the distributive consequences of different money injection mechanisms by only considering a sequence of bilateral meetings that happen one after the other. The absence of a Walrasian market forces individuals to store up large sums of money as a

4

precautionary measure for future liquidity shocks. In such an environment, it is shown that changes in the growth rate of money are neutral if new money is injected via proportional transfers, but not if it is injected by lump-sum transfers, as they generate a distributive effect. Furthermore, under the class of lump-sum transfers, an increase in the rate of money expansion tends to decrease the dispersion of money holdings and prices, and to improve welfare when inflation is low; however, when inflation is high enough, the opposite effect results. Chiu and Molico (2010) and Chiu and Molico (2011) extend this model along the lines of the Lagos-Wright alternating markets structure. Their results show that distributional effects matter for welfare as represantative agents models would tipically over-estimate the costs of inflation. Additionally, the welfare effects of inflation are non-linear in the inflation rate. In Menzio et al. (2013), they assume a directed search mechanism, instead of bilateral trading. In this sense, buyers know the term of trade before visiting a particular seller. This introduces the advantage of having a Block-Recursive equilibrium, i.e. policy functions and value functions do not depend on the aggregate distribution of money. By this means, they circumvent the problem of having to deal with the distribution of money as an aggregate state variable. However, none of this papers match the resulting distribution of money holdings from the model to the empirical one in the data. On a more recent set of papers, Rocheteau et al. (2018) includes heterogeneity in a model that is still analytically tractable using a Lagos-Wright environment where agents face constraints on their labor supply choices. As long as labor supply constraint binds, wealth effects start playing a role and contribute to having a non-degenerate distribution. The model exhibits one-sided heterogeneity since the roles of buyer and seller are permanent and exogenously imposed, thus there is no reason for sellers to hold money. On their part, Bustamante (2018) study the distributional implications of open market operations on a search theoretical monetary model based on Lagos and Wright (2005) but with persistent and endogenous heterogeneity in money holdings. In this economy agents cannot fully self-insure against idiosyncratic liquidity shocks and the heterogeneity in their marginal utilities makes nominal illiquid bonds useful. The presence of heterogeneity across agents and the existence of a fully fledged distribution of prices suggest that a one-time expansive open market operation has enough room to generate short-run non-neutralities. In particular, since individuals are not able to re-balance their portfolio instantaneously

after a liquidity shock, the distribution of assets and the distribution of prices would slowly respond to the increased liquidity. Lastly, Chiu and Molico (2021) deal with the short-run effects of monetary policy in a search-theoretic monetary model in which agents are subject to idiosyncratic liquidity shocks as well as aggregate monetary shocks. This is, they study the role of the endogenous non-degenerate distribution of liquidity, liquidity constraints, and decentralized trade, for the transmission and propagation of monetary policy shocks where money is injected through lump-sum transfers. My paper contributes to all the mentioned works in two margins. First, none of this papers match the distribution of money holdings from the model to the empirical one in the data, which is key to correctly quantify the costs of inflation in an heterogeneous economy. Secondly, I introduce persistence into households' earnings. This is particularly important in preventing households from easily shifting across the spectrum of money holdings, as it is the case in the mentioned previous works. In these search theoretic models, agents' positions in this distribution are entirely determined by the sequence of idiosyncratic liquidity shocks they encounter, dictating whether they become buyers or sellers in the decentralized market. With the addition of a productivity shock and its persistence, households don't transition between "poor" and "rich" states in terms of money holdings as frequently, and these states aren't solely determined by liquidity shocks. This feature is crucial in determining the welfare cost of inflation.

When it comes to the study of the effect of inflation on households' savings capacity. In a seminal work, İmrohoroğlu (1992) examines the welfare cost of inflation in an economy where agents hold money in order to smooth consumption in the face of income variability for which there is no insurance. This generates a welfare cost of inflation several times larger than the one studied previously in the transaction cost literature. When it comes to environments that include other assets apart from money, Kocherlakota (2002) studies the welfare effects of having illiquid nominal risk-free bonds in monetary economies. It is shown that individuals are better off when they can engage in additional intertemporal trades of money even though both money and nominal bonds have no intrinsic value. However, in the case that bonds are liquid, there are no welfare gains since there is no difference between bonds and money. Another example includes Erosa and Ventura (2000) where individuals allocate assets between capital and money and perform transactions with cash or costly credit.

In this scenario, money is a poor store of value since it is dominated by capital in rate of return, but agents hold money because of a cash in advance type of constraint. This paper suggests that inflation has important distributional effects since it is a regressive consumption tax. As a result, welfare costs of inflation can be higher for low-income individuals. On a more recent work, Cirelli (2023) studies the welfare cost of anticipated inflation in a model of non-competitive banks along with households that vary in financial sophistication, i.e. the possibility to get access to high-interest assets. Nevertheless, these papers either assume a reduced form cash-in-advance constraint or a cashless economy.

Finally, this paper is closely related to the research exploring the impact of household heterogeneity on the transmission and execution of monetary policy. This field of study has primarily relied on New Keynesian models in which households facing exogenous, uninsurable idiosyncratic earnings risk engage in centralized markets within cashless economies that feature nominal rigidities, for instance in Kaplan et al. (2018), Auclert et al. (2018), Auclert et al. (2020) and Auclert et al. (2021). The primary focus of this body of work has been on investigating the short-term stabilization properties of monetary policy, rather than considering its long-term effects in the context of explicit money demand, which is a central aspect of my examination. Moreover, in their papers, monetary policy is executed through a Taylor rule without actively altering the public provision of liquidity, which sets my research apart.

2   Motivating Facts

In this section, using the Survey of Consumer Finances (SCF), I document the facts that motivate my modeling choices and inform the quantitative analysis. First, I document that historically, around forty percent of U.S. households have stored all their transaction assets in the form of checking deposits and prepaid cards. Secondly, I show that checking and saving accounts constitutes a sizable share of total financial assets for the average U.S. household. In this sense, I document that almost two-thirds of U.S. households do not hold any interest-bearing financial asset. Altogether, I argue that nominal interest rates on checking and saving

deposits have remained low, and close to zero, for almost fifteen years now. [1]

Transaction Motive

Using the SCF from years 2001 to 2022, I look into the transaction account of households, defined by the survey as the sum of five components: checking accounts, saving accounts, call accounts, money market accounts and prepaid cards. I focus on the most liquid account so as to emphasize its role for transaction services.

Figure 1 shows that about forty percent of U.S. households hold all their transaction account in the form of checking deposits and prepaid cards. Interestingly, despite having considerable fluctuations in the inflation rate throughout this period, this proportion has remained mostly steady. That is, even when the opportunity cost of holding checking accounts is high, households do not appear to flee from their most liquid account, which manifest the necessity of liquidity for transaction purposes.

Figure 1: Share of households who only hold checking accounts and prepaid cards
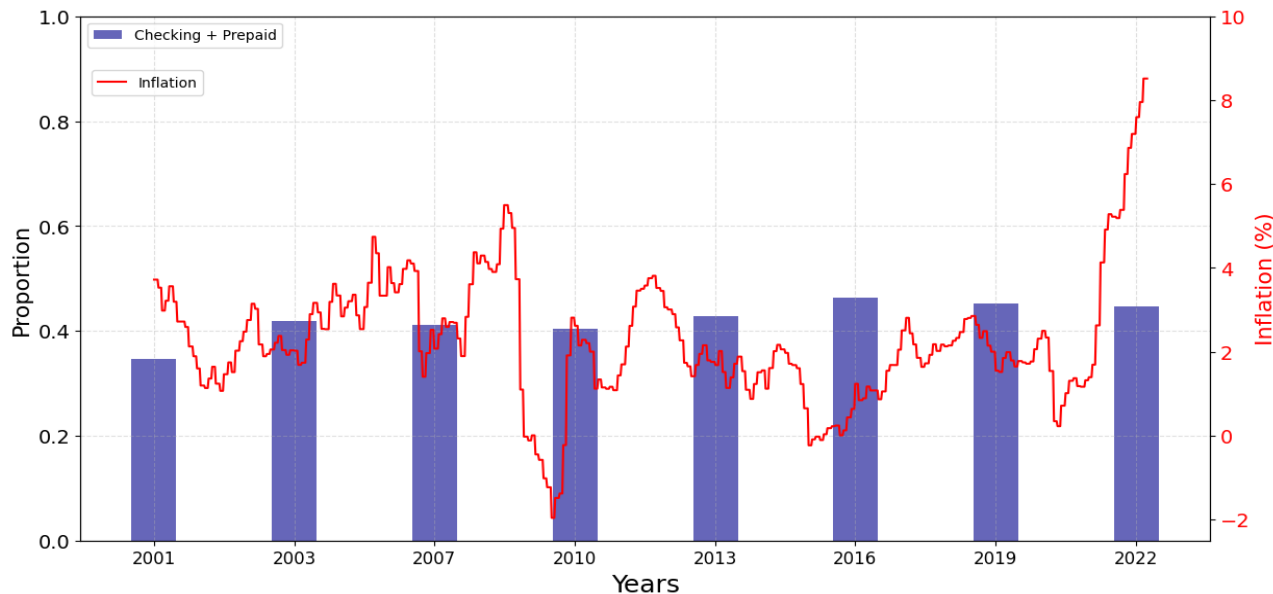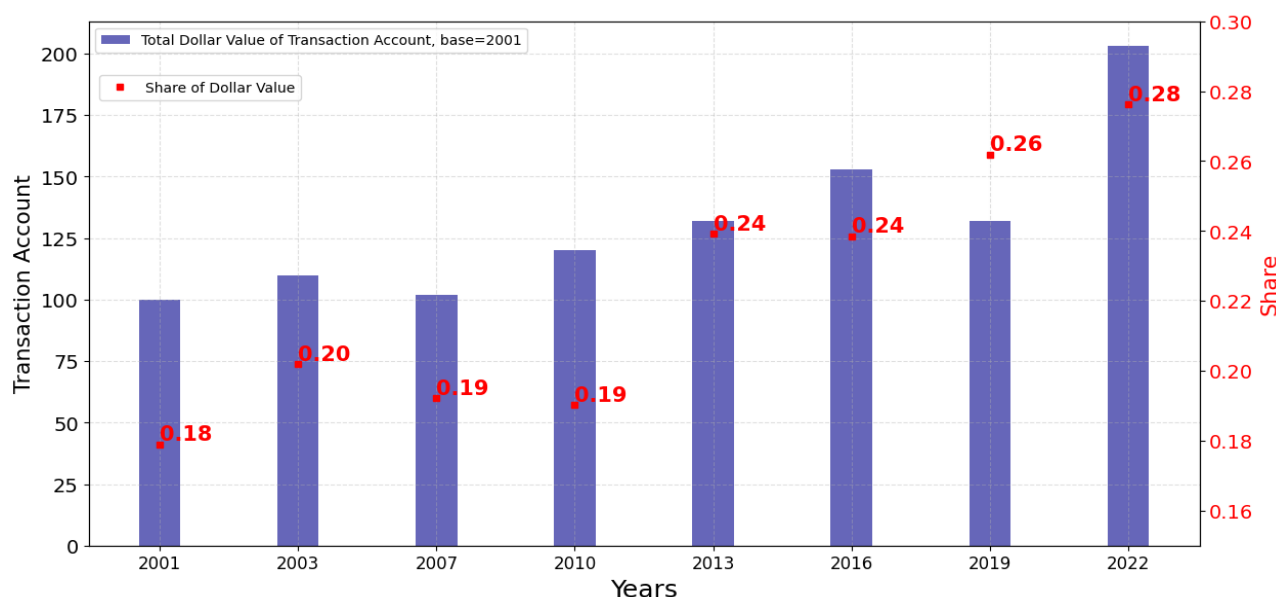

Note: The data come from the Survey of Consumer Finances and FRED, Federal Reserve Bank of St. Louis.

What is more, the share of the total dollar value in the transaction account that these households who only own checking accounts and prepaid cards, if anything, has been increasing

---

[1]For details on the SCF, sample selection and computations see Appendix A.

during the period covered. In fact, Figure 2 shows the total dollar value of the transaction account for each wave of the SCF, taking the year 2001 as the base year, so its total dollar value is equal to 100. Additionally, Figure 2 displays the share of the total dollar value in the transaction account that is own by the households who hold all their transaction account in the form of checking deposits and prepaid cards. It can be seen that, even when the total dollar-value of the transaction account has been growing over time, the share of it owned by this particular type of households has also increased, up to 28% in 2022, which indicates that their holdings are sizable inside this account.

Figure 2: Dollar-value hare of households who only hold checking accounts and prepaid cards

Note: The data come from the Survey of Consumer Finances.

Savings Motive

I will now focus on the role played by these most liquid accounts as saving vehicles. The focus on liquid assets is driven by recent findings, which highlight liquid assets as the primary type of assets that households rely on to maintain their consumption when they face unexpected changes in income. I define liquid assets as the whole universe of financial assets in the SCF excluding pension funds, life insurance, and other managed and miscellaneous assets. Thus, this definition of liquid assets includes, not only the entire transaction account, but also

certificate of deposits, directly held investment funds, saving bonds, directly held stocks and directly held bonds.

The data suggest that a large proportion of household hold all their liquid assets in the form of checking accounts, saving accounts and prepaid cards, despite these accounts having nearly zero nominal returns, as it will be shown later. That is, even when these accounts represent poor stores of value and do not offer the possibility to buffer against losses from inflation, households do not appear to abandon these accounts, not even for other financial liquid assets that bear higher rates such as money market accounts, bonds or investment funds. To make this point more explicit, Figure 3 shows that about sixty percent of U.S. households hold all their financial liquid assets only in the form of checking deposits, saving deposits and prepaid cards. Again, despite having considerable fluctuations in the inflation rate throughout this period, this proportion has remained mostly steady.

Figure 3: Share of households who only hold checking accounts, saving accounts and prepaid cards
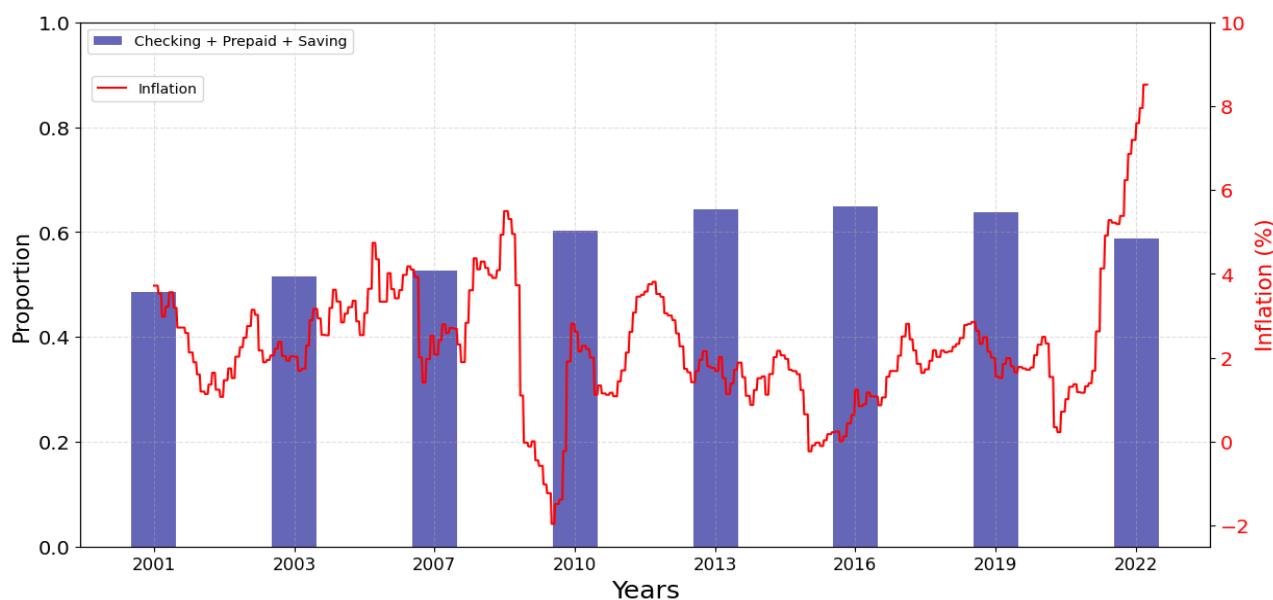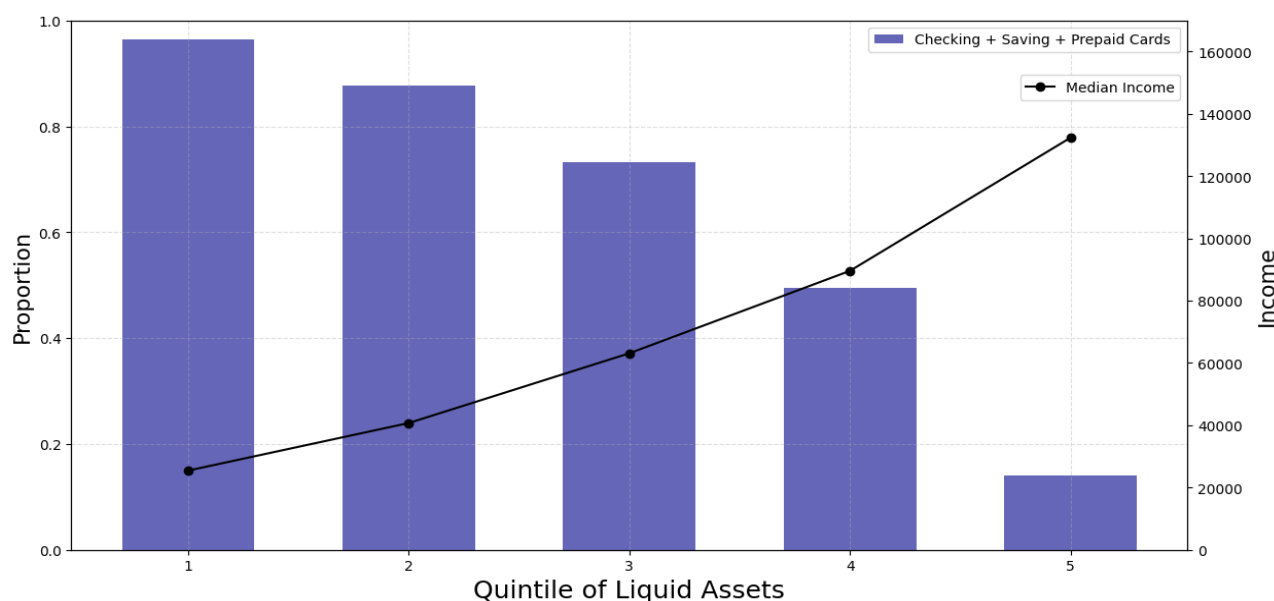

Note: The data come from the Survey of Consumer Finances and FRED, Federal Reserve Bank of St. Louis.

I will now restrict my attention to determining the composition of these households in terms of liquid wealth and income. For such purpose, I will delve into the SCF for the year 2019. This particular year exhibits averages values for inflation and nominal returns on assets in line with the time span being considered, as well as returns to checking and

saving deposits that are nearly zero. Figure 4 splits the 2019 population into liquid assets quintiles and calculates the share of households who have all their liquid wealth in terms of checking accounts, saving accounts and prepaid cards in each quintile. Additionally, it shows the median income of each group. [2] It is an striking fact that these particular households are not poor. In fact, 50% of households in the fourth quintile do not hold any asset at all apart from the three mentioned. Moreover, for this specific quintile, the median income is well above the median income of the entire U.S. population. This result is indicative of the fact that a large number of U.S. households use these three accounts as saving vehicles, as they do now hold any other liquid financial asset.

Figure 4: Share of households who only hold checking accounts, saving accounts and prepaid cards by quintile of liquid assets

Note: The data come from the Survey of Consumer Finances 2019. Average liquid assets over income for households in the bottom quintile is less than one week, for those in the middle, approximately four months, and in the top quintile, more than one year of household income.

Deposit Rates

While it is true that positive inflation represents a cost for those households who have checking and saving deposits, it is not the only cost they face. Regardless of the level of inflation,

_____

[2]This pattern is robust to the choice of year.

these households face an opportunity cost related the interest rate on other liquid assets. In other words, having deposits becomes costly when the return on them is less than what can be earned from a market asset with comparable features. Figure 5 displays data for the time span analyzed before, of the interest rates for the two different types of deposits: checking and savings accounts. It also includes the fed funds rate as a reference point for the safe return that households can achieve in the market. It is interesting to remark that after the Great Recession, deposit rates on both checking and saving accounts have remained steadily at zero. This is not a feature of a zero-lower-bound (ZLB) episode in the U.S., in fact, it can be noticed in the figure that when in 2016 Fed Funds rates start to increase, neither the rate on checking deposit nor the rate on saving deposit change in the same way. To reinforce this argument, the same can be concluded from the post COVID-19 period.

Figure 5: Deposit Rates and Fed Funds Rate
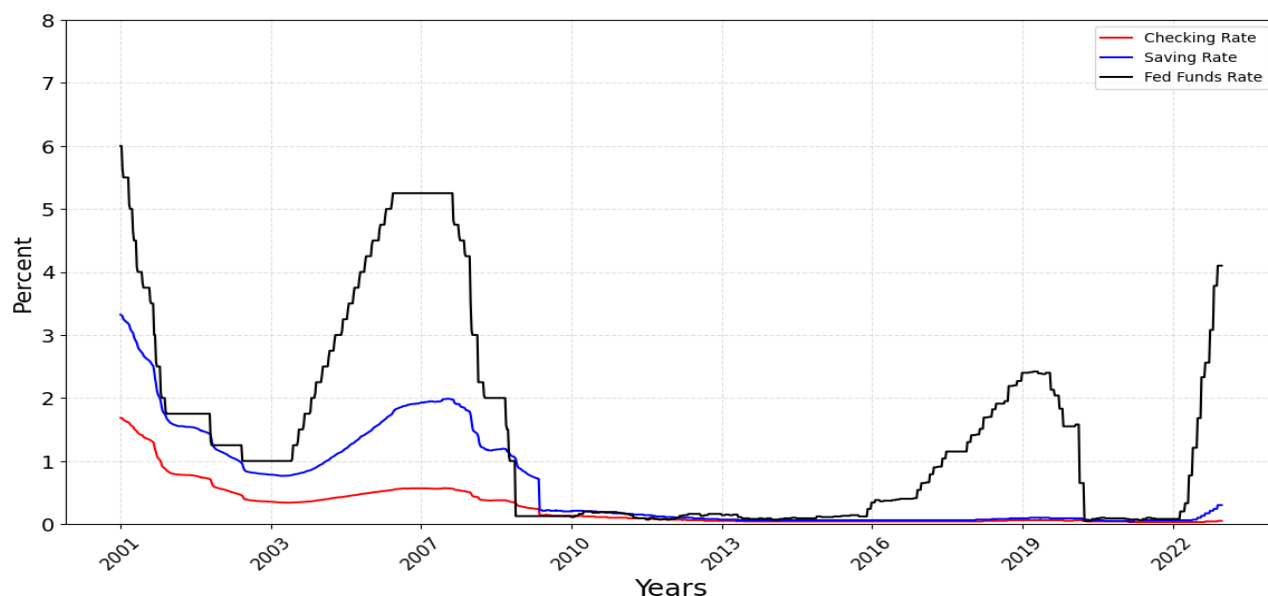
Note: The data come from Drechsler et al. (2017) and FRED, Federal Reserve Bank of St. Louis.

To summarize, for all the time span being considered here, approximately 40% of U.S. households, only hold checking deposits plus prepaid cards in the transaction account, which is indicative of the transaction role played by them. Moreover, 60% of all households, maintain all their liquid financial assets in the form of checking accounts, saving accounts and prepaid cards, that is to say, in zero-interest accounts, being this their only savings vehicle. In this

sense, almost two-thirds of the U.S. households do not hold any interest-bearing financial asset. What is striking is that these shares have remained stable, even in periods when inflation was higher or deposit rates on other assets were considerable larger. Finally, these households are not all poor; one-half of the households in the fourth liquid assets quintile lies inside this group of particular households.

3 The Model

Environment

The model presented here is based on money search models as in Lagos and Wright (2005), and Bewley-Aiyagari economies, (Bewley (1983), Aiyagari (1994)) but with a few critical differences that will be appropriately pointed out. Time is discrete. There is a $[0, 1]$ continuum of agents who live forever and discount the future with factor $\beta \in (0, 1)$. Each period is divided into two sub-periods (markets) that operate sequentially: first agents trade bilaterally in a decentralized market (DM), and afterwards they meet in a centralized market (CM). Agents consume and supply labor in both sub-periods. There is a unified fiscal-monetary authority which supplies the only asset available in this economy for the households: fiat money, which is perfectly divisible and storable in any non-negative quantity. Money, which has no intrinsic value, is essential in this model since meetings in the DM are anonymous, there is no record keeping and agent's histories are private information so exchange must be quid-pro-quo. Individual nominal money will be normalized with respect to the beginning of each period money supply, M. Then, if $\hat{m}$ represents individual nominal money holdings, $m = \hat{m}/M$ denotes the individual relative money holdings.

In the DM, agents are randomly matched in anonymous bilateral meetings where α is the probability of a match. Each individual is specialized in the production of a non-storable differentiated good, q. Each agent can transform labor one for one into one of these special goods that she herself does not consume. For any two individuals i and j matched at random, there are three different types of meetings. Either agent i consumes what agent j produces (a single-coincidence), agent j consumes what agent i produces (a single-coincidence) or neither

wants what the other produces (no coincidence). In a single-coincidence meeting, if i wants the special good that j produces, we call i the buyer and j the seller. Let the probability of being matched as a buyer be σ. Upon matching, agents must determine the terms of trade between them, this is, the quantities of the special good being traded and its price in units of money. Here we assume that these terms of trade are determined by a take-it-or-leave-it offer made from the buyer to the seller.

In the second sub-period agents trade in a centralized Walrasian market. With centralized trade, specialization does not lead to a double-coincidence problem, and so it is irrelevant whether the good c comes in many varieties or one; hence we assume that all agents produce and consume a general good in the CM. Agents are heterogeneous in the production technology that transforms labor into the general good. In this sense, each household periodically receives an idiosyncratic productivity shock that influences its labor income. This shocks follows a Markov process. The introduction of this feature into the standard search theoretic model is key because of two reasons. First, it adds persistence in households' earnings. This is most relevant in order to refrain households from moving easily throughout the money holdings distribution as in the standard search model in the literature, where the position of the agents in such distribution is completely given for the sequence of idiosyncratic liquidity shocks that they face, i.e. whether they get to be a buyer or a seller in the decentralized market. By incorporating this productivity shock with persistence, households do not just transit states of being "poor" and "rich" in terms of money holdings so frequently, and this states do not only depend on the liquidity shock. Secondly, it incorporates precautionary savings motives to the model. In this sense, being aware of the possibility of a change in their productivity, agents stock up savings in case of a bad shock. This is to capture the role of money holdings as a savings vehicle, given the fact that money is the only storable asset in this economy.

In this sub-period, agents also can adjust their money portfolio and receive transfers from the fiscal-monetary authority, if there are any. In addition, contrary to what is assumed in Lagos and Wright (2005), the agent's utility in this sub-period is not restricted to be linear in the labor argument, and as a result, there won't necessarily be a degenerate distribution of assets. Notice that special goods cannot be traded at the CM nor general goods during the DM because they are produced in only one sub-period and are not storable.

Decentralized Market

In the decentralized market, agents are randomly matched in bilateral meetings where they trade their specialized good for money. This differentiated good is nonstorable, can only be produced in the DM and its production technology is one-to-one with labor. The sub-period utility function is:

$$\mathcal{U}(q,h) = u(q) - v(h) \tag{1}$$

where q and h represent consumption and labor in the DM, respectively. Moreover, u and v are the utility and cost functions, respectively, which satisfies the following properties: $u(0) = v(0) = 0$, $u' > 0$, $v' < 0$, $u'' < 0$, $v'' \geq 0$ and are twice continuous and differentiable.

Let $F(m,s)$ and $G(m,s)$ be the probability measures that summarize the distribution over money and productivities for the DM and CM, respectively. These distributions are defined on the Borel algebra, $\mathcal{Z}$, generated by the open subsets of the space $Z = \mathcal{M} \times \mathcal{S}$. Given that the probability of being matched with someone with a particular portfolio (m,s) depends on how many individuals on the entire economy actually holds (m,s), the probability distribution is the aggregate state variable. Denote the law of motion for these distributions as $G = \Gamma_G(F,\theta)$ and $F' = \Gamma_F(G,\theta)$, where θ represents the vector of fiscal-monetary policy. Therefore, in equilibrium we must have:

$$\int_{\mathcal{M} \times \mathcal{S}} m \, dF([dm \times ds]) = M \tag{2}$$

Denote by m the amount of money with which an individual enters the CM, so we can define $x = m + \tau$ as cash on hand at the beginning of the CM, where τ corresponds to nominal transfers given by the fiscal-monetary authority. Note that cash on hand, x, and productivity states,s, are the relevant individual state variables in the CM. Besides, the distribution over money holdings and producivities, $G(m,s)$ represents the relevant aggregate state variable. In this context, define $V(m,s;F,\theta)$ and $W(x,s;G,\theta)$ as the value functions at the beginning of the DM and CM respectively.

Assume that when two individuals are matched in the DM, they decide over the terms of

trade with a take-it-or-leave-it offer made by the buyer to the seller, where the buyer offers to buy q units of the special good in exchange for d units of money. In a bilateral meeting where the buyer's state is (m_b, s_b) and that of the seller is (m_s, s_s), terms of trade are determined as the solution to the following generalized Nash bargaining problem:

$$\max_{q,d}[u(q)+W(m_b-d+\tau, s_b; G)-W(m_b+\tau, s_b)]^\theta[-v(q)+W(m_s+d+\tau, s_s; G)-W(m_s+\tau, s_s)]^{1-\theta}$$

subject to

$$-v(q) + W(m_s + d + \tau, s_s) \geq W(m_s, s_s; G)$$

$$u(q) + W(m_b - d + \tau, s_b) \geq W(m_b, s_b; G)$$

$$q \geq 0; \quad 0 \leq d \leq m_b$$

$$G(m, s) = \Gamma_G(F(m, s), \theta)$$

The first term in the maximization problem represents the surplus of the buyer. This is composed by the utility a buyer gets from consuming the specialized good in the DM, $u(q)$, and the corresponding continuation value, which is represented by the value in the CM of the cash on hand remaining after the DM transaction, i.e. $W(m_b - d, s_b)$. On the other hand, the threat point is given by the utility that corresponds to leave the bilateral meeting without doing any trading and just move onto the CM with the initial money holdings, $W(m_b, s_b)$. The second term represents the side of the seller. Given a successfully trade, the seller must produce the quantity accorded, which costs disutility $v(q)$. However, the seller moves into the CM with a higher stock of money, which is represented by $W(m_s + d, s_s)$. Finally, as in the case of the buyer, he threat point is given by the utility that corresponds to leave the bilateral meeting without doing any trading and just move onto the CM with the initial

money holdings. Then, the first two constraints are the participation constraints of the buyer and the seller, respectively. The third and fourth constraints indicate that quantities have to be non-negative and transaction of money holdings are bounded below by the non-negativity constraint as well, but are also bounded above by the total quantity of money holdings that the buyer has. In this problem, θ represents the bargaining power of the buyer. If it is assumed that the buyer makes a take-it-or-leave-it offer to the seller, the above maximization problem is reduced to:

$$\max_{q,d} u(q) + W(m_b + \tau - d, s_b; G)$$

subject to

$$-v(q) + W(m_s + \tau + d, s_s; G) = W(m_s + \tau, s_s; G)$$

$$q \geq 0$$

$$0 \leq d \leq m_b$$

$$G(m, s) = \Gamma_G(F(m, s), \theta)$$

Note that, in this case, the buyer takes all the surplus of the meeting and leave the seller indifferent between trading or not, which is embedded on the participation constraint with equality. More importantly, the continuation value of both buyer and seller takes into account the portfolio that they will have entering the CM. For each type of bilateral meeting in the DM, denote the solution to the terms of trade problem as $q(m_b, s_b, m_s, s_s; F, \theta)$ and $d(m_b, s_b, m_s, s_s; F, \theta)$.

Given the current environment, the expected lifetime utility at the beginning of the DM of an agent with individual states (m, s), this is, before any bilateral meeting takes place, can be defined as follows:

$$V(m, s; F, \theta) = \alpha\sigma \int_{\mathcal{M}\times\mathcal{S}} \{u\left(q\left(z, z_s\right)\right) + W\left(m - d\left(z, z_s\right) + \tau; G, \theta\right)\} F\left(d\left[m_s \times s_s\right]\right)$$

$$+ \alpha\sigma \int_{\mathcal{M}\times\mathcal{S}} \{-v\left(q\left(z_b, z\right)\right) + W\left(m + d\left(z_b, z\right) + \tau; G, \theta\right)\} F\left(d\left[m_b \times s_b\right]\right)$$

$$+ (1 - \alpha\sigma)W(m + \tau; G, \theta)$$

where $z_b = (m_b, s_b)$ and $z_s = (m_s, s_s)$.

This expression is composed by three terms. First, there is the expected value of being matched as a buyer, taking into account that there are different sellers they can meet with, according to their individual states. This is, it takes the probability of being a buyer in the DM bilateral meeting, $\alpha\sigma$, times the expected value, given the agent's individual states, of the result of the trading. This makes explicit the use of the distributions across money holdings and productivity states as an aggregate state variable. The second term is the expected value of being matched as a seller. Finally, the last term is the value of not being matched or being matched in a no-coincidence meeting.

Centralized Market

In the centralized market, individuals decide on how much to consume and produce of the general good, as well as the money holdings they want to carry to the next period. In the CM, Agents are heterogeneous in the production technology that transforms labor into the general good. In this sense, each household periodically receives an idiosyncratic productivity shock that influences its labor income. This shocks follows a Markov process. The sub-period utility function is:

$$U(c, h) \tag{3}$$

where U is twice continuously differentiable and strictly concave in both consumption and leisure. Notice that here U is not restricted to be linear in any argument.

Define the lifetime utility of an agent that enters the CM with cash on hand x and productivity state, s, as:

$$W(x, s; G) = \max_{c,h,m'}\{U(c, h) + \beta\mathbf{E}_s V(m', s'; F')$$

subject to

$$c + \phi_m[m'(1 + \mu)] = y(s)h + \phi_m x$$

$$F'(m', s') = \Gamma_F(G(m, s), \theta)$$

where ϕ_m is the price of money in terms of the general good, i.e. the inverse of the price of the general good and μ is the rate of money growth in the economy.

In this sub-period, given household's individual states, (x, s), they optimally choose how much to consume of the general good, c, how much labor supply, h, and how much money holdings to carry to the next period, m'. The expectation in the value function is taken across all the possible productivity states next period. The budget constraint of the households simply states that the real value of their expenses -consumption and future money holdings - has to be equal to the real value of their income plus cash-on-hand.

Fiscal-Monetary Authority

The fiscal-monetary authority is in charge of the supply of money, M. Assume money grows at the constant rate μ so that $M' = (1 + \mu)M$. Let $\theta \equiv (M, \mu)$ denote the state vector that characterizes monetary policy. The fiscal-monetary authority must maintain a balanced budget at all times. Therefore, it finances its unproductive expenditures services and transfers to households with seigniorage. Thus, the fiscal-monetary budget constraint is:

$$G_t + T_t = \phi_m(M_t - M_{t-1}) \tag{4}$$

Furthermore, transfers , T, (or taxes if negative) and government expenditures, G, are expressed in real terms, so that $T = \phi_m\tau$ and $G = \phi_m g$. So the government balanced budget constraint can be expressed in nominal term as follows:

$$g_t + \tau_t = \mu$$

The government has several options for balancing its budget through transfers and expenditures. Therefore, I formulate the model in a comprehensive manner. In the quantitative analysis, I explore two methods of closing the model: (i) a scenario involving government expenditure, where $g > 0$ and $\tau = 0$, and (ii) a scenario with lump-sum transfers, where $g = 0$ and $\tau > 0$. Despite the second case with positive lump-sum transfers being the most common among the literature, I incorporate the case with positive unproductive government expending to isolate from the case of transfers' distributional effects. In addition, $\pi = \frac{\phi_{m,t-1}}{\phi_{m,t}}$ denotes inflation.

Equilibrium

A monetary equilibrium is a set of functions for value $\{V(m,s;F,\theta), W(x,s;G,\theta)\}$, allocations $\{c(x,s;G,\theta), h(x,s;G,\theta), m'(x,s;G,\theta), q(z_b,z_s)\}$, prices $\{d(z_b,z_s), \phi_m(G)\}$ and distributions $\{F(m,s), G(m,s)\}$ such that, given the policy θ :

1. Values $V(m,s;F,\theta)$ and $W(x,s;G,\theta)$ and decision rules $c(x,s;G,\theta), h(x,s;G,\theta), m'(x,s;G,\theta)$ satisfy the definitions above, for any given $\{q,d,\phi_m\}$ and $\{F(m,s), G(m,s)\}$.

2. Terms of trade $\{q(z_b,z_s), d(z_b,z_s)\}$ in the decentralized market solve the Nash bargaining problem stated, given $V(m,s;F,\theta)$ and $W(x,s;G,\theta)$.

3. There is a monetary equilibrium: $\phi_m > 0$.

4. The government has a balanced budget:

$$G_t + T_t = \phi_m(M_t - M_{t-1})$$

5. The money market clears:

20

$$\int_{\mathcal{M}\times\mathcal{S}} m dF([dm \times ds]) = M$$

6. The law of motions for $F(m,s)$ and $G(m,s)$ are given by $\Gamma_F(\cdot)$ and $\Gamma_G(\cdot)$, respectively. These maps are consistent with the initial conditions and the evolution of money holdings implied by DM and CM trade, and productivities by its Markov process. This is,

$$G(m,s) = \Gamma_G(F,\theta) = \alpha\sigma \int_{\{z_s \in Z\}} \int_{\{(m-d(z,z_s)+\tau,s)\in Z\}} F([dm \times ds])F([dm_s \times ds_s])$$
$$+ \alpha\sigma \int_{\{z_b \in Z\}} \int_{\{(m+d(z_b,z)+\tau,s)\in Z\}} F([dm \times ds])F([dm_b \times ds_b])$$
$$+ (1-\alpha\sigma) \int_{\{(m+\tau,s)\in Z\}} F([dm \times ds])$$

and

$$F'(m,s) = \Gamma_F(G,\theta) = \Pi(s,s') \int_{\{(m'(x,s))\in Z\}} G([dm \times ds])$$

where, as before, $z_i = (m_i, a_i)$ for $i = b, s$ and $Z = \mathcal{M}\times\mathcal{S}$, and $\Pi(s,s')$ is the transition matrix of productivity states.

4 Calibration

The aim of the quantitative exercise is to assess the welfare costs of inflation across the money holdings distribution for those households who do not hold any interest-bearing financial asset. As it was shown in the empirical part, the share of such households in the U.S. economy is 64.39 % according to the SCF for 2019. The quantitative results of the model regarding the costs of inflation depend on how closely the model resembles the money holdings distribution among these households. Therefore, moments of this empirical distribution would be the natural target. To measure money in the data, I assign M to be the sum of checking deposits, saving deposits, and prepaid cards. All these elements have been shown to have close to zero nominal interest rate in the data.

The model is yearly, mainly to facilitate comparison with the existing literature. For all possible parameters that have a standard value in the literature or can be mapped directly

to the data, I do so. Remaining parameters are chosen to match informative steady state moments.

In terms of functional forms, I follow Lagos and Wright (2005). Therefore, the following utility and cost functions, respectively, in the DM are as follows:

$$u(q) = \frac{1}{1-\eta}[(q+b)^{1-\eta} - b^{1-\eta}]$$

$$v(h) = Bh^{\nu}$$

where $\eta > 0$, $B > 0$, $\nu > 0$ and $b \approx 0$.

As for the CM, I deviate from the quasi-linearity assumption as already stated. In particular, the utility function in the CM is concave in consumption, has a constant Frisch elasticity of labor supply, and is given by

$$U(c,h) = \mathcal{C}\log(c) + \kappa\frac{(h)^{1+\gamma}}{(1+\gamma)}$$

where γ is the inverse of the Frisch elasticity of labor supply and $\kappa > 0$ is a scale parameter.

The parameters determining the scale of utility of consumption in the DM, $\mathcal{C}$, the scale of the disutility of labor in the CM, κ, the Frisch elasticity of labor supply, γ, and curvature of the utility function in the DM, η, are jointly calibrated to match (i) each decile in the money holdings distribution, (ii) the average markup, (iii) the velocity of money and (iv) the semi-elasticity of money demand with respect to the nominal interest rate. In what follows, I discuss the calibration of the model with no transfers, $\tau = 0$, and positive unproductive government expenditure, $g > 0$, but the moments generated by the same parameter values in the model with positive transfers and no government expenditures are almost identical. Table 1 shows the calibrated parameters.

The money growth rate is taken to be consistent with an inflation rate in steady state equal to average of sample data. The Frisch elasticity of labor supply is set equal to 0.5, which means that $\gamma = 2$. This elasticity is in the range of what Chetty, Guren, Manoli, and Weber (2021) document. For the income process, I use the persistence estimated by Floden

Table 1: Parameter Values

Parameter	Description	Value	
Internally calibrated parameters			
γ	Inverse of Frisch elasticity	2	
κ	Scale of disutility of labor	5	
$\mathcal{C}$	Scale of utility of consumption	4	
η	Curvature of utility of consumption	0.45	
β	Discount factor	0.97	
Externally calibrated parameters			
$\bar{\mu}$	Growth rate of money	0.03	Average inflation
α	Probability of meeting	1	Lagos and Wright (2005)
σ	Probability of being a buyer	0.5	Lagos and Wright (2005)
B	Scale of cost of working	1	Lagos and Wright (2005)
b	Scale parameter in $u(q)$	0.0001	Lagos and Wright (2005)
ρ_s	Persistence of log income	0.91	Floden and Lindé (2001)
σ_s	Std. of log income innovations	0.7	Floden and Lindé (2001)

Note: Internally calibrated parameters are simultaneously calibrated to match the moments reported in Table 2.

and Lindé (2001) for yearly values. For the standard deviation of innovations, σ_s, the value targets the cross-sectional standard deviation of pre-tax log income of 0.7.

The model is able to replicate a velocity of money close to the one observed in the data for the sample period: 1.827 in the model versus 1.695 in the data. Moreover, the semi-elasticity of money demand with respect to the nominal interest rate is -0.069, close in range to the one documented in Lucas (2000), Aruoba et al. (2011), and Berentsen et al. (2011). Table 2 reports the aggregate targeted moments in the data and their model counterparts.

Table 2: Aggregate Moments Used in Calibration

|	**Data**	**Model**
Velocity of Money	1.695	1.827
Semi-elasticity of money demand	[-0.05, -0.1]	-0.069

Note: Velocity of money in the data is the average for the period 1999Q1–2022Q4. Money is equal to M2 taken from FRED, Federal Reserve Bank of St. Louis, computed as the sum of currency, demand deposits, saving deposits, and time deposits. The semi-elasticity of money demand is taken from Aruoba et al. (2011).

As it was previously stated, the most important object to match to perform the quantitative

exercise is the money holdings distribution. Its relevance is given by the fact that not only does it affect the computation of the aggregate welfare cost of inflation, but also it is the key element to assess any distributive effect. For this purpose, using the SFC 2019, I consider only the households who do not hold any interest-bearing asset, which represents 64.39% of U.S. households, and I compute the distribution of money holdings across them. In this case, money holdings are composed of the sum of checking accounts, saving accounts and pre-paid cards, which have been shown to bear near zero nominal interest rates in Section 2. Finally, to make the right comparison between the data and the model, I divide money holdings by the corresponding weighted average in the data.

Figure 6: Money Holdings Distribution: Model and Data
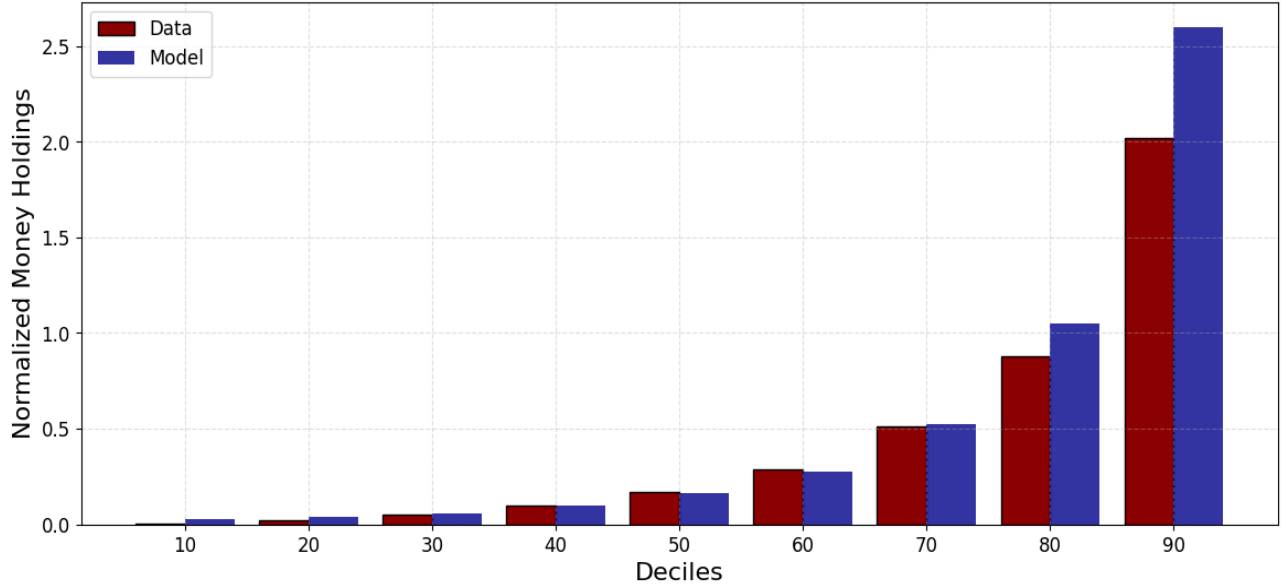
Note: Money holdings in the data are defined as the sum of checking accounts, saving accounts and pre-paid cards. These values are normalized by the aggregate weighted average, as in the model.

In this regard, the model mimics quite well the money holding distribution implied by the data. In particular, Table 3 shows the corresponding quintiles and the median of both distributions (data and model):

Table 3: Moments of Money Holdings Distribution

Money Holdings Distribution					
Percentile	20	40	50	60	80
Data	0.02	0.09	0.16	0.29	0.88
Model	0.04	0.09	0.16	0.28	1.05

5 Stationary Equilibrium

I begin by characterizing the steady-state equilibrium of the benchmark economy as this will help to better understand its properties. For this section, I consider the case of no transfers, $\tau = 0$, and positive unproductive government expenditure, $g > 0$. The alternative case with $\tau > 0$ and $g = 0$ yields very similar results as far as terms of trade, distribution of money, and decision rules are concerned.[3]

Figure 7 presents the terms of trade concerning both quantities and monetary payments involved in transactions between buyers and sellers with median productivity states. The graph illustrates how the terms of trade vary based on the buyer's money holdings, m_b, and the seller's money holdings, m_s. In any potential transaction, the quantities exchanged in the decentralized market (DM) increase with the buyer's money holdings and decrease with the seller's money holdings. To provide intuition, when the seller is wealthier, their opportunity cost of working becomes higher. Consequently, the seller demands more money in exchange for a lower level of output to maintain indifference towards trading. Conversely, the buyer's willingness to pay increases with her money holdings, ensuring higher consumption. In other words, the opportunity cost of spending an additional dollar in DM trading decreases with the seller's money holdings.

Delving into the buyer's terms of trade, Figure 8 reports the quantities of goods traded, as an average across all possible sellers, and the fraction of money holdings spent by the buyer (propensity to spend), both as a function of the buyer's real money holdings. Additionally, this figure shows these patterns for three different productivity levels: low, middle and high.

As it was pointed out before, the quantities traded are increasing and concave in the

---

[3]For all these numerical exercises, we use the solution method outlined in Appendix B

Figure 7: Terms of Trade: Quantities and Monetary Payments
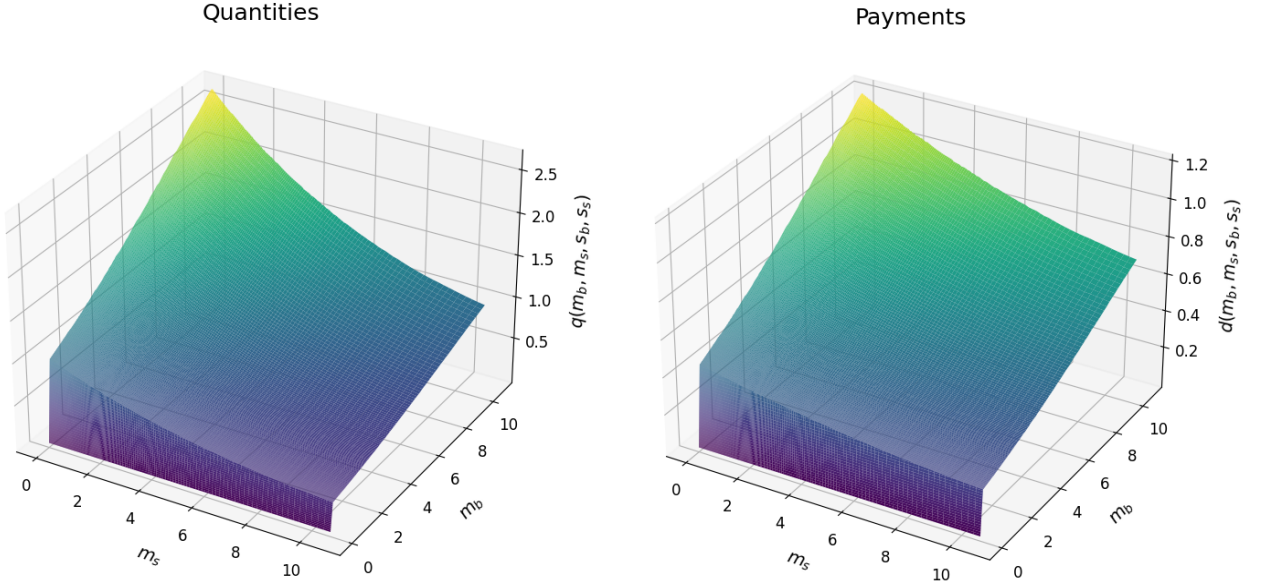
Note: For expositional purposes, the figures only present terms of trade as a function of (m_b, m_s) for meetings between both buyers and sellers with median productivity states, (s_b, s_s)

buyer's real money holdings. This is due to the fact that as the buyer gets richer, the more she wants to spend on current consumption. However, as the level of consumption increases, the seller's marginal value of money is driven down, and then the quantity the buyer is able to purchase will increase at a slower rate. As the seller becomes richer, the marginal amount of money holdings spent by the buyer becomes less valuable. As the buyer needs to pay more to buy the same quantity, production drops. Note that, as is standard in search-theoretic models of money, trading frictions in decentralized trading generate inefficient allocations. The dashed black line in the first panel in Figure 8 shows the efficient level for quantities exchanged, q^*, that solves $u'(q^*) = v'(q^*)$. Nevertheless, differences not only in money holdings but also in productivity levels give rise to inefficient outcomes. In this sense, buyers with less real balances typically consume less than the efficient quantity while buyers with more real balances typically consume more than the efficient quantity. In this regard, buyers with higher productivity levels consume more. This is because the continuation value out of a trade for a high-productivity agent is larger and, as such, the marginal value of an additional unit of money spent decreases which, in turn, implies higher consumption.

In line with this result, the right panel on Figure 8 depicts the buyer's propensity to spend,

defined as the total money payment made relative to the buyer's money holdings. It can be seen that buyers with low levels of real balances are liquidity constraint and thus they spend all their balances in each trade. Nonetheless, the fraction of money a buyer will spend at a meeting is decreasing in their money holdings. This is more apparent for low-productivity agents since, in the rare event that such an agent can stock up a large amount of real balances, their continuation value for spending it all is so low that they wish to accumulate most of it in order to maintain a higher consumption in the CM and enjoy better terms of trade in the following round of bilateral meetings. Such motive is much less strong in high-productivity agents because its easier for them to stock up larger amounts of real money holdings.

Figure 8: Terms of Trade - Buyer
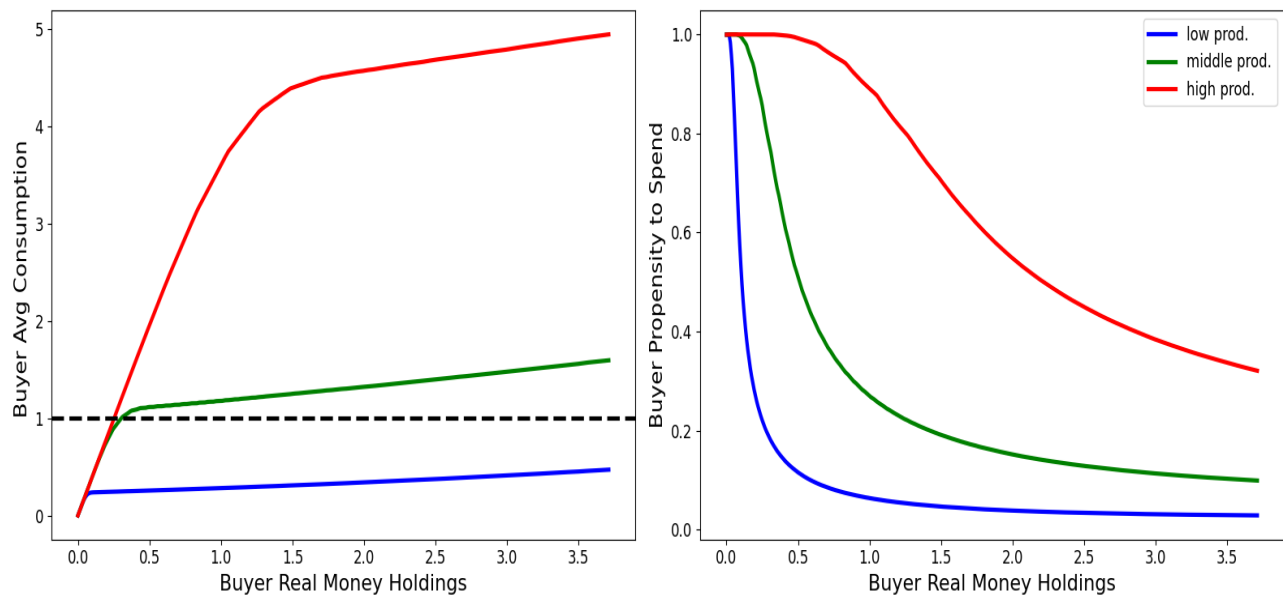
Note: Buyer's terms of trade. Quantities consumed are averaged across all possible sellers. Propensity to consume refers to the total money payment made relative to the buyer's money holdings. The dashed black line correspond to the efficient quantity exchanged.

As far as the seller is concerned, Figure 9 illustrates the terms of trade for the sellers in more depth. The panel on the left depicts the quantities of the good produced by the seller according to their real balances, as an average across all possible buyers in the economy. Again, note that poor sellers typically produce more than the efficient quantity while richer sellers will typically produce less than the efficient quantity. This means that a higher dispersion in money holdings generates a more inefficient equilibrium. Additionally, it is

worth noticing that low productivity agents produce more that high productivity ones. This is due to the fact that the marginal value of an extra unit of real balances for the first type of agents is generally greater than the cost of production, whereas, for the latter type of agents, the inverse is true. On the other hand, the right panel in Figure 9 plots the per-unit price at a meeting. As can be seen, this price is increasing in the real balances of the seller. The intuition is straightforward, the richer the seller, the lower their marginal value of money and thus, for the same amount of goods, the buyer will be willing to offer a a higher price and the seller will demand more money, and thus a higher price. Moreover, high-productivity agents are paid a higher price for their production given the fact that their marginal cost of working relative to the marginal benefit of an extra unit of real balances is higher than that of low productivity agents.

Interestingly, the diverse terms of trade arising from the bilateral meetings in the decentralized market result in buyers and sellers bringing different money holdings into the next centralized market. In contrast to models with quasi-linear preferences, the decisions of agents in the CM are notably influenced by wealth effects. Consequently, not all agents opt to accumulate identical amounts of money, leading to persistent heterogeneity generated by decentralized trading across periods. [4]

Finally, the diversity in money holdings among agents upon entering the decentralized market results in a fully-fledged distribution of prices, calculated as the ratio of monetary payments (d) to quantities exchanged (q). Figure 10 illustrates this price distribution. The dispersion in money holdings is sufficient to generate a standard deviation of 0.37 in the observed prices.

6 Welfare Costs of Inflation

The spirit of this section is to study what are the implications of having an heterogeneous distribution of money holdings in the welfare costs of inflation. This exercise is of interest since we can address whether the *aggregate* costs of inflation are the same as in the traditional models with degenerate money holdings distributions or the presence of heterogeneity changes

---

[4]Refer to Appendix for details in the policy functions in the CM.

Figure 9: Terms of Trade - Seller
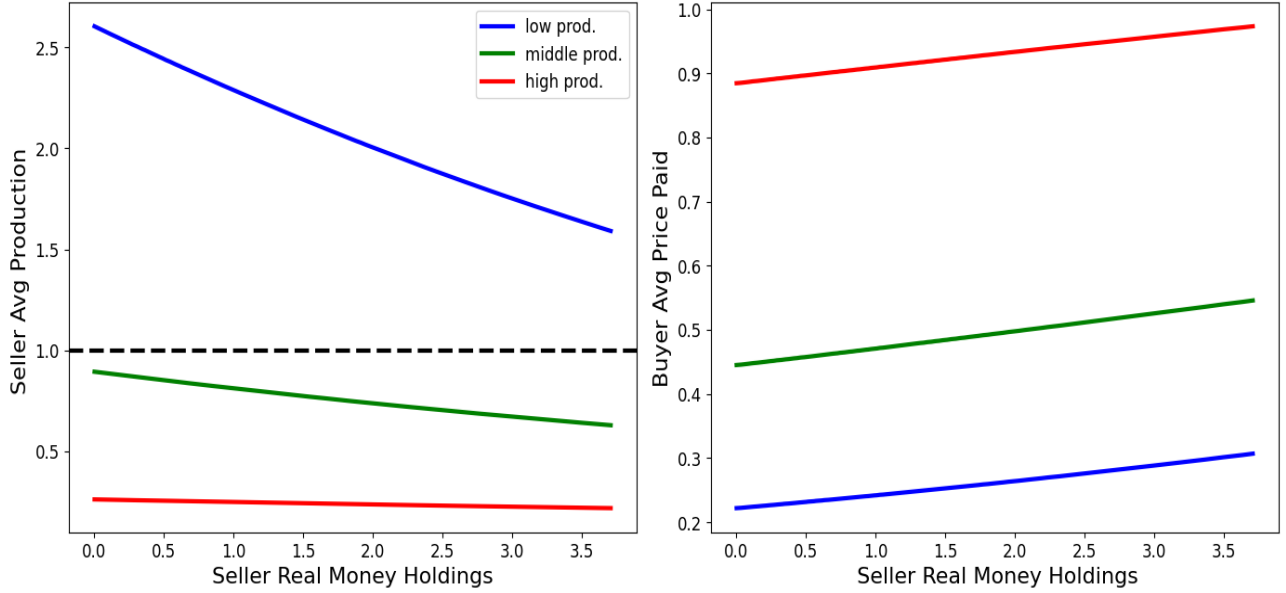
Note: Seller's terms of trade. Quantities produced are averaged across all possible buyers. Price paid refers to the total money payment made relative to the total quantity exchanged. The dashed black line correspond to the efficient quantity exchanged.

their result. Furthermore, we can decompose the aggregate welfare changes for different agents across the money holdings distribution. In this sense, the aim of this quantitative exercise is to assess the welfare costs of inflation across the money holdings distribution for those households who do not hold any interest-bearing financial asset.

To begin with, I start by computing the stationary equilibrium of the monetary economy under a 0% inflation regime . Then, I use the model to obtain the stationary equilibrium of the economy under a 10% inflation regime and I ask what are the welfare cost of living in the latter economy in comparison to the former. The way I measure the welfare costs is by asking how much total consumption agents would be willing to give up in order to move from a stationary equilibrium with 10% inflation to one with zero inflation.

For any given money holdings m and productivity state, s, I compute the value function at the 10% inflation steady-state as follows:

Figure 10: Stationary Distribution of Prices
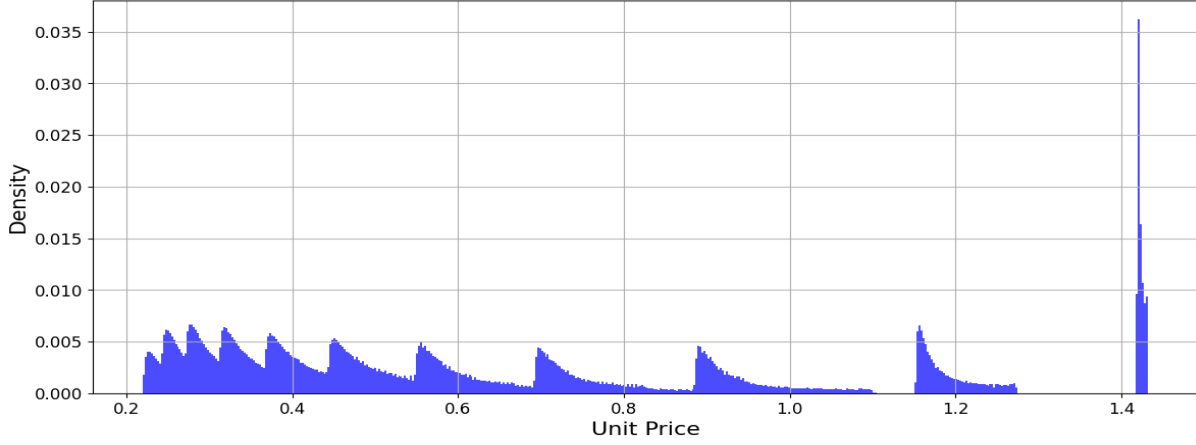
Note: Unit prices are defined as $p(m_b, s_b, m_s, s_s) = d(m_b, s_b, m_s, s_s)/q(m_b, s_b, m_s, s_s)$. The distribution of prices accounts for all possible meetings between buyers and sellers given the stationary distribution of agents over money holdings and productivity levels.

$$
V_{10}(m, s) = \alpha\sigma \int \Big[u(q(z, \hat{z})) + W_{10}(m - d(z, \hat{z})) + \beta\mathbf{E}_s V_{10}(m'(m - d(z, \hat{z})), s') \Big] dF(m \times s)
$$
$$
+ \alpha\sigma \int \Big[-v(q(\hat{z}, z)) + W_{10}(m + d(\hat{z}, z) + \beta\mathbf{E}_s V_{10}(m'(m + d(\hat{z}, z)), s') \Big] dF(m \times s)
$$
$$
+ (1 - \alpha\sigma)\Big[W_{10}(m) + \beta\mathbf{E}_s V_{10}(m'(m), s') \Big]
$$

where

$$
W_{10}(x, s) = \mathcal{C} \log(c(x, s)) - \kappa \frac{h(x, s)^{1+\gamma}}{1 + \gamma}
$$

and $x = m - d(z, \hat{z}) + tr$ when buyer and $x = m + d(\hat{z}, z) + tr$ when seller. As stated before, $z = (m, s)$.

In the same way, the value function at the 0% inflation steady-state is computed for all values of money holdings and productivities, obtaining :

$$V_0^\Delta(m,s) = \alpha\sigma \int \left[u(q(z,\hat{z})) + W_0(m - d(z,\hat{z})) + \beta\mathbf{E}_s V_0^\Delta(m'(m - d(z,\hat{z})), s') \right] dF(m \times s)$$
$$+ \alpha\sigma \int \left[-v(q(\hat{z},z)) + W_0(m + d(\hat{z},z) + \beta\mathbf{E}_s V_0^\Delta(m'(m + d(\hat{z},z)), s') \right] dF(m \times s)$$
$$+ (1 - \alpha\sigma) \left[W_0(m) + \beta\mathbf{E}_s V_0^\Delta(m'(m), s') \right]$$

where, as before,

$$W_0 = \mathcal{C}\log(\Delta c) - \kappa\frac{h^{1+\gamma}}{1+\gamma}$$

Notice that if we combine the two previous expressions we get:

$$V_0^\Delta(m,s) = \alpha\sigma \int \left[u(\Delta q(z,\hat{z})) + \mathcal{C}\log(\Delta c) - \kappa\frac{h^{1+\gamma}}{1+\gamma} + \beta\mathbf{E}_s V_0^\Delta(m'(m - d(z,\hat{z})), s') \right] dF(m \times s)$$
$$+ \alpha\sigma \int \left[-v(q(\hat{z},z)) + \mathcal{C}\log(\Delta c) - \kappa\frac{h^{1+\gamma}}{1+\gamma} + \beta\mathbf{E}_s V_0^\Delta(m'(m + d(\hat{z},z)), s') \right] dF(m \times s)$$
$$+ (1 - \alpha\sigma) \left[\mathcal{C}\log(\Delta c) - \kappa\frac{h^{1+\gamma}}{1+\gamma} + \beta\mathbf{E}_s V_0^\Delta(m'(m), s') \right]$$

Thus, after having computed both $V_{10}(m)$ and $V_0^\Delta(m)$, for each level of money holdings, m, and productivity states, s, all that is left to do is to solve for the one Δ such that [5]

$$V_{10}(m,s) = V_0^\Delta(m,s) \tag{5}$$

This computation is done for every pair (m,s) in the grid. So that the result is the "individual cost" for each possible pair of money holdings and productivity states, m, s, measured as $(1 - \Delta)$. Finally, the last step is to weight them for the corresponding density $F(m,s)$ in the zero-percent inflation steady state.

Before presenting the analysis on inflation costs, I proceed to illustrate the most significant differences between the steady-state economy with 0% inflation and its counterpart with 10%, both for the case with lump-sum transfers and unproductive government spending, since there are some crucial differences between these two. For this purpose, Table 4 summarizes

[5]For a more detailed explanation about this computation look at the appendix

the most important statistics for the two economies:

Table 4: Aggregate Welfare Costs of Inflation

	$\pi = 0$	$\pi_{tr} = 0.1$	$\pi_g = 0.1$
Real money balances	1.2787	0.2933	0.2893
Std. dev. money (DM)	2.5240	1.7630	1.7902
Avg. price (DM)	0.2416	1.0281	1.0252
$\%(d = m_b)$	0.7881	0.9117	0.9134
$\%(m = 0)$	0.1147	0.0145	0.0296
q (DM)	0.7441	0.8586	0.8675
Std. dev. q	1.3459	1.3068	1.3153
Welfare cost	0	1.79%	3.19%

There are some interesting results worth noticing. First, the value of money decreases sharply with inflation, as it is expected, thus real money balances go down. Along this line, the price of the specialized good rises in the decentralized market. Second, the reduced dispersion in money holdings in the decentralized market is due to the fact that higher inflation makes it less convenient for households to carry large amounts of money into the round of bilateral meetings, as inflation acts as a tax on those holdings. In the case of the economy with lump-sum transfers, this dispersion is even less, implying distributive effects from the rich to the poor. This, in turn, reduces the dispersion in the quantities traded, since buyers and sellers are less unequal in terms of money holdings. Third, output in the decentralized market increases by 16.23% when inflation increases from 0 to 10%. This result can be explain by two different forces. On the one hand, the number of liquidity constrained agents, i.e. agents that enter the round of decentralized trade with zero money holdings, is reduced significantly in a high-inflation steady-state. Since money is essential for trading in the DM, liquidity constrained agents don't consume at all in this sub-period. In this regard, liquidity constrained agents represent almost 12% of the economy in a 0% steady-state, whereas, this share reduces to 1.5% and 3% in an economy with 10% steady-state inflation, for the case of lump-sum transfers and government spending, respectively. The fact that the amount of liquidity constrained agents is lowest in the case of lump-sum transfers also denotes the redistributive effects of such policy. On the other hand, the proportion of bilateral meetings in which the buyers spend all their money in the trading opportunity out of the total trade matches, increases from 74.41% in a 0% inflation steady-state to around 91%

in its 10% inflation steady-state counterpart. Interestingly, for low levels of inflation, an increase in the rate of monetary expansion leads to a decrease in the dispersion of money holdings and prices, and an increase in the average quantity traded in the decentralized market. In particular, inflation as a tax on money holdings induce richer agents to reduce their average holdings of money, which causes a redistribution of liquidity from the rich to the poor. This, in turn, leads to a smaller dispersion in observed prices and in money holdings. Since there are fewer very poor and very rich agents, the average quantity traded in this sub-period increases. However, in equilibrium, a real balance effect also occurs. In this sense, the fraction of money paid per unit of the good at each bilateral meeting increases, given that agents are less willing to hold a high amount of money holdings, which leads to a decrease in the real money balances and in the quantity traded. Figure 11 exhibits this result more clearly. It can be seen that, for low levels of inflation, the average quantity traded in the decentralized market increases. Yet, when the level of inflation becomes higher, the average quantity traded in the decentralized market falls. The break-point level of inflation is around 4% for the benchmark economy.

Figure 11: Inflation - Output Relationship
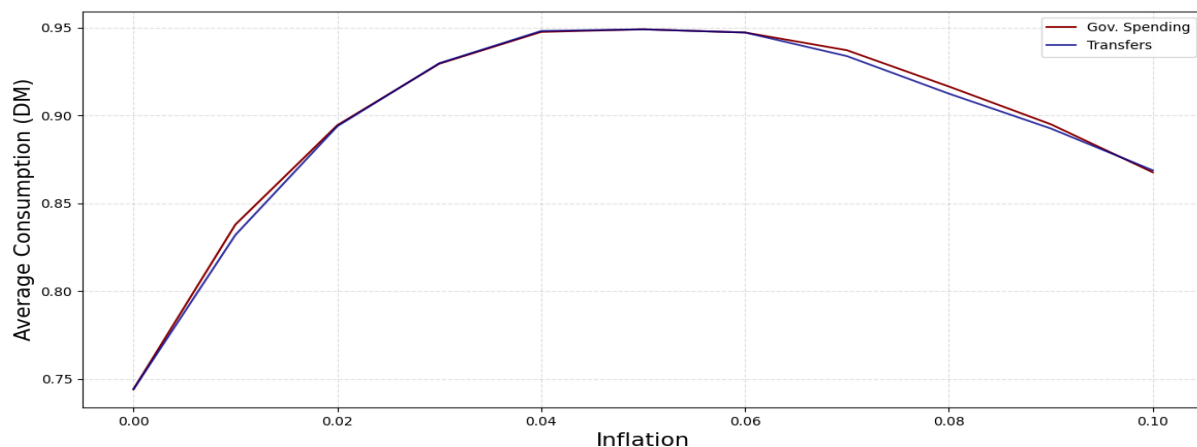
Note: The horizontal axis represents steady-states with different inflation levels, from 0% to 10%. The vertical axis represents average consumption in the decentralized market, given by the output outcome of each possible trade weighted by the corresponding density, $F(m.s)$. The blue line indicates the case with lump-sum transfers. The red line indicates the case with government expenditures.

The most important result of this exercise indicates that, on average, agents would be wiling to give up 3.19% of their total consumption in order to move away from an economy

with 10% steady-state inflation, where money is injected to finance unproductive government expenditures, into an economy with zero inflation. The corresponding cost for the case where money is injected via lump-sum transfers is 1.79% of consumption. First of all, these numbers are much higher than the one found in Lagos and Wright (2005) for the same bargaining structure, suggesting that the heterogeneity in money holdings and persistent productivity types play an important role in determining the costs of inflation. Secondly, there is a sharp difference in the welfare cost of inflation between the two different policies analyzed. When money injections are introduced to finance government expenditures, the costs of inflation in terms of consumption equivalent units are higher than in the case of lump-sum transfer. This is indicative of a strong redistributive effect of the latter policy, since these transfers work as a subsidy for the poor and a tax on the rich.

In order to assess this issue in more depth, I conduct a second exercise in which I disaggregate the total welfare costs by quintiles of the money holdings distribution. In this sense, Figure 12 shows the welfare costs of inflation in terms of consumption equivalent units by quintile of the money holdings distribution, for the two alternative policies. First, it is interesting to notice that inflation costs are increasing in money holdings. This is due to the fact that inflation deteriorates the two roles of money in this economy: as a transaction vehicle and as an instrument for precautionary savings. On the one hand, since inflation acts as a tax on money holdings, the opportunity cost of carrying money into the next round of decentralized market meetings increases. On the other hand, the opportunity cost on savings also goes up. This is particularly costly for more productive agents who, in the event of a bad realization of the productivity shock, rely on savings to smooth consumption. However, high inflation sharply harms the role of money as a saving instrument and distorts the consumption-savings decision of the agents. Even though more productive agents increase their consumption, on average, in steady-states with a higher inflation rate, in line with a higher velocity of money, the distortion created on their savings decision deteriorate their welfare. As far as low productivity agents are concerned, in higher inflation regimes, they find it relatively more costly to stock up money holdings for future liquidity shocks, as they face higher prices. However, given the decreased desire of richer agents to carry over large amount of money holdings, and the drop in the relative price of money to the

consumption good, steady-states with higher inflation exhibit a less dispersed distribution of money holdings, which improves term-of-trade outcomes. Additionally, there is a sharp contrast in the distribution of the welfare cost of inflation depending on how money is being injected into the economy. On the one hand, if the fiscal-monetary authority increases the aggregate money supply in order to finance its expenditures, agents would have to increase their labor supply, in view of the additional demand for the general good by the government. In particular, poor agents increase their labor supply more, leading to a higher disutility. On the other hand, in the case of lamp-sum transfers, the fiscal-monetary authority just hands over the new issuance of money to the agents. As such, this policy has a strong redistributive effect since it acts as a tax on the rich and a subsidy to the poor. This result is evident in Figure 12 where the welfare costs of inflation are negligible for the lowest quintiles of the money holdings distribution when money injections are made via lump-sum transfers, whereas, under the alternative policy, these costs are much higher. In the above-presented model, two opposing dynamics come into play. Firstly, high inflation leads to a more rapid erosion in the real value of portfolios for individuals holding larger amounts of nominal assets. Consequently, the impact of the inflation tax is not uniformly distributed across the population, especially in the presence of a non-degenerate distribution of money. Given that, in such scenarios with high inflation, individuals typically hold only the necessary amount of money for transactional purposes, the concentration in money holdings plays a crucial role in determining which individuals bear a greater burden of the inflation tax. However, this also erodes the possibility for agents to buffer against bad productivity shocks. This higher vulnerability increases welfare costs of inflation, especially for high productivity agents. On the flip side, higher inflation rates are linked to larger lump-sum transfers, when this is the mechanism that increases inflation in the economy, it can have important distributive effects. All these features of the model underscore the significance of considering distributional effects when assessing the costs associated with inflation.

Figure 12: Welfare Costs of Inflation by Quintile
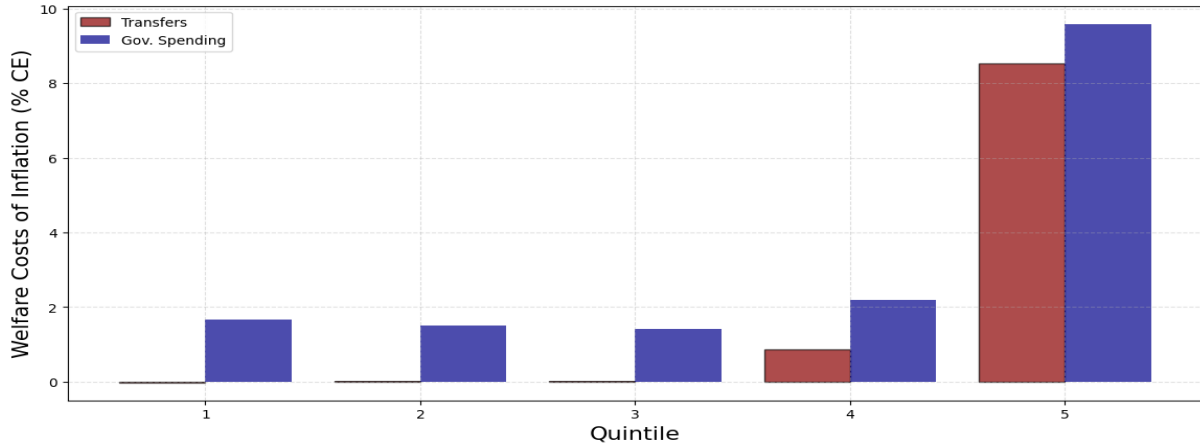
Note: Welfare costs of inflation are expressed as percentages in terms of consumption equivalent units. The blue bars represent the case with government spending. The red bars represent the case with lump-sum transfers.

7   Concluding Remarks

This paper emphasizes a stark feature of the U.S. economy: around 60% of the U.S. households hold all their liquid financial wealth in the form of checking accounts, saving accounts and pre-paid cards, and these instruments have paid a nominal interest rate on its deposits that has been close to zero for the last twenty years. Surprisingly, the share of such a particular type of households has been stable for the last twenty-five years, even when inflation rates have been higher than usual, as it is the case of the post COVID-19 crisis and the current global economic situation. This economic environment with high inflation rates poses a threat to the large proportion of the U.S. households who do not bear any interest-bearing liquid financial asset. In this paper, I study the long-run welfare costs of inflation for this particular type of households. To do this, I develop a search-theoretic model of money where agents are ex-ante heterogeneous in their productivity level, and money serves two roles: as a medium of exchange and as an instrument for precautionary savings. Then, a consolidated fiscal and monetary authority can use the aggregate supply of money to manage the public provision of liquidity and the inflation rate. The model generates a nondegenerate distribution of money in steady-state that matches the one observed in the data. This last point is crucial to assess the unequal costs of inflation across the money holdings distribution. My results show that a

permanent increase in the rate of inflation ruins household's ability to buffer against liquidity and income shocks, distorting optimal savings-consumption decisions. In this sense, average welfare costs of inflation are higher than in previous studies, where the precautionary savings motive is absent. This result advocates for grater financial integration. Moreover, my results show that inflation hits differently to agents across the distribution of money holdings, and this crucially depends on how money is being injected into the economy. In this regard, if lump-sum transfers are used, the poorest and more liquidity-constrained agents lose the least -or even gain- warfare, as some resources get redistributed out of the wealthiest agents in the economy. Alternatively, if the fiscal-monetary authority uses the new issuance of money to finance its expenditures, average welfare costs are much larger for the bottom quintiles of the money holdings distribution. This indicates that the distributional effects of inflation cannot be overlooked.

References

R. Aiyagari. *Uninsured Idiosyncratic Risk and Aggregate Saving.* The Quarterly Journal of Economics, 1994. ISBN 9781417642595. URL https://www.jstor.org/stable/2118417.

B. Aruoba, J. Waller, and R. Wright. *Money and Capital.* Journal of Monetary Economics, 2011. ISBN 9781417642595. URL https://ideas.repec.org/a/eee/moneco/v58y2011i2p98-116.html.

A. Auclert, M. Rognlie, , and L. Straub. *The intertemporal keynesian cross.* Working Paper, National Bureau of Economic Research, 2018. ISBN 9781417642595. URL http://www.nber.org/papers/w25020.

A. Auclert, M. Rognlie, , and L. Straub. *Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model.* Working Paper, National Bureau of Economic Research, 2020. ISBN 9781417642595. URL http://www.nber.org/papers/w26647.

A. Auclert, B. Bardóczy, M. Rognlie, and L. Straub. *Using the sequence-space jacobian to solve and estimate heterogeneous-agent models.* Econometrica, 2021. ISBN 9781417642595. URL https://doi.org/10.3982/ECTA17434.

A. Berentsen, G. Menzio, and R. Wright. *Inflation and Unemployment in the Long Run.* The American Economic Review, 2011. ISBN 9781417642595. URL https://www.aeaweb.org/articles?id=10.1257/aer.101.1.371.

T. Bewley. *A Difficulty with the Optimum Quantity of Money.* Econometria, 1983. ISBN 9781417642595. URL https://www.jstor.org/stable/1912286.

C. Bustamante. *More Money for Some: Monetary Policy Meets a Rich and Persistent Household Wealth Distribution.* Working paper, 2018. ISBN 9781417642595. URL https://cbustamante.co/research.html.

R. Chetty, A. Guren, D. Manoli, and A. Weber. *Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins.*

American Economic Review, 2021. ISBN 9781417642595. URL https://www.aeaweb.org/articles?id=10.1257/aer.101.3.471.

J. Chiu and M. Molico. *Liquidity, redistribution, and the welfare cost of inflation.* Journal of Monetary Economics, 2010. ISBN 9781417642595. URL http://www.sciencedirect.com/science/article/pii/S0304-3932(10)00036-X.

J. Chiu and M. Molico. *Uncertainty, Inflation, and Welfare.* Journal of Money, Credit and Banking, 2011. ISBN 9781417642595. URL https://doi.org/10.1111/j.1538-4616.2011.00448.x.

J. Chiu and M. Molico. *Short-Run Dynamics in a Search-Theoretic Model of Monetary Exchange.* Review of Economic Dynamics, 2021. ISBN 9781417642595. URL https://www.sciencedirect.com/science/article/abs/pii/016518899290006Z.

F. Cirelli. *Bank-Dependent Households and the Unequal Costs of Inflation.* JMP, 2023. ISBN 9781417642595. URL https://fercirelli.github.io/personal_website/Cirelli_JMP.pdf.

I. Drechsler, A. Savov, and P. Schnabl. *The Deposit Channel of Monetary Policy.* The Quarterly Journal of Economics, 2017. ISBN 9781417642595. URL https://academic.oup.com/qje/article-abstract/132/4/1819/3857743?redirectedFrom=fulltext&login=false.

A. Erosa and G. Ventura. *On inflation as a regressive consumption tax.* Journal of Monetary Economics, 2000. ISBN 9781417642595. URL http://www.sciencedirect.com/science/article/pii/S0304-3932(02)00115-0.

M. Floden and J. Lindé. *Idiosyncratic Risk in the United States and Sweden: Is There a Role for Government Insurance?* Review of Economic Dynamics, 2001. ISBN 9781417642595. URL https://www.sciencedirect.com/science/article/abs/pii/S1094202500901212.

G. Kaplan, B. Moll, and G. Violante. *Monetary Policy According to HANK.* American Economic Review, 2018. ISBN 9781417642595. URL https://www.aeaweb.org/articles?id=10.1257/aer.20160042.

N. Kiyotaki and R. Wright. *On Money as a Medium of Exchange.* The University of Chicago Press, 1989. ISBN 9781417642595. URL https://www.jstor.org/stable/1832197?seq=1#metadata_info_tab_contents.

N. Kocherlakota. *Money: What's the Question and Why Should We Care about the Answer?* The American Economic Review, 2002. ISBN 9781417642595. URL https://www.jstor.org/stable/3083377?seq=1#metadata_info_tab_contents.

R. Lagos and R. Wright. *A Unified Framework for Monetary Theory and Policy Analysis.* Journal of Political Economy, 2005. ISBN 9781417642595. URL https://www.jstor.org/stable/10.1086/429804?seq=1#metadata_info_tab_contents.

R. Lucas. *Inflation and Welfare.* Econometrica, 2000. ISBN 9781417642595. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00109.

G. Menzio, S. Shi, and H. Sun. *A Monetary Theory with Non-Degenerate Distributions.* Journal of Economic Theory, 2013. ISBN 9781417642595. URL https://ideas.repec.org/p/red/sed010/598.html.

M. Molico. *The Distribution of Money and Prices in Search Equilibrium.* International Economic Review, 2006. ISBN 9781417642595. URL https://doi.org/10.1111/j.1468-2354.2006.00393.x.

G. Rocheteau and R. Wright. *Money in Search Equilibrium, in Competitive Equilibrium, and in Competitive Search Equilibrium.* Econometrica, 2005. ISBN 9781417642595. URL https://doi.org/10.1111/j.1468-0262.2005.00568.x.

G. Rocheteau, W. Pierre-Olivier, and W. Tsz-Nga. *A tractable model of monetary exchange with ex post heterogeneity.* Theoretical Economics, 2018. ISBN 9781417642595. URL https://doi.org/10.1111/j.1538-4616.2011.00448.x.

S. Shi. *Money and Prices: A Model of Search and Bargaining.* Journal of Economic Theory, 1995. ISBN 9781417642595. URL http://www.sciencedirect.com/science/article/pii/S0022-0531(85)71081-2.

A. Trejos and R. Wright. *Search, Bargaining, Money, and Prices.* The University of Chicago Press, 1995. ISBN 9781417642595. URL http://www.jstor.org/stable/2138721.

A. İmrohoroğlu. *The welfare cost of inflation under imperfect insurance.* Journal of Economic Dynamics and Control, 1992. ISBN 9781417642595. URL https://www.sciencedirect.com/science/article/abs/pii/016518899290006Z.

Appendix

A   Data Appendix

This section complements the evidence presented in Section 2 and show details on data computations.

A.1   Data Sources

Data on households' portfolios comes from the Survey of Consumer Finances (SCF), a U.S. households survey sponsored by the Federal Reserve Board. The survey is a repeated cross-sectional survey of U.S. families that collects information on household balance sheets, income, and demographic characteristics. Post-1983 data of the SCF is available on the website of the Board of Governors of the Federal Reserve System. In the modern version of the survey around 6500 families are interviewed every three years with particular attention to capturing top wealthy families. I keep the entire sample of households in the SCF without any demographic or income restrictions.

B   Solution Algorithm

The solution algorithm used here is based on solving iteratively for the agents' decision rules, the terms of trade in decentralized trading, the level of lump-sum transfers consistent with the government balance, and the prices that clear the markets of bonds and money.

1 Discretize the state space $\mathcal{M} \times \mathcal{S}$ and $\mathcal{X}$, so that $F(m,s), V(m,s;F), < Q, D > (m_b, m_s, s_b, s_s)$ and $W(x,s;G)$ are defined over an infinite number of points.

2 Start with some initial guess for $F(m, s), W(x, s; G)$ and $< Q, D > (m_b, m_s, s_b, s_s)$.

3 Start with some arbitrary prices ϕ_m and transfers tr or government expenditures, g, if any .

4 Solve for $V(m, s; F)$ and $W(x, s; G)$ by updating them sequentially until convergence. In this step, we obtain policy functions $m' = g_m(x, s; G)$, $h = g_h(x, s; G)$ and $c = g_c(x, s; G,)$ in the CM.

All functions are being parameterized using cubic splines.

5 Solve for $G(m, s)$ using $F(m, s)$ and $< Q, D >$. The, using $G(m, s)$, and policies m' compute $F'(m, s)$. Update F and G until convergence.

6 Verify if markets for money clears. If not, adjust prices and return to step 1. Repeat until the market clears.

7 Check if the government budget constraint is satisfied. If not, go back to step 1. Repeat until the government budget constraint is satisfied.

8 Finally update the terms of trade $< Q, D >$ using the value function $W(x, s)$. Check if the newly calculated terms of trade are close enough to their previous values. If not, return to Step 1. If so, it is done.