The Classical Cost of Inflation Along the Income Distribution

Jackson Mejia*

December 6, 2022

Abstract

I study the welfare cost of deviating from the Friedman rule along the income distribution. As a first step, I provide a structural framework for studying heterogeneous agent money demand in the context of a shopping time model with idiosyncratic risk. With that framework, I estimate money demand curves for low-, middle-, and high-income households using microdata from the Survey of Consumer Finance over the period 1989-2019. Low-income money demand is perfectly interest-inelastic, while middle-income demand is roughly as elastic as aggregate money demand and half as elastic as high-income money demand. Using these demand curves, I quantify the welfare cost of deviating from the Friedman rule. Numerically, a 5% nominal interest rate costs 0.54%, 0.66%, and 0.76% of consumption for low-, middle-, and high-income households, respectively.

*Department of Economics, Massachusetts Institute of Technology. Email: jpmejia@mit.edu. I am grateful to Ricardo Caballero, Tomàs Caravello, Joel Flynn, Basil Halperin, Iván Werning, Christian Wolf, and the MIT Macro Lunch for useful comments and suggestions.

1 Introduction

As inflation once more rears its ugly head, questions of distributional impact come to the fore. Naturally, these questions come in two forms. First, how are costs and benefits distributed? Second, what causes the dispersion in costs? Unsurprisingly, the answers to these questions are inextricably linked. For example, *borrowers* benefit from an *unexpected, transitory* inflation at the expense of *creditors* (Doepke and Schneider 2006). More generally, a focus on short-run unexpected inflation through sticky wages and prices (Yang 2022), quality adjustment, consumption bundles (Jaravel 2019; Baqaee, Burstein, and Koike-Mori 2022), and so on, has left largely unresolved the question of who is hurt the most in the long run.

The fact that the distribution of the long-run, classical cost of inflation—which can be rephrased as the distributional cost of deviating from the Friedman (1969) rule—has been left largely unaddressed is perhaps unsurprising. Although aggregate estimates of the classical cost are comparable to or dwarf in magnitude other such costs (Bils and Chang 2003; Ascari, Phaneuf, and Sims 2018), estimation relies on a concept that many economists cast aside long ago. In particular, quantifying the welfare cost via the classical channel requires a stable money demand curve; it is roughly the area under the inverse demand curve. Hence the inattention stems to some extent from the perceived collapse of stable money demand (Ireland 2009). However, the resurrection of stable money demand by Lucas and Nicolini (2015) has reinvigorated research in the classical monetarist program (Benati et al. 2021; Alvarez, Lippi, and Robatto 2019; Gao, Kulish, and Nicolini 2020; Benati and Nicolini 2021), giving credence to a "distributional monetarism" which considers the composition of money demand and the resulting distribution of welfare cost.

I bring new evidence to light on the cost of deviating from the Friedman rule along the income distribution. Using microdata from the Survey of Consumer Finance (SCF) over the period 1989-2019. I estimate relationships between real balances, consumption, and nominal interest rates over the income distribution. Under unitary consumption elasticity, the money-interest relation can be summarized by two parameters: the interest elasticity and the average level of the money demand function. For a broad class of theoretical models, this is sufficient to recover the welfare cost of inflation. Observationally, there is a clear relationship between these parameters and the income class a household belongs to. First, the money-consumption ratio, *i.e.*, the average level of money demand, is increasing in income. Second, the interest semi-elasticity of money demand is likewise increasing in income. Taken together, these observations suggest considering how the composition of money demand affects aggregate money demand and how that, in turn, affects the distributional cost of deviating from the Friedman rule.[1]

I rationalize these observations about the relationship between money holdings and interest rates along the income distribution with an Aiyagari-style shopping time monetary model augmented with bonds in which a share of the population has access to credit markets while the remainder do not. This gives rise to state-dependent money demand, with hand-to-mouth households exhibiting perfectly inelastic demand. Building on this, I develop a general framework for analyzing the classical cost of inflation with heterogeneous agents. From the household's perspective, the relevant object is the expected present value of the sequence of welfare costs under the stationary distribution of income states. In the limiting case where income states are absorbing—*i.e.,* a rich household is always rich—quantification coincides exactly with the representative agent approach. One can simply estimate a rich household money demand curve and compute the area under its inverse. However, in modern economies there is substantial economic mobility. Rich households may become poor and vice versa. In the process of shifting states, households may submit different money demand curves. Consequently, the welfare cost may differ for each household period-by-period even in the long run when the aggregate economy is stationary.

Toward quantifying welfare costs, I categorize households as low-, middle-, and high-income and estimate money demand curves for each group following the implied regression

1. Despite the fact that many central banks do not in practice set the return on bonds equal to the return on money, this is optimal policy in many monetary models. da Costa and Werning (2008) show that this result continues to hold in quite general settings for heterogeneous agent models.

specification from the shopping time model. Low-income money demand is perfectly inelastic, while middle-income money demand is about half as elastic as high-income demand. If viewed in the context of recent empirical work on credit constraints along the income distribution (Kaplan and Violante 2014; Campbell and Hercowitz 2019), this is unsurprising. However, middle-income households are as elastic as the aggregate economy. Given those estimates, a 5% nominal interest costs low-income households around 0.54% of consumption, middle-income households 0.66%, and high income households 0.76%. All of these estimates are within the range of aggregate estimates provided by Benati and Nicolini (2021).

My approach to the distributional costs of inflation marries two literatures. On the one hand, there is a long tradition of computing the welfare cost of inflation using the area under the money demand curve, starting with Bailey (1956) and continuing to Benati and Nicolini (2021). The consensus welfare cost is around 1% of income. Lucas (2000) and Alvarez, Lippi, and Robatto (2019) show that this is a valid approach for many classes of monetary models, including money in the utility function (MIU), shopping time, and inventory theoretic models. With heterogeneous agents, da Costa and Werning (2008) show that this constitutes a lower bound on the welfare cost. On the other hand, a more recent literature quantifies distributional welfare costs in the context of calibrated models.[2] Until recently, studies in the vein of Erosa and Ventura (2002), Akyol (2004), Wen (2015), and Allais et al. (2020) were limited in their ability to accurately compute welfare costs given the paucity of data. In general, they find very large welfare costs for low-income households. Part of this stems from a specification error. With cash-in-advance (CIA) models—which many of these are—it is difficult to match money-consumption ratios with actual data. In particular, although the money-consumption ratio is empirically increasing in income, the opposite tends to happen in CIA models, resulting in overstated costs for low-income households. Compared with my study, Cao et al. (2021) and Cirelli (2022) are most closely related. The former use microdata from Canada to calibrate an overlapping generations

2. Perhaps the earliest attempt to quantitatively answer this question is from Budd and Seiders (1971), but the literature did not take off until recent decades.

model. Although their policy experiment differs from mine because it studies unexpected inflation, they find very large welfare costs for low-income households. The latter studies the distribution of the welfare cost of inflation via the precautionary savings channel and finds that while the average is approximately zero, low-wealth households are hurt the most. However, because this cost is an order of magnitude smaller than what I find and the average is zero, I ignore the precautionary motive for the remainder of the paper.

2 Two Stylized Observations on Cross-Sectional Money Demand

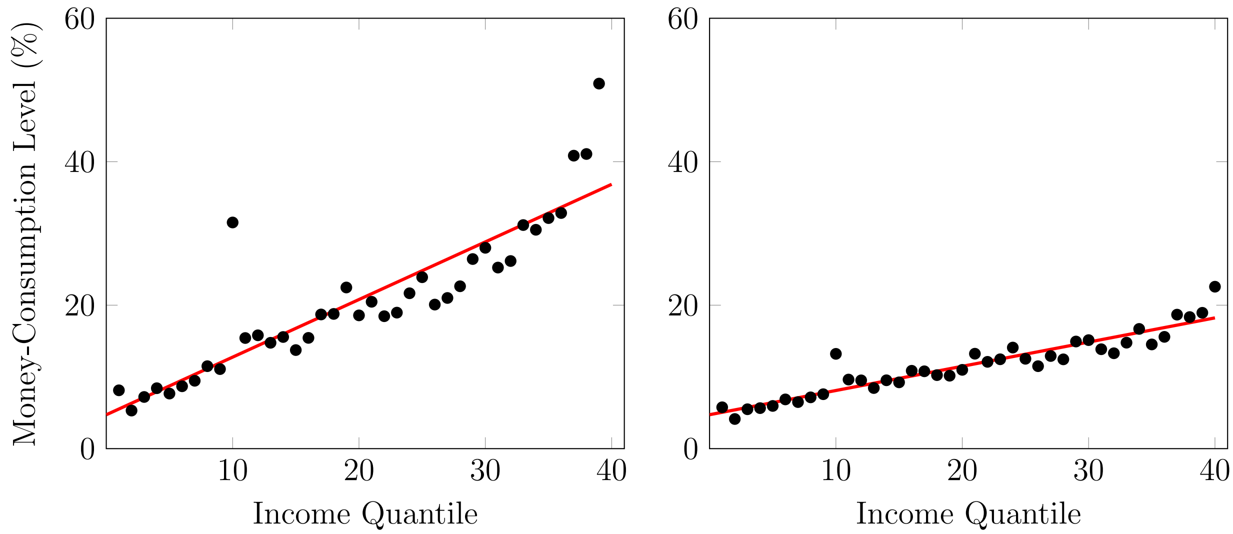
I start by making two stylized observations on the relationship between the money-consumption ratio and income. First, I show that the money-consumption ratio is increasing in income. Second, I show that the interest semi-elasticity is increasing in income. Taken together, these observations suggest that both the level and slope parameters for any money demand function will be increasing in income.

2.1 Observation 1: The Money-Consumption Ratio is Increasing in Income

Using data from the Survey of Consumer Finance (SCF) and the Consumer Expenditure Survey (CEX) from 1989-2019, I proceed in four steps to show that the money-consumption ratio is increasing in income.[3] First, I group households into forty income quantiles and compute the annual average money-consumption ratio for each group separately. Second, I take the average money-consumption ratio within each income quantile over the whole time sample. Third, I plot the unconditional mean of the money-consumption ratio against income quantile. Fourth, I compute standard errors via wild bootstrap. I provide results

3. See Appendix E for details on data construction.

for both the traditional measure of M1, i.e., the sum of checkable deposits and currency, as well as for NewM1 from Lucas and Nicolini (2015), which adds interest-bearing checkable deposits. Results are in Figure 1, where Figure 1a plots the NewM1 money-consumption ratio and Figure 1b plots the money-consumption ratio excluding interest-bearing checkable deposits.



(a) Average level of NewM1

(b) Average level of M1

Figure 1: Average level of NewM1 and M1 over the income distribution with line of best fit plotted in red, where M1 excludes interest-bearing checkable deposits.

Clearly, the money-income ratio is increasing in income. To check statistical significance, I regress the money-consumption ratio on income quantile and a constant:

$$m_j = \hat{\alpha} + \hat{\beta} \times \text{Income Quantile}_j + \epsilon_j,$$

where m_j is the money-consumption ratio for quantile j. Results for each of NewM1 and M1 are in Table 1. The 95% confidence intervals for each of the intercept and slope are calculated via wild bootstrap with 10,000 draws.

In general, there is a positive and statistically significant relationship between income and money holdings. These results seem to fly in the face of intuition. In general, we may think that either money demand is non-homothetic—in which case the money-consumption ratio should be constant—or that it should be *decreasing* in income. In particular, if some subset of households are hand-to-mouth and cash is required for transactions, then the money-consumption profile should be decreasing in income. This is exactly the result in the calibrated models of Erosa and Ventura (2002) and Akyol (2004). However, that is not the empirical relationship in the data.

Dep. Variable	Intercept	95% CI	$\hat{\beta}$	95% CI
$\frac{\text{NewM1}_j}{\text{Consumption}_j}$	4.723	[1.621, 7.828]	0.803	[0.647, 0.959]
$\frac{\text{M1}_j}{\text{Consumption}_j}$	4.739	[4.318, 5.813]	0.338	[0.284, 0.356]

Table 1: Regression of level of money holdings m_j on income quantile and a constant. 95% confidence interval calculated via wild bootstrap.

Moreover, the relationship holds regardless of which monetary aggregate is used, *i.e.*, exclusion of interest-bearing checkable deposits has no bearing on the qualitative result. However, it is clearly true that the quantitative impact of including interest-bearing checkable deposits is large and points in the direction of an even starker contrast along the income distribution. Some of these accounts require a minimum balance, with the predictable result that only high-income people hold them. Hence it is unsurprising that the relationship between average money holdings and income would level up with their inclusion.

2.2 Observation 2: The Interest Semi-elasticity is Increasing in Income

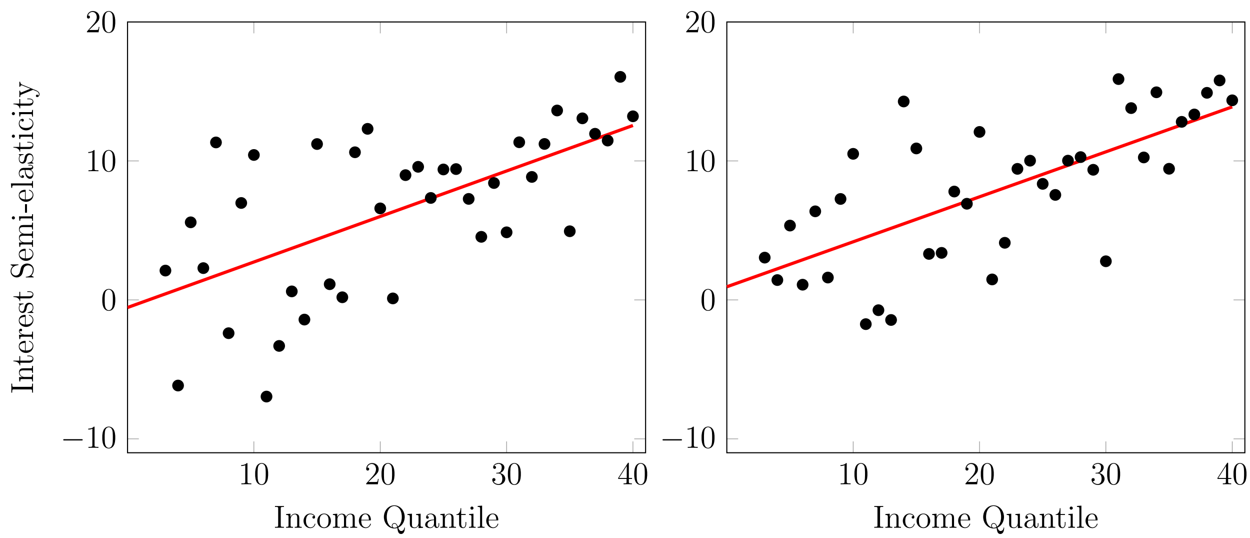
Next, I document a relationship between the interest semi-elasticity of money holdings and income. The procedure is straightforward. First, within each income quantile j, I regress the money-consumption ratio on a constant and the level of the short rate i^d

$$\log m_{j,t} = \alpha_j - \eta_j i^d_t + \epsilon_{j,t}$$

for each of NewM1 and M1. In Figures 2a and 2b I plot the point estimate of the semi-elasticity $\hat{\beta}_j$ for each income quantile for NewM1 and M1, respectively. Then, to get an idea of the statistical relationship between money holdings and income, I regress the estimated semi-elasticity η_j for each income quantile on income quantile and a constant:

$$\hat{\eta}_j = \hat{\gamma} + \hat{\kappa} \times \text{Income Quantile}_j + \epsilon_j,$$

This estimated relationship is the line of best fit in each plot. Next, to check that the positively sloped relationship between interest semi-elasticity and income has statistical significance, I compute a 95% confidence interval for each of the intercept and the slope via wild bootstrap with 10,000 draws. Results for the point estimates and associated confidence intervals are in Table 2.



(a) Interest semi-elasticity of NewM1

(b) Interest semi-elasticity of M1

Figure 2: Interest semi-elasticity of NewM1 and M1 over the income distribution with line of best fit plotted in red, where M1 excludes interest-bearing checkable deposits.

The relationship between these variables is clear. As income grows, households become more interest-elastic. Indeed, this is true (with essentially the same coefficient) regardless of whether M1 or NewM1 is used as the measure of money. This is an unsurprising result light of recent research on the behavior of hand-to-mouth and liquidity constrained households over the income distribution. Standard results imply that low-income households—which are often hand-to-mouth—have little financial flexibility and have very high marginal propensity to consume, which suggests that their behavior would be largely invariant to movements in nominal interest rates and inflation. But even middle class households seems relatively interest-inelastic, which may be slightly more surprising. However, Campbell and Hercowitz (2019) highlight that as much as 80% of the middle class behave *as if* they are liquidity constrained, largely due to lumpy term savings motives. This behavior would make them likewise inelastic to changes in the nominal interest rate and inflation. Similarly, results from Sahm, Shapiro, and Slemrod (2012) and Kaplan and Violante (2014) indicate that the middle class is highly constrained, which would again make them less sensitive than high-income households to interest rate changes.

Dep. Variable	Intercept	95% CI	$\hat{\kappa}$	95% CI
η^{NewM1}	0.052	[0.049, 0.055]	0.328	[0.241, 0.413]
η^{M1}	0.038	[0.035, 0.041]	0.325	[0.274, 0.379]

Table 2: Regression of interest semi-elasticity of money holdings η_j on income quantile and a constant. 95% confidence interval calculated via wild bootstrap. κ is the slope of the line of best fit in each plot and can be thought of as the increase in semi-elasticity given a one unit increase in income quantile.

Throughout the remainder of the paper, my preferred specification considers NewM1 since it behaves qualitatively similarly to M1.

3 The Classical Channel with Heterogeneity

The stylized observations of the previous section may be statistically significant, but that has little to do with economic significance. Correlations, even statistically significant ones, are meaningless if not grounded primarily in economic theory and secondarily in sound econometric methods. Since "giving colorful names to statistical relationships is not a substitute for economic theory," (Lucas 2000, p. 253) I set up a benchmark heterogeneous agent monetary model with idiosyncratic risk to study the classical welfare cost of inflation. The goals of this are threefold. First, I extend the representative agent shopping time framework into a heterogeneous agent model to give a foundation for a cross-sectional money demand function. Second, using this framework, I motivate via credit constraints how cross-sectional interest elasticities may differ. This has important general equilibrium implications for monetary policy, namely that if there are changes in the share of hand-to-mouth agents, then the aggregate money demand function may change as well. Finally, I provide sufficient statistics for measurement of the welfare cost of inflation from the household's perspective. In particular, with heterogeneous households, the relevant welfare cost is no longer simply the area under the money demand curve. Because households have some probability of switching states and hence switching to a different demand curve, the relevant metric is now the discounted expected welfare cost generated by a sequence of money demand curves at a particular inflation rate rather than simply the area under the money demand curve at a single point in time.

3.1 Environment

The environment is a closed labor-only production economy with ex-ante identical households facing idiosyncratic risk. There is no aggregate risk. Time is discrete and runs from $t = 0$ to ∞ and households have rational expectations. Households can be in a discrete number of income states indexed by s in which they have access to a state-dependent transactions technology. I assume that the mass of agents in idiosyncratic state s is always equal to $\pi(s)$, the probability of state s in the stationary distribution of Ψ.

Production. There is a continuum of households of unit mass supplying labor. A representative firm uses linear labor technology to produce a final consumption good with price

p_t so that aggregate output is given by

$$Y_t = \int A_t^j n_t^j \mathrm{d}j = \bar{A}\bar{n}, \tag{1}$$

where A_t^j is the productivity of household j at time t, n_t^j is the labor supply of that household, and Y_t is aggregate output. Under the ergodic distribution of the process for productivities, average productivity and average labor supply are assumed constant, which means that output and aggregate consumption are likewise constant.

Households. There is a unit mass of ex-ante identical households indexed by j which have preferences over consumption according to

$$\mathbb{E}\sum_{t=0}^{\infty} \beta^t U\left(c_t^j\right), \tag{2}$$

where β is the discount factor and U is differentiable, increasing, and concave.

At the beginning of every period, household j observes its state and enters a financial market to purchase money and bonds. Starting with nominal wealth W_t^j, the household allocates wealth between money M_t^j and interest bearing bonds B_t^j such that

$$M_t^j + B_t^j \leq W_t^j. \tag{3}$$

Bonds pay net interest rate i_t. However, households have state-dependent access to credit markets pinned down by a parameter $\phi^j \geq 0$. Hence the household-specific net interest rate is given by $i_t/(1+\phi^j)$, where a positive ϕ^j corresponds to a lower household-specific return on bonds and a correspondingly higher bond price. Note that as ϕ^j goes to infinity, the bond return goes to zero and the price goes to infinity so that the household essentially becomes hand-to-mouth. This is a reduced-form way to capture credit constraints without appealing to a borrowing constraint. Later on, to put more structure on the problem and connect it to the empirical literature on credit constraints, I consider some share of households with $\phi^j = 0$, so that they are standard agents on their Euler equation, while the remaining share has $\phi^j \to \infty$. This essentially prices those households out of the bond market.

After observing its state and allocating wealth, the household splits in two to produce output and shop. In particular, given a time endowment of one, the household can allocate time between shopping time z_t^j and labor n_t^j:

$$z_t^j + n_t^j = 1. \tag{4}$$

With stochastic productivity A_t^j, households therefore earn total labor income $A_t^j n_t^j$. In addition to shopping time z_t^j, households transact using two additional inputs: the real value of bonds B_t^j/p_t and real balances M_t^j/p_t. State-dependence enters the transaction technology by altering the relative efficiency of using money versus other means of transacting. For example, one can imagine that in practice higher-income households employ more money than time spent shopping because the opportunity cost of spending time shopping is higher than simply demanding more money. I assume the transaction technology is Cobb-Douglas with state-dependent share parameters ν^j on real balances and γ^j on alternative transaction methods and level parameter ζ^j raised to the power $1/\alpha^j$:

$$c_t^j \leq \left[\zeta^j \left(\frac{B_t^j}{p_t} + A_t^j z_t^j \right)^{\gamma^j} \left(\frac{M_t^j}{p_t} \right)^{\nu^j} \right]^{\frac{1}{\alpha^j}} \tag{5}$$

There are three important notes to make about the transactions technology here. First, I model bonds and shopping time as perfect substitutes while they together smoothly substitute against real balances (depending on the values of the parameter γ^j and ν^j). The idea here is that money must be a primary means of transacting, but that households will choose to either use bonds or shopping time as a second means of transacting. The perfect substitutability between bonds and shopping time underscores the notion that rich households which suddenly become low-income may wish to use existing wealth to transact via bonds, whereas poorer households may wish instead to use time. Second, raising the technology to the power $1/\alpha^j$ allows for additional generality; households may not have a unitary consumption elasticity. This becomes important during the estimation procedure. Third, when the constraint binds (as it must in equilibrium), the problem becomes isomorphic to a bonds-in-utility and money-in-utility model. When $\gamma^j = 0$ and $\zeta^j = \alpha^j = 1$, the model is then a cash-in-advance model.

Households also receive a state-dependent lump-sum transfer T_t^j. Putting these inflows and outflows together, next period wealth is given by

$$W_{t+1}^j = M_t^j + B_t\left(1 + \frac{i_t}{1+\phi^j}\right) + T_t^j + p_t A_t^j n_t^j - p_t c_t^j. \tag{6}$$

Putting all of this together, households maximize utility (2) by choosing consumption c_t^j, nominal balances M_t^j, nominal bonds B_t^j, shopping time z_t^j, labor time n_t^j, and next-period nominal wealth W_{t+1}^j subject to the portfolio allocation constraint (3), the time endowment (4), the transactions technology (5), and the asset evolution constraint (6).

Government. The government has a simple role: make lump-sum transfers T_t^i to households financed by increases in the money supply so that average transfers are equivalent to increases in the money supply per agent:

$$\int T_t^j \mathrm{d}j = \int M_t^j \mathrm{d}j - \int M_{t-1}^j \mathrm{d}j.$$

That is, given an aggregate supply of money M_0, which is equivalent to the average supply of money $\int M_0^i \mathrm{d}i$ at time zero, the government finances a sequence of transfers. Government bonds are in zero net supply.

or analytical simplicity, I assume that some share ψ of the households have $\phi^j = 0$ and the remaining $1-\psi$ have $\phi^j \to \infty$. This makes it similar to a spender-saver model. As noted earlier this is akin to cutting off access to the credit market for $1-\psi$ of the population. Given recent empirical evidence on the fact that many low- and middle-income households are borrowing constrained, this is a reasonable assumption. Note that when $\phi^j \to \infty$, it effectively makes bonds perfectly interchangeable with money. As is well-known, this effectively means that the credit-constrained share of the population exhibit a constant demand for money regardless of the nominal interest rate (Lucas and Stokey 1987). I formalize this notion in the context of this model in Proposition 1.

Proposition 1. *The money demand function for households in an unconstrained group is given by*

$$\frac{M_t^j}{p_t}(i_t) = \Upsilon^j (c_t^j)^{\Gamma^j} \left(1 + i_t\right)^{-\sigma^j}, \tag{7}$$

while the money demand function for the constrained group is given by a constant

$$\frac{M_t^j}{p_t}(i_t) = \Upsilon^j (c_t^j)^{\Gamma^j}, \tag{8}$$

where Υ^j, σ^j, and Γ^j are composite parameters.

Proof. See Appendix A.1

Note first that money demand for unconstrained households is a function of the gross nominal interest rate, which suggests finite money holdings at the zero lower bound. This is a particularly interesting outcome of the shopping time model augmented with bonds because it suggests also that the interest elasticity is a function of the nominal interest rate rather than a constant. Indeed, the interest elasticity is given by

$$\varepsilon^j = \sigma^j \frac{i_t}{1 + i_t}, \tag{9}$$

so that households become more interest elastic at higher interest rates. This is perhaps a compelling result in light of evidence from Sims (2003) and Braitsch and Mitchell (2022), namely that attentiveness to inflation is a nonlinear function of the inflation rate. In this model, attentiveness can perhaps be mapped directly to elasticity. Moreover, the result stands in sharp contrast to much of the modern literature, which typically focuses on money demand functions with constant elasticity. Finally, note that since households are predicted to hold finite money holdings at the zero lower bound—which recent experience shows is an empirical reality—then the welfare cost will be strictly lower than the case where money demand exhibits constant elasticity (as in the log-log case).

Next, clearly when households vary in their credit access, their interest elasticities will likewise vary. In particular, $\phi^j > 0$ mutes the response of money demand to a change in the nominal rate. In the limit where, in the shorthand of this problem, some unconstrained households have perfect access to credit markets while the remainder have no access (constrained), the former exhibit a money demand function that looks quite similar to the representative agent formulation of Lucas (1988) and Teles and Zhou (2005), while the latter are perfectly interest inelastic. This finding is, in a sense, obvious. If I cannot buy a product, then of course I do not care about its price. And yet, these findings point to an underappreciated notion, namely that aggregate money demand depends on the composition of money demand, which depends in turn on the share ψ of unconstrained households.

Proposition 2. *An increase in the share ψ who are not credit-constrained increases the elasticity of the aggregate money demand curve.*

Proof. See Appendix A.2

This follows almost immediately from Proposition 1, but it has a surprising implication for the aggregate money demand function. In particular, it suggests that a necessary condition for the stability of an aggregate money demand curve in any economy is constancy in the composition of household access to credit markets. This in itself is a function of both technological innovation and regulatory policy. Given that both have shifted substantially in the last century, the results of Lucas and Nicolini (2015) and Benati et al. (2021), namely that the long-run aggregate money demand curve is stable both in the United States and in a cross-section of countries, is surprising. However, it is also reassuring because it provides support for the idea that there is, broadly at least, stability in the composition of money demand, which in turn provides support for considering cross-sectional money demand curves.

That the composition of money demand has implications for aggregate money demand suggests that optimal conduct of monetary policy depends not only on the maintaining a stable growth rate of money (or managing the money supply indirectly via interest rate changes), but on managing credit market access. This echoes a strand of literature emphasizing the importance of the credit channel but also highlights a difficulty in the practice of actually managing policy. Access to credit markets, which I model in a highly-reduced form way, depends on many factors outside the control of the monetary authority. Although it is true that the Federal Reserve has both regulatory and monetary authority, its ability to exert tight control over both is dubious regardless of monetary instrument.

3.2 Sufficient Statistics

A natural starting point for computing the welfare cost of inflation is the canonical approach of Bailey (1956) and Lucas (2000). Under superneutrality in a representative agent economy, given some money demand function, the welfare cost of inflation is defined by considering the consumer surplus that could be gained by changing the steady state nominal interest rate from i_0 to i_1. Conveniently, after scaling money demand by consumption, we can interpret this as the fraction of consumption households would require as compensation to make them indifferent between living in a steady state with interest rate i_0 versus the otherwise identical steady state with interest rate i_1 (p. 250). If the comparison is between a zero-percent nominal rate and i_0, then the estimate also gives the welfare cost of ignoring the Friedman (1969) rule. In a representative agent context, for some money demand curve scaled by consumption $m(c, i)$, the welfare cost of inflation is

$$W(i) = \int_0^i m(c, i)\mathrm{d}i - im(c, i),$$

where the second term rebates seigniorage lump-sum back to the household. Under a unitary consumption elasticity assumption, the information requirements are small: only an estimate of a level and a slope parameter is necessary. If the consumption-elasticity is not unity, knowledge of that parameter is likewise required. The inflation cost computed via the area under the demand curve is purely *classical* in the sense that it is a long-run cost equilibrium cost absent any frictions.[4]

Compared to the representative agent setting, there are two key differences when we wish to consider heterogeneity. First, we have to be more careful about redistribution of seigniorage revenues. In general, it is not the case that seigniorage will be rebated lump-sum to the household it was taken from. Second, although the economy itself may be in a stationary equilibrium, the agents themselves may move in and out of states. If there were no economic mobility, then households would remain in their state for all time, so we could simply compute welfare in the standard way as the area under a single demand curve. However, because households may move from state to state, the area under the demand curve changes (in expectation) period by period. Hence the relevant object for measurement of welfare is no longer the area under a static money demand curve, but the expected discounted value of future welfare loss, where the welfare loss may differ in each period because of the possibility of shifting to a different demand curve. In that sense, this is a "dynamic" welfare metric in comparison to the "static" representative agent formulation. For example, an agent may start poor today under the stationary distribution and potentially benefit from high steady state inflation via redistribution, but later move into a new state with a new money

4. However, the area under the inverse demand curve is technically neither an exact measure of welfare loss nor is it the cost of inflation. More precisely, it measures the welfare loss generated by positive interest rates using the area under the Marshallian demand curve. Positive interest rates are a violation of the Friedman rule regardless of the inflation rate. Next, although the area under the Marshallian demand curve is usually an approximation welfare, Alvarez, Lippi, and Robatto (2019) and Benati and Nicolini (2021) show that it is close to exact for a large class of monetary models.

demand curve. Consequently, I measure welfare loss for each household as

$$\frac{1}{\bar{c}}\mathbb{E}\sum_{t=0}^{\infty}\beta^t W_{t,k}(i,c), \tag{10}$$

where

$$W_{t,k}(i,c) = \int_0^i M_{t,k}(c,i)\mathrm{d}i + T_{t,k}, \tag{11}$$

where $W_{t,k}$ is the welfare loss from interest rate $i > 0$ for a household in state k at time t under money-demand curve $M_{t,k}(c,i)$ net of seigniorage transfers $T_{t,k}$. $\bar{c}$ is average consumption under the stationary distribution, so welfare loss is in units of permanent consumption.[5] Therefore, compared to the representative agent case, we need two extra pieces of information: how transfers from the government flow to each household and a transition matrix for households. This is in addition to the consumption and interest elasticities and the level parameter required in the simpler representative-agent case.

4 Data and Identification

4.1 Data

Money demand estimation requires three components at a minimum: data on money balances, consumption, and nominal interest rates. Since I study cross-sectional differences in money demand with the ultimate objective of measuring welfare, I require two additional elements missing from aggregate studies of money demand: microdata and a transition matrix between states. In particular, it is necessary to have data on household-level consumption and income to make the relevant cross-sectional distinctions.

5. In general, studies of aggregate money demand are framed in terms of the money-income ratio (Lucas 2000; Ireland 2009; Benati and Nicolini 2021). However, following Mankiw and Summers (1986) and Lucas (1988), I use consumption as a scale variable because consumption is plausibly more closely related to permanent income than realized income. If the objective is long-run money demand estimation, then consumption is more sensible. Also, if the primary purpose of money is for transactions, and transactions at the household level are for consumption, then consumption is again the relevant object.

4.1.1 Data: Money Balances and Consumption

Proposition 1 suggests simple procedure for estimation. The idea is to bin households into groups which are ex ante likely to exhibit similar money demand. Given the model, that means splitting along the lines of who is likely to be credit constrained. But even within these two groups, there is likely to be substantial heterogeneity. Recent research from Campbell and Hercowitz (2019) and Kaplan and Violante (2014) point to as much as 80% of the middle class being credit constrained. With that in mind, we would expect much of the middle class—which has much higher consumption than their low-income counterparts—to be somewhat credit constrained. From our earlier observations, the middle class appears more interest-elastic and holds hgher real balances than low-income households, which suggests a significantly different money demand curve. As a result, I group households into low-, middle-, and high-income. In addition to this structural reasoning, it is also practical: policy is often defined over its costs and benefits for low-, middle-, and upper-income households, distinctions which naturally flow into definitions of welfare for politically and economically relevant groups. In the data, I distinguish between income groups using the Current Population Suvey (CPS) ASEC supplement. For every year in the sample, I group households as low-income if they are in the first income quintile for that year, high-income if in the top decile, and middle-income otherwise.[6] These definitions are not particularly important and are robust to changes in where the lines are.

With those cutoffs, I quantify a household-group money aggregate using data from the Survey of Consumer Finance (SCF). The SCF has been conducted every three years since 1989 and contains detailed data on household finances. I use the definition of NewM1 from Lucas and Nicolini (2015) for my group-level monetary aggregate.[7]. In particular, I define

6. Ideally, I would split households more narrowly, but data quality from the SCF is already somewhat dubious even if it is the best we have outside of administrative data (Bhandari et al. 2020).

7. The results are qualitatively similar excluding interest-bearing checkable deposits, which would correspond to the traditional M1 aggregate.

money for each household k as

$$\text{NewM1}_{k,t} = \text{Checking}_{k,t} + \text{Call}_{k,t} + \text{Prepaid}_{k,t} + \text{MMDA}_{k,t}.$$

Ideally I would use a Divisia measure of money since my definition of money includes interest-bearing checking accounts, but its construction at the household level would be very difficult, so I instead proceed with a simple sum.

Next, I construct income group consumption using data from the Consumer Expenditure Survey (CEX), which is conducted annually with a focus on consumption expenditures. In particular, I group households into income quantiles using cutoffs from the CPS ASEC survey, then add up weighted consumption within each income group. See Appendix E for more details.

An obvious issue with this so far is data scarcity. Given that the SCF is administered every three years, there are only eleven years of data available. In an effort to rectify this, I carry out a state-space imputation procedure common in time series macroeconometrics. Following Stock and Watson (2016) and McGrattan (2020), I use related higher frequency data series and first apply a Kalman filter and then a Kalman smoother on detrended data. In particular, I use aggregate NewM1 as well as data on liquid assets from the CEX from the corresponding low-, middle-, and high-income households to impute the missing values in the sample.[8] See Figures 5a, 5b, 5c for plots of real balances scaled by consumption plotted against nominal interest rates for low-, middle-, and high-income households, respectively. Casual observation suggests an increasing relationship between income and the level and elasticity of the money-consumption ratio.

8. See Appendix E.2 for details on the procedure.

4.1.2 Data: Nominal Interest Rates

In lieu of using household-specific interest rates, I use aggregate nominal short rates for three reasons. First, it is very difficult to obtain data on household-specific interest rates, whereas aggregate rates are readily available. The second reason is more subtle. A household-specific nominal interest rate is presumably endogenous to the decisions of the household, whereas an aggregate interest rate is clearly exogenous. Moreover, significant variation in household-specific interest rates will be captured by variation in aggregate rates. Third, the relevant variable for the classical channel is an aggregate interest rate because a household-specific rate will move around for reasons that have nothing to do with policy, whereas the correlation between the short rate and policy is particularly strong. Consequently, I rely on aggregate nominal rates.[9]

Because the time period I examine contains episodes at or near zero percent nominal interest rates, the data reveal important information about the behavior of money demand near the ZLB, something that was not available in many earlier studies. However, there are three complications. First, components of the money aggregate pay slightly different returns. Evidence from Alvarez and Lippi (2009) indicates that cash pays a negative 2% nominal return, while MMDAs pay a weakly positive nominal return. Second, Kurlat (2019) provides strong empirical evidence that deposit spreads depend linearly on the nominal interest rate; this follows from well-documented evidence on market power of banks. Third, many countries have gone far below the ZLB without the consequences predicted by economic theory. For example, Switzerland experienced rates that fell to -1.85%. Following Benati and Nicolini (2021), I use the first two facts to rationalize the third and assume that the relevant nominal interest rate is a functional relationship between the return on bonds and

9. On the other hand, omitted variable bias may be an issue. In particular, it may be wise to include some measure of creditworthiness. However, because I am measuring at the group level rather than the household level, it would be quite difficult to aggregate that upward even if some metric were available. Moreover, I am tentatively confident that the middle class in this sample is perhaps financially similar enough that simply running a separate regression from them is sufficient.

money given as

$$i^d \equiv i^{bonds} - i^{money} = i^{bonds} + a - bi^{bonds} \geq 0,$$

where i^{bonds} is the return on three-month Treasurys and i^{money} is the return on money. For non-zero values of a and b, this implies that the effective lower bound is different from zero, where a non-zero a follows from the findings of Alvarez and Lippi (2009) and non-zero b follows from Kurlat (2019). However, since the U.S. did not experience negative rates and too low of an effective lower bound would push up the welfare cost of inflation, I employ slightly more cautious parameters for a and b than Benati and Nicolini (2021).[10] Hence, consistent with an effective lower bound of approximately -0.25%, I assume $a = 0.225$ and $b = 0.10$. The lower value for a than in Benati and Nicolini (2021) can be rationalized by the relative importance of Alvarez and Lippi (2009) being lower in the U.S. than in the rest of the world. The lower value for b can be justified from a time series perspective: whereas my sample is only after interstate banking laws became liberalized and hence more competitive, aggregate studies examine a longer time period with greater variability in market power. See Cacciatore, Ghironi, and Stebunovs (2015) for the relevant regulatory analysis.

4.1.3 Data: Transition Probabilities

Ultimately, it is necessary to calibrate transition probabilities between low, middle, and high states. I use the longitudinal component of the CPS ASEC to compute a transition matrix. The resulting matrix is

$$\begin{array}{cc} & \text{Next State} \\ & \begin{array}{ccc} \text{L} & \text{M} & \text{H} \end{array} \\ \text{Current State} \begin{array}{c} \text{L} \\ \text{M} \\ \text{H} \end{array} & \begin{bmatrix} 0.6428 & 0.3487 & 0.0085 \\ 0.0906 & 0.8475 & 0.0618 \\ 0.0180 & 0.4331 & 0.5489 \end{bmatrix} \end{array}$$

10. Since Benati and Nicolini (2021) study this problem in an international context, they are forced to be more generous in their parameterization.

which has a stationary distribution of $(0.186, 0.713, 0.101)$. Additional details on construction are given in Appendix E.3.

4.2 Identification

Conditional on having the relevant cross-sectional data, I can estimate a relationship between real balances, consumption, and interest rates. In particular, equation (7) calls for estimating a relationship of the form:

$$\log m_{k,t} = \log \Upsilon_k + \Gamma_k \log c_{k,t} - \sigma_k \log(1 + i_t^t), \tag{12}$$

where $m_{k,t}$ is the real balances of household group k at time t, $c_{k,t}$ is real consumption of household group k at time t, and i_t^d is the nominal interest rate. Υ_k is a constant, Γ_k is the consumption elasticity of group k and σ_k is an interest elasticity parameter.[11] In keeping with traditional analysis of money demand, I interpret these elasticities as structural parameters.

Identification of supply and demand is generally a challenging econometric task, but two assumptions are sufficient to identify the money demand function from observational data. First, in keeping with monetarist thought, I assume that the money supply is exogenously controlled by the monetary authority (Papademos and Modigliani 1990; Bischoff and Belay 2001; Bae and Jong 2007). In that case, the money supply curve is vertical and so if the demand curve is stable, then a simple regression is sufficient to identify the relevant parameters (Stock and Watson 1993). Of course, this is a perhaps contentious assumption since banks create inside money in practice. Even so, it is far from clear that the fact that banks create their own money creates a difficulty for us; banks create their own money according to rules fixed by the monetary authority. Moreover, in the long run, the central bank has control over the money supply, which is precisely the horizon in question. Additionally, without this assumption, the entirety of the money demand literature wilts away. Second, I

11. Different structural models could likewise generate a log-log or a semi-log relationship.

must assume that the income groupings give rise to a stable money demand. This relies on an argument that low-, middle-, and high-income households submit systematically different money demands by virtue of belonging in a particular state of the world. Based on empirical evidence that these groups have historically exhibited different degrees of credit market access and the theoretical demonstration that this will give rise to quite different elasticities of money demand, I find this argument convincing. Then, conditional on believing that the income groupings are sound, if both money demand and interest rates are nonstationary and are cointegrated, then the money demand function for each group can be estimated via the dynamic OLS estimator of Stock and Watson (1993).

5 Parameter Estimation

Before computing parameter estimates, I first perform two checks. First, I test for non-stationarity for each of real balances, real consumption, the money-consumption ratio, and the gross interest rate within each income group. Second, I perform a joint cointegration test between the three variables, as well as a cointegration test between the money-consumption ratio and the log gross interest rate.

5.1 Econometric Tests

Unit Root Test

In Table 3, I report results for the augmented Dickey-Fuller test (with drift) against a null hypothesis of a unit root. I test the stationarity of the log-level of real balances, the log-level of real consumption, the log money-consumption ratio, and the log gross interest rate for each of the three income groups as well as the entire SCF sample and the aggregate economy. Per the results of the test, none of the series are stationary. On the other hand, with such a small time series, the Dickey-Fuller test has low power to detect stationarity. Given that, I also perform a Phillips-Perron test with drift for stationarity. Results are in

Table 4. In this case, only the low-income variables exhibit stationarity, while the series for middle-income households, high-income households, all households, and the whole economy are non-stationary. It is unsurprising that these variables are non-stationary and so we should not be excessively concerned about low power.

Given the discrepancy between the Phillips-Perron and the augmented Dickey-Fuller test, I appeal to economic theory for how to evaluate the stationarity of the low-income variables. Given evidence on weak income and consumption growth over the time period covered for low-income families, it seems quite plausible that these series would be stationary (CRS 2020). Moreover, Figure 5a suggests an essentially constant money-consumption ratio. This is largely consistent with the theory presented in Section 3, namely that households shut out of the bond market will hold a largely invariant amount of money as a share of consumption. Since low-income households *are* empirically credit-constrained, it then seems relevant to apply the theory developed above to them. Hence I apply the results of the Phillips-Perron test and call the income-group-specific variables stationary.

Cointegration Test

Again because of concerns about power driven by the small sample size, I consider two different tests of cointegration. First, I apply the Johansen test. Second, I apply the Phillips-Ouliaris $\hat{P}_u$ test. I do this for two different specifications. The first tests for a cointegrating vector between the log of real balances, the log of consumption, and the log gross interest rate, while the second tests for cointegration between the log money-consumption ratio and the log gross interest rate. Results are in Table 5 and Table 6, respectively. Both tests agree at the 5% level that there is a cointegrating relationship between these variables for middle-income, high-income, the whole SCF sample (all households), and the aggregate variables for both specifications. For completeness, I also report results for low-income households, but of course those are made irrelevant by the conclusion that the low-income variables are stationary.

5.2 Estimation

Estimation: Low Income Households

The shopping time model in Section 3 calls for a regression of log real balances on a constant, log real consumption, and the log interest rate. The issue here is that the interest rate series is non-stationary whereas it *is* stationary for real balances and consumption for low-income households. Consequently, running the original regression would lead to spurious results. I make the interest rate series stationary via first-differencing and then regress first-differenced real balanced on a constant, first-differenced real consumption, and first-differenced nominal interst ratess:

$$\log m_{t,L} - \log m_{t-1,L} = \alpha^L + \beta_c^L \left(\log c_{t,L} - \log c_{t-1,L}\right) + \beta_i^L \left(\log(1+i_t^d) - \log(1+i_{t-1}^d)\right) + \nu_{t,L}. \tag{13}$$

Standard results show that the coefficients β_L^c and β_L^i recover the elasticity parameters. Note that although this estimation fails to recover the level parameter, that turns out not to matter if the elasticity is zero. In this case, the estimated elasticity parameters are not significantly different from zero, which is equivalent to saying that low-income households have perfectly inelastic demand for money. Results are in Table 7. I also report results for regressing the first-differenced log money-consumption ratio on the first-differenced log interest rate.

These results are unsurprising in lighting of both econometric and economic theory. In particular, we should not expect a short-run relationship between money holdings and the interest rate if we cannot plausibly expect a long-run relationship between them either. Moreover, that household variables are stationary while the interest rate is clearly non-stationary strongly suggests that there is not a relationship between them. Standard results imply that low-income households—which are often hand-to-mouth—have little financial flexibility and have very high marginal propensity to consume, which suggests that their behavior would be largely invariant to movements in nominal interest rates and inflation.

Given that result, I instead estimate a money demand curve for each group by simply regressing the level of money demand on an intercept. This implies that households in this group keep a constant share of income as highly liquid money. Results, given in the bottom row of Table 7, indicate that low-income households exhibit a constant money-consumption ratio of around 0.07.

Estimation: Middle and High Income

Next, given the cointegration results for the middle and high income groups, I estimate money demand curves for both of them. This consists of two different types of regressions. The first is static OLS, linking income group j's current real balances m_t^j to a constant, consumption c_t^j and the current nominal interest rate i_t^d and a constant:

$$\ln m_{t,j} = \alpha_H + \beta_c^j \ln c_{t,j} + \beta_i^j \ln\left(1 + i_t^d\right) + \epsilon_{t,j}, \tag{14}$$

where β_c^j and β_i^j correspond to the consumption and interest elasticity parameters for each income group j. Note that this is exactly the regression equation implied by equation (7). I report results for this specification in the top row of Tables 8 and 9.[12]

However, results from Stock and Watson (1993) indicate that dynamic ordinary least squares (DOLS) is a more efficient estimator than static OLS in this setting, which suggests a closer reliance on this specification than static OLS. In particular, I estimate

$$\ln m_{t,j} = \alpha^j + \beta_c^j \ln c_{t,j} + \beta_i^j \ln\left(1 + i_t^d\right) + \sum_{t=-p}^{t=p} \ln\left(1 + i_t^d\right) + \sum_{t=-p}^{t=p} \ln c_{t,j} + \epsilon_t. \tag{15}$$

for up to $p = 2$ leads and lags of the nominal interest rate and consumption. Adding lags and leads controls for possible correlation between the interest rate, consumption, and the residual from the cointegrating relationship linking moneyholdings and the interest rate.

12. I also report results for a more traditional specification that links the money-consumption ratio to a constant and the nominal interest rate. Additionally, I report results for all households and the whole economy in Tables 10 and 11.

However, this not account for serial correlation in the error term, so I compute Newey-West standard errors for various values of the lag truncation parameter q. Estimates are in Tables 8 and 9. The standard errors vary little depending on the lag truncation parameter and the estimates for the relevant parameters are statistically significant.

The elasticity parameter is about twice as large for high-income households than for middle-income households. This result is unsurprising in light of the theory presented earlier. In particular, Campbell and Hercowitz (2019) highlight that as much as 80% of the middle class behave *as if* they are liquidity constrained, largely due to lumpy term savings motives. This behavior would make them likewise somewhat inelastic to changes in the nominal interest rate and inflation. Similarly, results from Sahm, Shapiro, and Slemrod (2012) and Kaplan and Violante (2014) indicate that the middle class is constrained, which would again make them less sensitive to changes in nominal rates.

Going forward, I rely on the specification with $p = 1$. This implies an elasticity parameter of around 10 for middle-income households and around 18 for high-income households. Since the actual computed elasticity varies with the interest rate, that implies middle-income households have an elasticity of 0.2 at a 2% nominal interest rates, while high-income households have an elasticity of 0.35. At very high nominal rates—around 10%—these jump to 0.9 and 1.63, respectively. Given the sample average for the nominal interest rate, the high-income households have an interest elasticity of 0.5, which corresponds exactly with Baumol-Tobin. Hence, despite the difference in specification closely mirrors Lucas's (2000) estimate as well as Benati and Nicolini (2021). Interestingly, those estimates are made entirely on the basis of aggregate data, while my estimates rely only on a small slice of households which nevertheless hold a substantial fraction of cash in the economy.

As a robustness check, I perform the same regressions for the entire SCF sample and the aggregate economy. The interest elasticity is consistently similar between the two regardless of specification, while the consumption elasticity tends to be higher for the aggregate economy. However, the interest elasticity estimates are quite similar to the estimate for the middle income group. Overall, the similarity is surprising because firms also hold a substantial amount of money, yet they are not at all reflected in the SCF sample.

Finally, note that the consumption elasticity is significantly different from one for practically all specifications, so I am forced to rely on money demand specifications which depend non-linearly on consumption.

5.2.1 Summing up

It is most convenient to write the money demand functions as a ratio of real balances to consumption. After exponentiating the preferred specifications, that simply corresponds to raising consumption to its estimated elasticity minus one. Hence the results give rise to the following money-consumption ratio demand functions:

$$m_L(i^d) = 0.0707 \tag{16}$$

$$m_M(i^d) = \exp\{-1.98\}c_M^{0.42}\left(1+i^d\right)^{-9.95} \tag{17}$$

$$m_H(i^d) = \exp\{0.33\}c_H^{-0.54}\left(1+i^d\right)^{-18.29}. \tag{18}$$

I use these parameterizations for the remainder of the paper.

6 Welfare Estimates

In a heterogeneous agent model, redistribution takes center stage. Using the estimated money demand functions and the Friedman rule as a benchmark, I estimate the welfare cost of inflation for different income groups according to the framework developed in Section 3. At each period t, before accounting for redistribution, the welfare cost of moving from i_1^d to

i_2^d for each income group (using the estimates from the empirical section) is given by

$$w_L(i^d) = \Upsilon_L(i_2^d - i_1^d) - T(z) \tag{19}$$

$$w_M(i^d) = \frac{\Upsilon_M \bar{c}_M^{\Gamma^M - 1}}{1 - \sigma_M} \left((1 + i_2^d)^{1-\sigma_M} - (i_1^d)^{1-\sigma_M}\right) - T(z) \tag{20}$$

$$w_H(i^d) = \frac{\Upsilon_H \bar{c}_H^{\Gamma^H - 1}}{1 - \sigma_H} \left((1 + i_2^d)^{1-\sigma_H} - (1 + i_1^d)^{1-\sigma_H}\right) - T(z), \tag{21}$$

where $T(z)$ denotes an undefined redistribution policy.

Seigniorage Accounting

At each point in time, the welfare cost of moving from i_1^d to i_2^d is the consumer surplus less the seigniorage revenue returned lump-sum to households. Because there is almost surely heterogeneity in how seigniorage revenue is used by the government, it is necessary to do some accounting. Under the stationary distribution of constant aggregate and group-level consumption, total seigniorage revenue from each group is given by

$$S^j = \left(i_2^d - i_1^d\right) \left[\Upsilon_j (\bar{c}_j)^{\Gamma_j} \left(1 + i_2^d\right)^{-\sigma_j}\right],$$

where j indexes household type in low, middle, and high. Note that this is not in terms of the money-consumption ratio. Now, the total seigniorage revenue the government collects is

$$S = \left(i_2^d - i_1^d\right) \left[\sum_{j \in \{L,M,H\}} \Upsilon_j (\bar{c}_j)^{\Gamma_j} \left(1 + i_2^d\right)^{-\sigma_j}\right].$$

Supposing that each household receives a share of redistribution equivalent to ω_j, the total seigniorage revenue received by each household group j is given by $\omega_j S$. Then the welfare cost at each point in time of moving from i_1^d to i_2^d, written as a share of consumption and conditional on being in state j, is given by

$$w_j(i_2^d) = \frac{\Upsilon_j \left(\bar{c}_j\right)^{\Gamma_j - 1}}{1-\sigma_j} \left((1+i_2^d)^{1-\sigma_H} - (1+i_1^d)^{1-\sigma_j}\right) - \omega_j S. \tag{22}$$

Conditional on being in state j at time $t = 0$ under the stationary distribution, the dynamic welfare cost is a discounted expected sequence of such costs as described earlier.

An Empirical $T(z)$

Finally, I require a policy rule $T(z)$ that takes some quantity of seigniorage revenue and distributes shares ω_j to each income group. Note that these do not necessarily need to add up to one; some may be wasted. In the case where government spending enters the utility function, then it is possible for the utility-relevant redistribution to be nonlinearly increasing. However, in keeping with the literature and for the sake of simplicity, I assume that government spending does not enter the utility function and simply enters linearly back into the consumption-equivalent welfare cost.

As my benchmark case, I assume that half of seigniorage revenue is wasted, 25% of it goes to low-income households, 22.5% to middle-income households, and 2.5% to high-income households. This distribution is calibrated from Congressional Budget Office (2018). I also present results without any redistribution.

Numerical Estimates

Before computing the correct welfare object, namely the expected present value of future welfare losses scaled by permanent consumption, I first compute the static welfare cost for each group assuming the same redistribution weights as above, which is akin to assuming that each income state is absorbing and the transition matrix is diagonal. That is, I simply compute the area under the estimated demand curve net of redistribution without attention to dynamic concerns. I present welfare costs for each group in Figure 3.

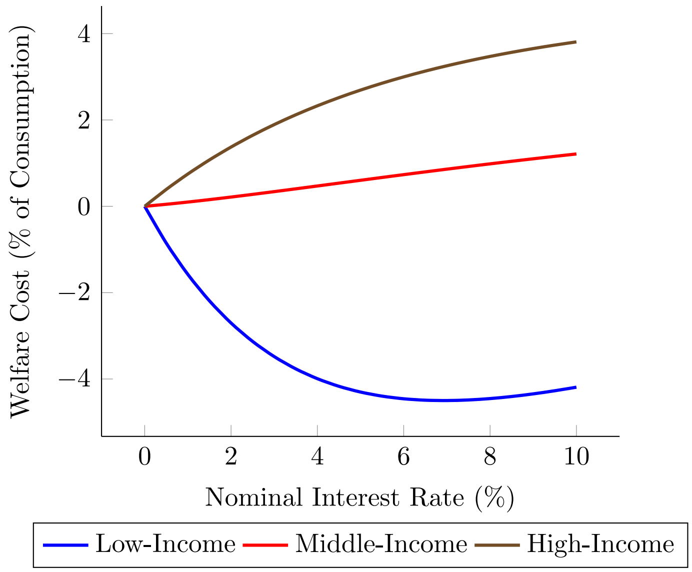


Figure 3: Static welfare cost of inflation for varying nominal interest rates.

In Figure 3, costs for each group can be interpreted as the welfare loss if there were zero economic mobility. A 5% nominal interest rate costs middle-income households about 0.6% of their annual consumption and high-income households around 2.7% of annual consumption, while this policy is worth 4.3% of annual consumption to low-income households. Hence, low-income households benefit enormously from positive steady-state inflation due to positive redistribution, most of which is taken from high-income households. A result like this—namely that positive inflation may be good for redistributive reasons—reflects arguments that inflation benefits low-income households more than it hurts high-income households, with the normative implication that a positive inflation tax is beneficial on net (Yang 2022).

Next, I compute dynamic welfare costs taking account of economic mobility, namely that households may switch between states. Calibrating permanent consumption under the stationary distribution from the CEX and setting the discount factor $\beta = 1/1.04$, I compute the classical cost of inflation as the expected present value of future welfare loss scaled by permanent consumption. Results for various nominal interest rates are in Figure 4. Here, a 5% nominal interest rate costs low-income households 0.54% of consumption, middle-income households 0.66% of consumption, and high-income households 0.76% of consumption. In particular, note that while welfare cost is increasing in income, *all* types of households are hurt significantly by inflation. This stands in contrast to implied distributional costs from Erosa and Ventura (2002) and Akyol (2004). The welfare cost estimates for a 5% nominal rate are well within the range of estimates for aggregate welfare loss in the state-of-the-art literature (Benati and Nicolini 2021). Moreover, comparing Figures 3 and 4, middle-income households face essentially the same welfare cost curve statically and dynamically. This is sensible, but indicates that most of the gains from correctly accounting for dynamics affect the tails of the distribution.

The large difference in the distribution of costs between Figures 3 and 4 prompts some consideration of what the correct welfare metric should be. Clearly, since the long-run nominal rate is a function of policy, this is an important normative question. I argue that the dynamic approach is correct. An alternative interpretation of Figure 3 is that it is a static intraperiod cost for each income group, which is the opposite of the relevant object, namely the long-run cost. This argument resolves the tension between short-run studies which argue for a positive inflation tax for the purposes of short-run redistribution (e.g., Akyol (2004)) and the present study. In contrast, the dynamic welfare metric *is* a long-run cost because it reflects not only the household's state today, but the expected sequence of all such future states. Finally, the dynamic metric aligns with the theoretical result from da Costa and Werning (2008), *i.e.*, that the Friedman rule is optimal in a broad class of monetary economies.

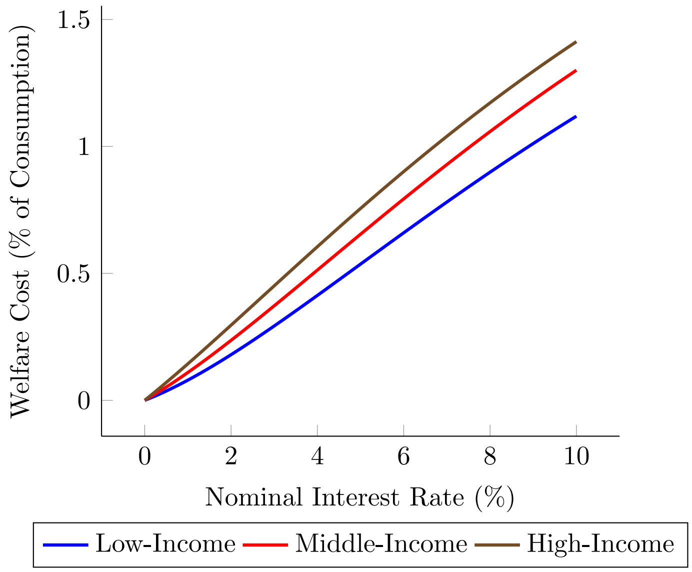


Figure 4: Dynamic welfare cost of inflation for varying nominal interest rates.

Of course, one may make different assumptions about the redistribution function $T(z)$. An interesting case may be without *any* redistribution. That is, all seigniorage revenue is wasted. In Figure 8, I present welfare cost estimates for each income group for varying nominal interest rates. The low-income group's welfare cost more than doubles, the middle-income group's welfare cost doubles, and the high-income group's less than doubles, but the ranking of costs remains constant. The costs squeeze together because redistribution is statically beneficial to low-income groups, but high-income groups have little chance of becoming low-income. That the variance of costs between groups declines significantly is indicative of the importance of assumptions about redistribution.

Because my specification is somewhat nonstandard to account for heterogeneity, I compute welfare loss for the SCF as a whole and the aggregate economy. I use the standard area under the demand curve formula net of seigniorage. Results are in Figure 7. A 5% nominal interest rate has a welfare cost of 0.44% of consumption for the money demand curve generated by the SCF alone and 0.59% for the aggregate economy. Both of these are within the aggregate estimates presented by Benati and Nicolini (2021) and in fact are bounded below by the semi-log and above by the log-log specifications. Moreover, the fact that firms hold money and these are largely held by high-income households points to an underestimate of their welfare loss.

7 Concluding Remarks

Household heterogeneity in money demand generates differing classical welfare costs from inflation. Given that the primary expositor of modern monetarist analysis looks askance at considerations of inequality in economics, it is incumbent on me to present some defense of a sort of "distributional monetarism."[13]

Distributional monetarism has two meanings. First, it means considering, in a positive sense, the composition of the aggregate money demand curve. As I show in Section 3, the aggregate money demand curve—possibly the most important element of monetarism—is itself influenced by household-level factors. As these change, so does the aggregate money demand curve. Second, distributional monetarism means taking proper account of optimal monetary policy in the context of heterogeneity. Even in this setting, the Friedman rule holds and so do the standard monetarist prescriptions. Yet the distribution of welfare costs has important implications for how policymakers form long-term policy. If my results were different, namely that the welfare cost gap is vast between low-income and high-income households, then one could reasonably advocate for a policy based on redistribution. But given the apparent similarity in a dynamic sense, the prescription is the same as if we only analyzed the aggregate economy.

More generally, the notion that long-run distributional welfare costs should be evaluated dynamically rather than statically has broader implications for wealth taxation. Inflation in the long-run is merely a tax on real balances, *i.e.,* a wealth tax on a particular type of wealth. With economic mobility, the prospect of a wealth tax is perhaps positive in a static

13. Robert Lucas is well-known for the following line: "Of the tendencies that are harmful to sound economics, the most seductive, and in my opinion the most poisonous, is to focus on questions of distribution" (Lucas 2004).

sense from the perspective of a low-income household, but is likely negative from a dynamic perspective. This, perhaps, should be the starting point for future research on the subject.

References

Akyol, Ahmet. 2004. "Optimal monetary policy in an economy with incomplete markets and idiosyncratic risk." *Journal of Monetary Economics* 51, no. 6 (September): 1245–1269. ISSN: 03043932. https://doi.org/10.1016/j.jmoneco.2003.09.005.

Allais, Olivier, Yann Algan, Edouard Challe, and Xavier Ragot. 2020. "The Welfare Cost of Inflation Risk under Imperfect Insurance." *Annals of Economics and Statistics,* no. 138, 1. ISSN: 21154430. https://doi.org/10.15609/annaeconstat2009.138.0001.

Alvarez, Fernando, Francesco Lippi, and Roberto Robatto. 2019. "Cost of inflation in inventory theoretical models." *Review of Economic Dynamics* 32 (April): 206–226. ISSN: 10942025. https://doi.org/10.1016/j.red.2018.11.001.

Alvarez, Fernandon, and Francesco Lippi. 2009. "Financial Innovation and the Transactions Demand for Cash." *Econometrica* 77 (2): 363–402. ISSN: 0012-9682. https://doi.org/10.3982/ECTA7451.

Ascari, Guido, Louis Phaneuf, and Eric R. Sims. 2018. "On the welfare and cyclical implications of moderate trend inflation." *Journal of Monetary Economics* 99 (November): 56–71. ISSN: 03043932. https://doi.org/10.1016/j.jmoneco.2018.06.001.

Bae, Youngsoo, and Robert M. de Jong. 2007. "Money demand function estimation by nonlinear cointegration." *Journal of Applied Econometrics* 22, no. 4 (June): 767–793. ISSN: 08837252. https://doi.org/10.1002/jae.915.

Bailey, Martin J. 1956. "The Welfare Cost of Inflationary Finance." *Journal of Political Economy* 64, no. 2 (April): 93–110. ISSN: 0022-3808. https://doi.org/10.1086/257766.

Baqaee, David, Ariel Burstein, and Yasutaka Koike-Mori. 2022. *A New Method for Measuring Welfare with Income Effects using Cross-Sectional Data.* Technical report. Cambridge, MA: National Bureau of Economic Research, October. https://doi.org/10.3386/w30549.

Benati, Luca, Robert E. Lucas, Juan Pablo Nicolini, and Warren Weber. 2021. "International evidence on long-run money demand." *Journal of Monetary Economics* 117 (January): 43–63. ISSN: 03043932. https://doi.org/10.1016/j.jmoneco.2020.07.003.

Benati, Luca, and Juan-Pablo Nicolini. 2021. "The Welfare Costs of Inflation."

Bhandari, Anmol, Serdar Birinci, Ellen R. McGrattan, and Kurt See. 2020. "What Do Survey Data Tell Us about US Businesses?" *American Economic Review: Insights* 2, no. 4 (December): 443–458. ISSN: 2640-205X. https://doi.org/10.1257/aeri.20190304.

Bils, Mark, and Yongsung Chang. 2003. "Welfare costs of sticky wages when effort can respond." *Journal of Monetary Economics* 50, no. 2 (March): 311–330. ISSN: 03043932. https://doi.org/10.1016/S0304-3932(03)00004-7.

Bischoff, Charles W., and Halefom Belay. 2001. "The Problem of Identification of the Money Demand Function." *Journal of Money, Credit and Banking* 33, no. 2 (May): 205–215.

Braitsch, Hana M., and James Mitchell. 2022. "A New Measure of Consumers' (In)Attention to Inflation." *Economic Commentary (Federal Reserve Bank of Cleveland),* nos. 2022-14 (October). ISSN: 2163-3738. https://doi.org/10.26509/frbc-ec-202214.

Budd, Edward C., and David F. Seiders. 1971. "The Impact of Inflation on the Distribution of Income and Wealth." *American Economic Review* 61 (2): 128–138.

Cacciatore, Matteo, Fabio Ghironi, and Viktors Stebunovs. 2015. "The domestic and international effects of interstate U.S. banking." *Journal of International Economics* 95, no. 2 (March): 171–187. ISSN: 00221996. https://doi.org/10.1016/j.jinteco.2014.12.001.

Campbell, Jeffrey R., and Zvi Hercowitz. 2019. "Liquidity Constraints of the Middle Class." *American Economic Journal: Economic Policy* 11, no. 3 (August): 130–155. ISSN: 1945-7731. https://doi.org/10.1257/pol.20180070.

Cao, Shutao, Césaire A. Meh, José-Víctor Ríos-Rull, and Yaz Terajima. 2021. "The welfare cost of inflation revisited: The role of financial innovation and household heterogeneity." *Journal of Monetary Economics* 118 (March): 366–380. ISSN: 03043932. https://doi.org/10.1016/j.jmoneco.2020.11.004.

Cirelli, Fernando. 2022. "Bank-Dependent Households and the Unequal Costs of Inflation."

Congressional Budget Office. 2018. *The Distribution of Household Income, 2018.* Technical report. Congressional Budget Office.

Costa, Carlos E. da, and Iván Werning. 2008. "On the Optimality of the Friedman Rule with Heterogeneous Agents and Nonlinear Income Taxation." *Journal of Political Economy* 116, no. 1 (February): 82–112. ISSN: 0022-3808. https://doi.org/10.1086/529397.

CRS. 2020. *Real Wage Trends, 1979 to 2019.* Technical report. Congressional Research Service, December.

Doepke, Matthias, and Martin Schneider. 2006. "Inflation and the Redistribution of Nominal Wealth." *Journal of Political Economy* 114, no. 6 (December): 1069–1097. ISSN: 0022-3808. https://doi.org/10.1086/508379.

Erosa, Andrés, and Gustavo Ventura. 2002. "On inflation as a regressive consumption tax." *Journal of Monetary Economics* 49, no. 4 (May): 761–795. ISSN: 03043932. https://doi.org/10.1016/S0304-3932(02)00115-0.

Friedman, Milton. 1969. "The Optimum Quantity of Money." In *The Optimum Quantity of Money and Other Essays.* Chicago: Aldine Publishing Company.

Gao, Han, Mariano Kulish, and Juan Pablo Nicolini. 2020. "Two Illustrations of the Quantity Theory of Money Reloaded."

Ireland, Peter N. 2009. "On the Welfare Cost of Inflation and the Recent Behavior of Money Demand." *American Economic Review* 99, no. 3 (May): 1040–1052. ISSN: 0002-8282. https://doi.org/10.1257/aer.99.3.1040.

Jaravel, Xavier. 2019. "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector*." *The Quarterly Journal of Economics* 134, no. 2 (May): 715–783. ISSN: 0033-5533. https://doi.org/10.1093/qje/qjy031.

Kaplan, Greg, and Gianluca Violante. 2014. "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica* 82 (4): 1199–1239. ISSN: 0012-9682. https://doi.org/10.3982/ECTA10528.

Kurlat, Pablo. 2019. "Deposit spreads and the welfare cost of inflation." *Journal of Monetary Economics* 106 (October): 78–93. ISSN: 03043932. https://doi.org/10.1016/j.jmoneco.2019.07.006.

Lucas, Robert E. 1988. "Money demand in the United States: A quantitative review." *Carnegie-Rochester Conference Series on Public Policy* 29 (January): 137–167. ISSN: 01672231. https://doi.org/10.1016/0167-2231(88)90009-7.

———. 2000. "Inflation and Welfare." *Econometrica* 68, no. 2 (March): 247–274. ISSN: 0012-9682. https://doi.org/10.1111/1468-0262.00109.

Lucas, Robert E. 2004. *The Industrial Revolution: Past and Future.* Technical report. Federal Reserve Bank of Minneapolis, May.

Lucas, Robert E., and Juan Pablo Nicolini. 2015. "On the stability of money demand." *Journal of Monetary Economics* 73 (July): 48–65. ISSN: 03043932. https://doi.org/10.1016/j.jmoneco.2015.03.005.

Lucas, Robert E., and Nancy L. Stokey. 1987. "Money and Interest in a Cash-in-Advance Economy." *Econometrica* 55, no. 3 (May): 491. ISSN: 00129682. https://doi.org/10.2307/1913597.

Mankiw, N. Gregory, and Lawrence H. Summers. 1986. "Money Demand and the Effects of Fiscal Policies." *Journal of Money, Credit and Banking* 18, no. 4 (November): 415. ISSN: 00222879. https://doi.org/10.2307/1992462.

McGrattan, Ellen R. 2020. "Intangible capital and measured productivity." *Review of Economic Dynamics* 37 (August): S147–S166. ISSN: 10942025. https://doi.org/10.1016/j.red.2020.06.007. https://linkinghub.elsevier.com/retrieve/pii/S1094202520300466.

Papademos, Lucas, and Franco Modigliani. 1990. "Chapter 10 The supply of money and the control of nominal income." In *Handbook of Monetary Economics,* 1:399–494. https://doi.org/10.1016/S1573-4498(05)80013-X.

Sahm, Claudia R, Matthew D Shapiro, and Joel Slemrod. 2012. "Check in the Mail or More in the Paycheck: Does the Effectiveness of Fiscal Stimulus Depend on How It Is Delivered?" *American Economic Journal: Economic Policy* 4, no. 3 (August): 216–250. ISSN: 1945-7731. https://doi.org/10.1257/pol.4.3.216.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50, no. 3 (April): 665–690. ISSN: 03043932. https://doi.org/10.1016/S0304-3932(03)00029-1.

Stock, James H., and Mark W. Watson. 1993. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica* 61:783–820.

———. 2016. "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," 415–525. https://doi.org/10.1016/bs.hesmac.2016.04.002.

Teles, Pedro, and Ruilin Zhou. 2005. "A stable money demand: Looking for the right monetary aggregate." *Economic Perspectives* 29 (Q1): 50–63.

Wen, Yi. 2015. "Money, liquidity and welfare." *European Economic Review* 76 (May): 1–24. ISSN: 00142921. https://doi.org/10.1016/j.euroecorev.2015.01.012.

Yang, Yuchen. 2022. "Redistributive Inflation and Optimal Monetary Policy."

A Shopping Time Model Derivation

All households j choose $c_t^j, n_t^j, z_t^j, M_t^j, B_t^j, W_{t+1}^j$ to maximize utility

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t U(c_t^j) \tag{A.1}$$

subject to the time endowment

$$1 = z_t^j + n_t^j, \tag{A.2}$$

the transactions technology

$$\left(c_t^j\right)^{\alpha^j} = \zeta^j\left(\frac{B_t^j}{p_t} + A_t^j z_t^j\right)^{\gamma^j}\left(\frac{M_t^j}{p_t}\right)^{\nu^j}, \tag{A.3}$$

the intra-period portfolio allocation constraint

$$B_t^j + M_t^j \leq W_t^j \tag{A.4}$$

and the evolution of wealth

$$W_{t+1}^j \leq M_t^j + B_t^j\left(1 + \frac{i_t}{1+\phi^j}\right) + p_t A_t^j n_t^j + T_t^j - p_t c_t^j. \tag{A.5}$$

Substituting $1 - z_t^j$ for n_t^j and letting μ_t^j, δ_t^j, and λ_t^j denote the Lagrange multipliers on constraints (A.3), (A.4), and (A.5), respectively, first-order conditions are given by

$$\left[c_t^j\right]: \quad \beta^t U'(c_t^j) = p_t \lambda_t^j + \alpha^j (c_t^j)^{\alpha^j - 1} \delta_t^j \tag{A.6}$$

$$\left[z_t^j\right]: \quad \gamma^j A_t^j \zeta^j \left(\frac{B_t^j}{p_t} + A_t^j z_t^j\right)^{\gamma^j - 1} \left(\frac{M_t^j}{p_t}\right)^{\nu^j} \mu_t^j = p_t A_t^j \lambda_t^j \tag{A.7}$$

$$\left[B_t^j\right]: \quad \delta_t^j = \lambda_t^j \left(1 + \frac{i_t}{\phi^j}\right) + \frac{\gamma^j \zeta^j}{p_t} \left(\frac{B_t^j}{p_t} + A_t^j z_t^j\right)^{\gamma^j - 1} \left(\frac{M_t^j}{p_t}\right)^{\nu^j} \mu_t^j \tag{A.8}$$

$$\left[M_t^j\right]: \quad \delta_t^j = \lambda_t^j + \frac{\nu^j \zeta^j}{p_t} \left(\frac{B_t^j}{p_t} + A_t^j z_t^j\right)^{\gamma^j} \left(\frac{M_t^j}{p_t}\right)^{\nu^j - 1} \mu_t^j \tag{A.9}$$

$$\left[W_{t+1}^j\right]: \quad \lambda_t^j = \mathbb{E}_t \delta_{t+1}^j \tag{A.10}$$

Using the first-order conditions, the necessary equilibrium condition relating the opportunity cost of holding money and consumption can be written as

$$\frac{\nu^j}{\gamma^j} \left(\frac{M_t^j}{p_t}\right)^{-1} \left(\frac{B_t^j}{p_t} + A_t^j z_t^j\right) = 1 + \frac{i_t}{1 + \phi^j}. \tag{A.11}$$

Then, using the fact that in equilibrium consumption is equal to the transactions technology so that

$$\frac{B_t^j}{p_t} + A_t^j z_t^j = \left(\frac{(c_t^j)^{\alpha^j}}{\zeta^j}\right)^{\frac{1}{\gamma^j}} \left(\frac{M_t^j}{p_t}\right)^{-\frac{\nu^j}{\gamma^j}},$$

the required equilibrium relationship between real balances, consumption, and the opportunity cost of holding money is given by

$$\frac{M_t^j}{p_t} = \Upsilon^j (c_t^j)^{\Gamma^j} \left(1 + \frac{i_t}{1 + \phi^j}\right)^{-\sigma^j}, \tag{A.12}$$

where

$$\Upsilon^j \equiv \left[\frac{1}{\zeta^j} \left(\frac{\nu^j}{\gamma^j}\right)^{\gamma^j}\right]^{\frac{1}{\gamma^j + \nu^j}}$$

is a constant determining average money holdings,

$$\Gamma^j \equiv \frac{\alpha^j}{\gamma^j + \nu^j}$$

is the consumption elasticity of money demand, and

$$\sigma^j \equiv \frac{\gamma^j}{\nu^j + \gamma^j}$$

is a composite parameter critical for determining the interest elasticity of money demand.

A.1 Proof of Proposition 1

Start with a representative agent of the ψ unconstrained households. With $\phi^j = 0$, these households submit money demand

$$\frac{M_t^j}{p_t}(i_t) = \Upsilon^j (c_t^j)^{\Gamma^j} (1 + i_t)^{-\sigma^j}, \tag{A.13}$$

Now consider a representative agent of the $1 - \psi$ constrained households for which $\phi^j \to \infty$. Applying equation (A.12), we know

$$\lim_{\phi^j \to \infty} \frac{M_t^j}{p_t} = \Upsilon^j (c_t^j)^{\Gamma^j} \left(1 + \frac{i_t}{1 + \phi^j}\right)^{-\sigma^j} = \Upsilon^j (c_t^j)^{\Gamma^j}.$$

That is, money demand for constrained households is given by a constant

$$\frac{M_t^j}{p_t} = \Upsilon^j (c_t^j)^{\Gamma^j} \tag{A.14}$$

and hence is invariant to changes in the level of the nominal interest rate.

A.2 Proof of Proposition 2

Note that the private demand for money is simply the horizontal summation of individual demands. In the simplest case, suppose there are only two states, denoted 1 and 2, one of which corresponds to being in the high state and having access to $\phi^1 = 1$ and the other corresponds to being in the low state with $\phi^2 \to \infty$. Suppose also that the consumption level and elasticity are the same across states and the constant Υ^j are the same across groups, i.e.,

$$(c_t^1)^{\alpha^1} = (c_t^2)^{\alpha^2} \equiv \bar{C} \quad \text{and} \quad \Upsilon^1 = \Upsilon^2 \equiv \bar{\Upsilon}$$

and letting $\bar{B} \equiv \bar{C}\bar{\Upsilon}$. The purpose of this assumption is to make clear that I am principally analyzing changes in money demand emerging from heterogeneous elasticity rather than heterogeneous levels. For what follows, assume the parameter ν_t^1 is time-invariant and drop the group superscript. Then market demand is given by a weighted average of the unconstrained and the credit-constrained money demands.

$$\begin{aligned} \frac{M^{\text{market}}}{p_t} &= \psi\left[\bar{B}(1+i_t)^{-\sigma}\right] + (1-\psi)\bar{B} \\ &= \bar{B}\left(\psi(1+i_t)^{-\sigma} + 1 - \psi\right), \end{aligned} \tag{A.15}$$

Under this formulation, the interest elasticity of aggregate money demand is given by[14]

$$\varepsilon = \frac{\sigma\psi i_t(1+i_t)^{-\sigma-1}}{\psi(1+i_t)^{-\sigma} + 1 - \psi} \tag{A.16}$$

Taking a partial derivative of equation A.16 with respect to ψ, the share of unconstrained households, gives

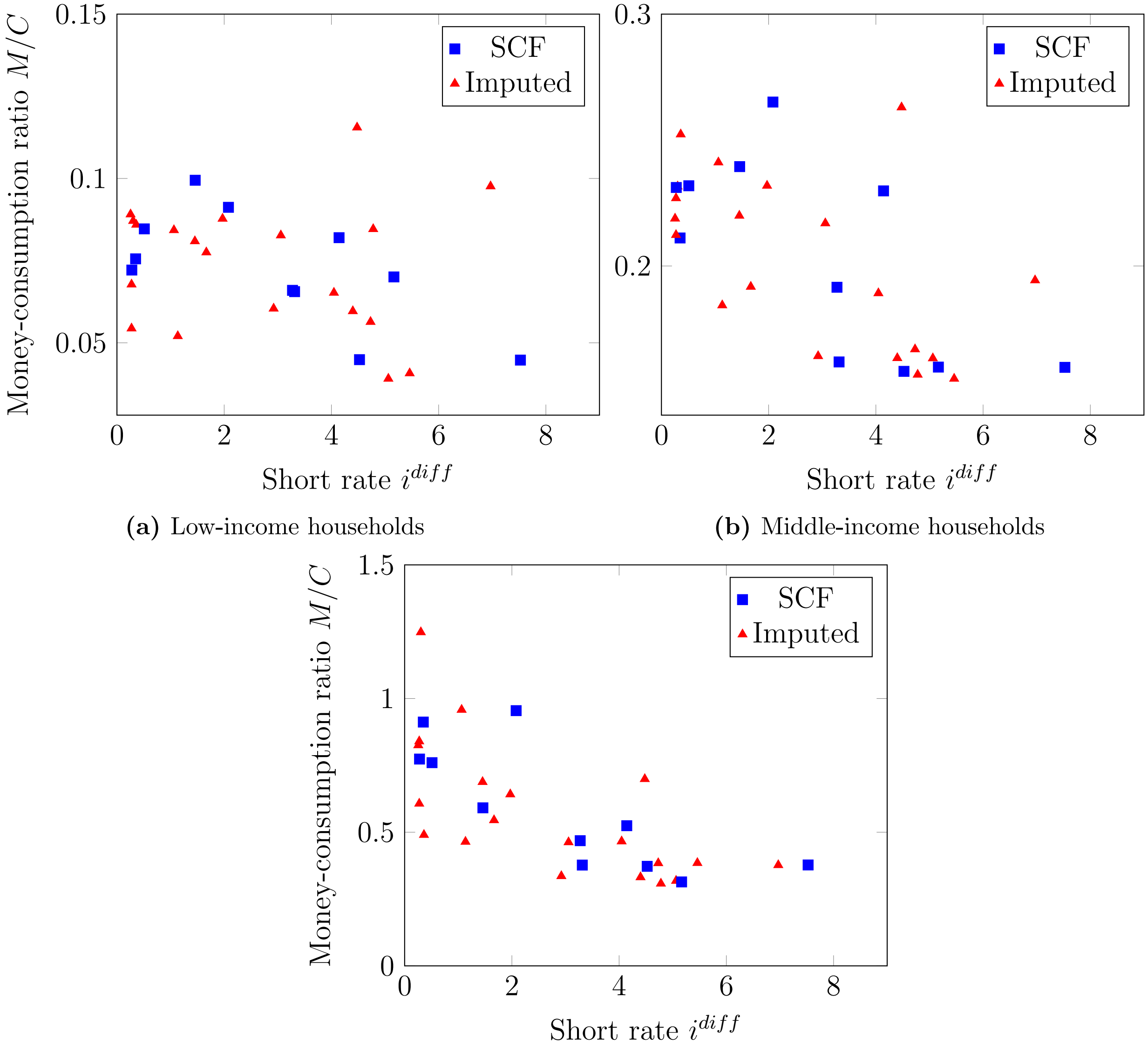
$$\frac{\partial\varepsilon}{\partial\psi} = \frac{\sigma i_t(1+i_t)^{-\sigma-1}}{\left[\psi\left((1+i_t)^{\sigma} - 1\right) - (1+i_t)^{\sigma}\right]^2} > 0. \tag{A.17}$$

14. Elasticity is computed as $\varepsilon = -\frac{Px'(P)}{x(P)}$, where P is a price and $x(P)$ is a demand function.

That is, the interest elasticity is increasing in the share of interest-elastic households. If we relaxed the assumptions of having two discrete states and having equivalent constants, the results would still go through but would not be as clear.

B Money Demand Plots

In each of the following figures, I plot the money-consumption ratio against the short rate. The "Whole Economy" are the ratio of NewM1 from Gao, Kulish, and Nicolini (2020) to nominal aggregate consumption from Line 2 of NIPA Table 1.5.

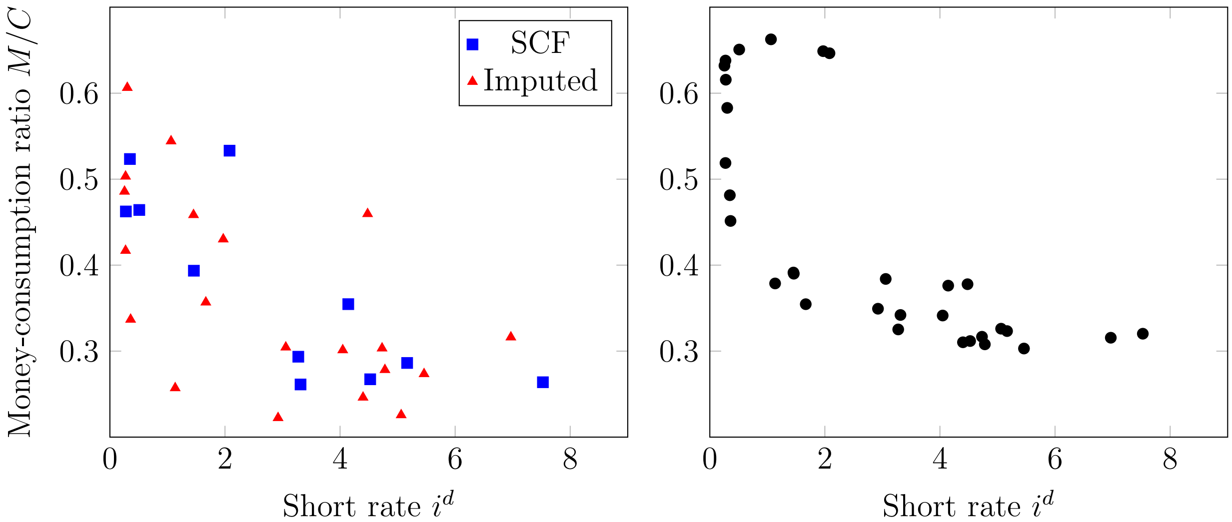


(a) Low-income households

(b) Middle-income households

(c) High-income households

Figure 5: Money-consumption ratios for each income category plotted against the nominal interest rate. Imputed values are denoted with a red triangle, while realizations are blue squares.



(a) Households

(b) Whole Economy

Figure 6: Household money-consumption ratio (left) and economy-wide money-consumption ratio plotted against the nominal short rate.

C Econometric Tests

Unit Root Tests

	Low-Income	Middle-Income	High-Income	All Households	Economy
Consumption	-1.700	0.242	-1.516	-1.403	-1.853
Real Balances	-1.778	-1.553	-0.573	-1.472	-0.392
M/C	-1.766	-0.550	-0.520	-0.381	-0.670
Interest Rate					-2.913

Table 3: Results of augmented Dickey-Fuller test with drift against null hypothesis of a unit root. The critical value (at the 5% level) is -3.00.

	Low-Income	Middle-Income	High-Income	All Households	Economy
Consumption	-5.279	-2.464	-1.042	-2.047	0.482
Real Balances	-7.638	-1.718	-6.443	-5.941	-5.941
M/C	-5.632	-1.399	-1.828	-2.091	0.123
Interest Rate					-2.428

Table 4: Results of Phillips-Perron test with drift against null hypothesis of a unit root. The critical value (at the 5% level) is -3.00.

Cointegration Tests

	Low-Income	Middle-Income	High-Income	All Households	Economy	Crit. Value
S1	25.404	45.278	44.846	51.251	52.983	34.91
S2	19.138	23.907	39.492	39.542	26.368	19.96

Table 5: Results of Johansen test for cointegration. S1 corresponds to a test between the log-level of real balances, the log-level of consumption, the log-level of the gross interest rate, and a constant. S2 corresponds to a cointegration test between the log of the money-consumption ratio and the log-level of the gross interest rate with a constant.

	Low-Income	Middle-Income	High-Income	All Households	Economy	Crit. Value
S1	32.889	13.415	11.355	16.908	2.553	40.525
S2	34.327	12.567	21.080	19.133	0.419	33.713

Table 6: Results of Phillips-Ouliaris $\hat{P}_u$ test for cointegration. S1 corresponds to a test between the log-level of real balances, the log-level of consumption, the log-level of the gross interest rate, and a constant. S2 corresponds to a cointegration test between the log of the money-consumption ratio and the log-level of the gross interest rate with a constant.

C.1 Parameter Estimates

Dep. Var	$\hat{\alpha}$	s.e.($\hat{\alpha}$)	$\hat{\beta}_c$	s.e.($\hat{\beta}_c$)	$\hat{\beta}_i$	s.e.($\hat{\beta}_i$)
$\Delta \frac{M}{p}$	0.013	0.021	-0.65	0.55	-0.43	3.68
$\Delta \frac{M/p}{C}$	0.024	0.020			0.29	5.38
$\frac{M/p}{C}$	-2.65	0.06				

Table 7: OLS Estimates for Low-Income Households. p corresponds to the leads and lags of consumption and the interest rate included. $\hat{\beta}_c$ is the coefficient on consumption, while $\hat{\beta}_i$ is the coefficient on the nominal interest rate. The dependent variable column corresponds to whether the LHS variable is the log money-consumption ratio or the log-level of real balances. Note that the final regression corresponds to simply regressing the money-consumption ratio on a constant. Standard errors are robust to heteroskedasticity and autocorrelation.

Regression	Dep. Var	p	$\hat{\alpha}$	s.e.($\hat{\alpha}$)	$\hat{\beta}_c$	s.e.($\hat{\beta}_c$)	$\hat{\beta}_i$	s.e.($\hat{\beta}_i$)
Static OLS	$\frac{M}{p}$	0	-0.80	0.40	-0.48	0.28	-3.47	1.46
	$\frac{M/p}{C}$	0	-1.46	0.04			-5.42	1.31
DOLS	$\frac{M}{p}$	1	-1.98	0.42	1.42	0.32	-9.95	2.33
	$\frac{M/p}{C}$	1	-1.40	0.05			-7.93	1.72
	$\frac{M}{p}$	2	-2.32	0.13	1.71	0.09	-12.51	0.67
	$\frac{M/p}{C}$	2	-1.38	0.05			-8.77	1.64

Table 8: OLS Estimates for Middle-Income Households. p corresponds to the leads and lags of consumption and the interest rate included. $\hat{\beta}_c$ is the coefficient on consumption, while $\hat{\beta}_i$ is the coefficient on the nominal interest rate. The dependent variable column corresponds to whether the LHS variable is the log money-consumption ratio or the log-level of real balances.

Regression	Dep. Var	p	$\hat{\alpha}$	s.e.($\hat{\alpha}$)	$\hat{\beta}_c$	s.e.($\hat{\beta}_c$)	$\hat{\beta}_i$	s.e.($\hat{\beta}_i$)
Static OLS	$\frac{M}{p}$	0	0.74	0.16	-0.22	0.17	-10.21	1.35
	$\frac{M/p}{C}$	0	-0.24	0.08			-14.18	1.97
DOLS	$\frac{M}{p}$	1	0.33	0.26	0.46	0.35	-18.29	2.01
	$\frac{M/p}{C}$	1	-0.08	0.02			-20.79	0.68
	$\frac{M}{p}$	2	-0.32	0.28	1.31	0.36	-22.88	1.97
	$\frac{M/p}{C}$	2	-0.05	0.03			-21.91	1.07

Table 9: OLS Estimates for High-Income Households. p corresponds to the leads and lags of consumption and the interest rate included. $\hat{\beta}_c$ is the coefficient on consumption, while $\hat{\beta}_i$ is the coefficient on the nominal interest rate. The dependent variable column corresponds to whether the LHS variable is the log money-consumption ratio or the log-level of real balances.

Regression	Dep. Var	p	$\hat{\alpha}$	s.e.($\hat{\alpha}$)	$\hat{\beta}_c$	s.e.($\hat{\beta}_c$)	$\hat{\beta}_i$	s.e.($\hat{\beta}_i$)
Static OLS	$\frac{M}{p}$	0	21.77	3.22	0.01	0.14	-6.64	1.01
	$\frac{M/p}{C}$	0	-0.78	0.06			-9.55	1.53
DOLS	$\frac{M}{p}$	1	14.14	5.58	0.35	0.24	-10.85	1.74
	$\frac{M}{p}$	2	6.54	5.20	0.68	0.23	-13.96	1.63
	$\frac{M/p}{C}$	2	-0.61	0.029			-15.85	0.97

Table 10: OLS Estimates for all households. p corresponds to the leads and lags of consumption and the interest rate included. $\hat{\beta}_c$ is the coefficient on consumption, while $\hat{\beta}_i$ is the coefficient on the nominal interest rate. The dependent variable column corresponds to whether the LHS variable is the log money-consumption ratio or the log-level of real balances.

Regression	Dep. Var	p	$\hat{\alpha}$	s.e.($\hat{\alpha}$)	$\hat{\beta}_c$	s.e.($\hat{\beta}_c$)	$\hat{\beta}_i$	s.e.($\hat{\beta}_i$)
Static OLS	$\frac{M}{p}$	0	-0.08	11.87	0.95	1.34	-3.24	10.28
	$\frac{M/p}{C}$	0	-0.59	0.08			-10.75	7.87
DOLS	$\frac{M}{p}$	1	2.38	2.07	0.69	0.21	-10.33	5.56
	$\frac{M/p}{C}$	1	-0.49	0.06			-14.91	2.06
	$\frac{M}{p}$	2	4.19	1.80	0.51	0.19	-14.56	5.13
	$\frac{M/p}{C}$	2	-0.48	0.04			-15.42	1.45

Table 11: OLS Estimates for the whole economy. p corresponds to the leads and lags of consumption and the interest rate included. $\hat{\beta}_c$ is the coefficient on consumption, while $\hat{\beta}_i$ is the coefficient on the nominal interest rate. The dependent variable column corresponds to whether the LHS variable is the log money-consumption ratio or the log-level of real balances.

D Welfare

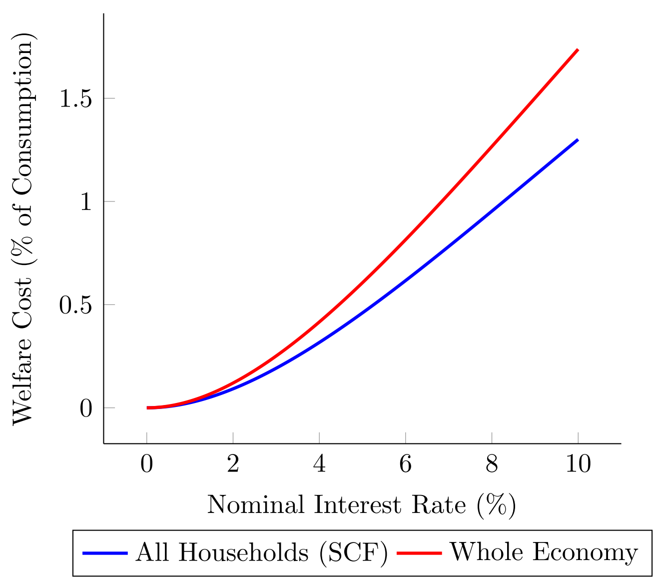


Figure 7: Aggregate welfare cost of inflation for the SCF and the whole economy.

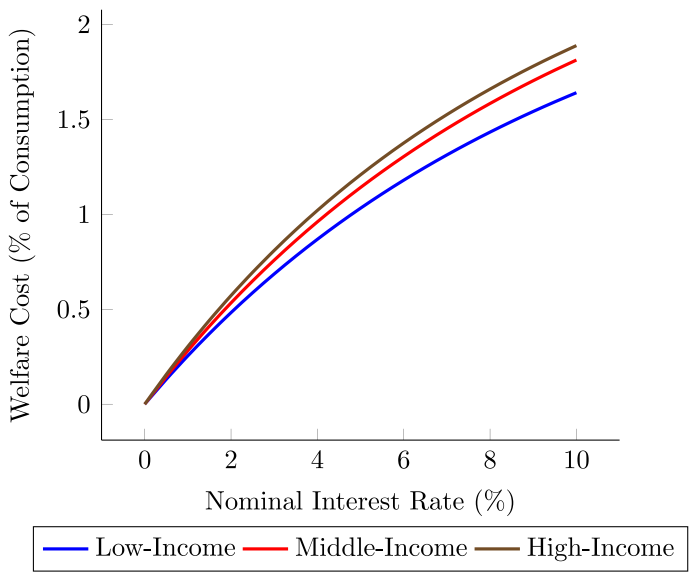


Figure 8: Dynamic welfare cost of inflation with no redistribution for varying nominal interest rates.

E Data Appendix

E.1 Money Data

I use the Survey of Consumer Finance (SCF) to construct estimates of real balances ratios for low-, middle-, and high-income households. For the variables I require, there is complete coverage over the period 1989-2019 comprising eleven survey years. Luckily, these also correspond to periods for which there is significant variation in nominal interest rates. To construct a household-level version of Lucas and Nicolini's NewM1 money measure, I sum the value checkable deposits, call accounts, prepaid cards, and money market deposit accounts. All of these come from the summary-level extract of the SCF.

Since my focus is on working age households, I filter out retirees. After that, I sort households into their income categories for each year based on the income category provided by the SCF. These correspond to the CPS income cutoffs. Then, within each group for each year, I computed the weighted sum of household-level NewM1.

E.2 Imputation Methodology

SCF asset data is published every three years, which presents an issue for considering any kind of long-run money demand function over an appropriately lengthy window. To obtain estimates of money demand as a share of income at an annual frequency, I follow the methodology of Stock and Watson (2016) and McGrattan (2020). That is, I use a Kalman filter to compute forecasts of annual money-income ratios given related series available at a higher frequency and then apply a Kalman smoother on the forecasts. Specifically, I use the annual aggregate money-income ratio and household-level liquid asset-income ratios from the CEX, both of which are available annually.

Let Z_t be a variable available every three years. Select X_t variables published from other sources available at annual frequency and used to make inferences about the annual value of Z_t, which I call $\hat{Z}_t$. The first step is to detrend all time series Z_t and X_t using cubic splines. Then, to obtain annual estimates of $\hat{Z}_t$, I estimate A and B in the following state space system via maximum likelihood:

$$\begin{aligned} x_{t+1} &= Ax_t + B\epsilon_{t+1} \\ y_t &= Cx_t, \end{aligned} \tag{E.1}$$

where $x_t = \left[X_t, \hat{Z}_t, X_{t-1}, \hat{Z}_{t-1}, X_{t-2}, \hat{Z}_{t-2}, X_{t-3}, \hat{Z}_{t-3}\right]^T$, $y_t = \left[X_t, Z_t\right]^T$, and ϵ_t are normally distributed shocks. Coefficients are given by

$$A = \begin{bmatrix} a_1 & a_2 & \dots & a_j \\ I & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \; B = \begin{bmatrix} b \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; C = I \tag{E.2}$$

With estimates for $\left(\hat{A}, \hat{B}\right)$, it becomes possible to estimate forecasts $\hat{Z}_t = \mathbb{E}[Z_t | y_1, \dots, y_T]$

of annual data at a quarterly frequency by applying a Kalman smoother and then adding back the low-frequency trend to the estimated time series. I replace imputed estimates with their realizations where applicable.

I utilize the procedure outlined above to obtain annual real balances by income class. In particular, I use aggregate data on money demand as well as Consumer Expenditure Survey (CEX) data on liquid assets grouped by the appropriate income thresholds. I detail construction of each series below.

Additionally, one problem is that there is some scaling disagreement between the SCF and the CEX. Rather than directly use CEX consumption data, I first multiply group-level income in the SCF by the corresponding consumption-income ratio from the CEX. Then, using aggregate consumption from Line 2 of NIPA Table 1.5 and CEX consumption for the corresponding group, I impute consumption using the same state-space methodology. This results in the corresponding money-consumption ratio series I plot in Figure 5 against the nominal short rate.

E.2.1 Aggregate Money Real Balances

Aggregate moneyholdings up until 2015 come from Benati et al. (2021), while moneyholdings after that were provided privately by Han Gao as part of Gao, Kulish, and Nicolini (2020). These are aggregate holdings of currency and checkable deposits as well as money-market deposit accounts (NewM1 as defined by Lucas and Nicolini (2015)).

E.2.2 CEX Liquid Asset Data

For almost the entirety of its duration, the Consumer Expenditure Survey (CEX), conducted by the Bureau of Labor Statistics, has collected some level of data on liquid asset holdings from households. Unfortunately, the level of aggregation changes substantially over time and in recent years has entirely merged savings and checkings accounts into one category without paying particular attention to important components of liquid assets which would be ideal to include in a household measure of M1. Naturally, this is largely because the CEX is not designed to measure asset holdings, whereas the SCF is. Because the definition of liquid holdings becomes more aggregated over time, I try to likewise remain aggregated in my measurements so that there are not random jumps due to definitional changes rather than actual changes in balances.

All data come from the interview file (fmli) of the CEX. For each year to get some measure of a liquid assets-income ratio and a household's distance from the poverty line, several variables are sufficient: age, household size, gross (pre-tax and -transfer) household income, household weight, and some measure of liquid assets. There are three distinct periods: 1989-2003, 2004-2012, and 2013-2019. To compute a liquid assets-income ratio, the procedure is largely the same for each period. First, compute a household weight variable based on FINLWT21 and the variables QINTRVMO and QINTRVYR. The latter two determine how long the household is in the sample, while the first weights the household.[15] Following that, I remove observations for which the population weight is undefined or zero, for which there are no observations of the relevant variables, and for which the age of the household is greater than 65. Next I group households according to their income. Finally, I summed weighted liquid holdings for each income category. Differences between each of the three periods are detailed below. These differences are driven largely by study design and ideally they would be constant across all years. The fact that they are not is a major impetus for using the SCF instead.

- **1989-2003:** For liquid assets, I sum the variables CKBKACTX (checking account value) and SAVACCTX (savings account value). Income is defined as the sum of EARNINCX (earned income) and NO_EARNX (unearned income).

- **2004-2012:** The liquid assets and household size variables are the same but I use the variable FINCBTXM (family income) for household income.

15. Instructions for this procedure can be found in Section 6.3.1 of this BLS document..

- **2013-2019:** The household size and income variables are the same. At the this point, the CEX began lumping together checking and savings accounts into a single variable called LIQUIDX, which I use for this period. In cases where this is unavailable, I use LIQUIDBX.

E.3 Transition Probabilities

Every year, the U.S. Census Bureau publishes the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). The ASEC also has a useful longitudinal component, in which households interviewed as part of the last year's cohort are interviewed once more in the following year. Among other relevant characteristics, ASEC contains data on household income and household size. These are sufficient to construct estimates of the household's income relative to the poverty line for a household of their characteristics, where the poverty line is defined by the Census Bureau. I use these guidelines to define each household's status relatve to the poverty line for each year from 1989-2019 and then categorize them. In particular, I use whether or not the head of the household transferred groups from year to year. After filtering out retirees and households that enter retirement in the following year, that consists of dividing the variable OFFTOTAL (official poverty line income) by the variable OFFCUTOFF (official poverty line cutoff) for each year.[16] After that, I count the number of households which transferred into each category, weighted by the longitudinal weight LNKFW1YWT_1. Then I counted the weighted total of each initial group (low-, medium-, and high-income) and computed the share of each that went into the low-, middle-, and high-income categories. Within each group, I average equally over all years in the sample. These averages form the transition probabilities. Note that the years 1994-1995 are not in the sample because the Census Bureau could not properly link the ASEC supplement in those years.

16. Note that people may leave the labor force for other reasons (e.g., injury). I leave them in the sample because the transitions are largely about idiosyncratic shocks, to the extent I can account for them. Retirement is not a shock.