



Research Division
Federal Reserve Bank of St. Louis
Working Paper Series



**Liquidity Demand and Welfare in a Heterogeneous-Agent
Economy**

Yi Wen

Working Paper 2010-009A
<http://research.stlouisfed.org/wp/2010/2010-009.pdf>

March 2010

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

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

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to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Liquidity Demand and Welfare in a Heterogeneous-Agent Economy*

Yi Wen

Federal Reserve Bank of St. Louis
& Tsinghua University

First Version: April, 2009. This Version: March 12, 2010

Abstract

This paper provides an analytically tractable general-equilibrium model of money demand with micro-foundations. The model is based on the incomplete-market model of Bewley (1980) where money serves as a store of value and provides liquidity to smooth consumption. The model is applied to study the effects of monetary policies. It is shown that heterogeneous liquidity demand can lead to sluggish movements in aggregate prices and positive responses from aggregate output to transitory money injections. However, permanent money growth can be extremely costly: With log utility function and an endogenously determined distribution of money balances that matches the household data, agents are willing to reduce consumption by 8% (or more) to avoid 10% annual inflation. The large welfare cost of inflation arises because inflation destroys the liquidity value and the buffer-stock function of money, thus raising the volatility of consumption for low-income households. The astonishingly large welfare cost of moderate inflation provides a justification for adopting a low inflation target by central banks and offers an explanation for the empirical relationship between inflation and social unrest in developing countries.

Keywords: Liquidity Preference, Money Demand, Incomplete Markets, Velocity, Welfare Costs of Inflation.

JEL codes: E12, E31, E41, E51.

*This paper is a significantly revised version of Wen (2009a). I thank Pengfei Wang for discussions on issues related to this project; Costas Azariadis, Jinghui Bai, Aleksander Berentsen, Gabriele Camera, Mark Huggett, Qing Liu, Rodolfo Manuelli, Pengfei Wang, Steve Williamson, and seminar participants at Georgetown University and St. Louis Fed for comments; Judy Ahlers for editorial assistance; and Yu man Tam and Luke Shimek for research assistance. The views expressed do not reflect official positions of the Federal Reserve Bank of St. Louis and the usual disclaimer applies. Correspondence: Yi Wen, Research Department, Federal Reserve Bank of St. Louis, P.O. Box 422, St. Louis, MO, 63166. Phone: 314-444-8559. Fax: 314-444-8731. Email: yi.wen@stls.frb.org.

1 Introduction

In developing countries, liquid money (cash and checking accounts) is the major form of household financial wealth and a vital tool of self-insurance to buffer idiosyncratic shocks because of the lack of a well-developed financial system. Based on recent data in China and India, more than 90% of the household financial wealth is held in the form of cash and checking accounts.¹ Even in developed countries, because of borrowing constraints, money remains one of the most important assets to provide liquidity to smooth consumption, especially for low-income households. In the United States, money demand is highly heterogeneous: The Gini coefficient of the distribution of money across households is greater than 0.85. This degree of heterogeneity in money demand closely resembles the distribution of financial wealth instead of consumption (with a Gini coefficient less than 0.3). This suggests that the liquidity motive of money demand is at least as important (if not more so) as the transaction motive of money demand, even in developed countries like the United States.²

When money is essential for consumption smoothing and is unequally distributed across households, largely because of idiosyncratic needs for liquidity and lack of sophisticated risk sharing, inflation can be far more costly than recognized by the existing literature. Historical evidence also suggests that moderate inflation (around 10% to 20% a year) may be significant enough to cause widespread social and political unrest in developing countries.³ Yet, the existing monetary literature suggests that the cost of inflation is surprisingly small. For example, Lucas (2000) recently estimated that the welfare cost of increasing inflation from 4% to 14% is less than 1% of aggregate output.⁴ Such results are disturbing; for if this is true, then the commonly accepted inflation target of 2% a year by most central banks in developed countries may be too conservative and not well justified, and this policy may have forgone too large social benefits of potentially higher employment through faster money growth.

¹Townsend (1995) points out that currency and crop inventory are the major forms of liquid assets to provide self-insurance against idiosyncratic shocks for farmers in India and Thailand, and surprisingly, purchases and sales of real capital assets, including livestock and consumer durables, do not play a role in smoothing income fluctuations.

²Ragot (2009) reports that this stylized fact holds for other developed countries and argues that this is a problem for theories that directly link money demand to consumption, such as cash-in-advance (CIA), money-in-the-utility (MIU), or shopping-time models, but consistent with incomplete-market models in which money is held as a form of financial asset that provides liquidity to smooth consumption.

³See Cartwright, Delorme, and Wood (1985) and Looney (1985) for empirical studies on the relationship between inflation and revolutions in recent world history. Using data from 54 developing countries, Cartwright, Delorme, and Wood (1985) find that inflation is the most significant economic variable to explain the probability and duration of social unrest and revolution, and is far more important than other economic variables, such as income inequality, GDP per capita, income growth, unemployment rate, and degree of urbanization. Their estimates show that a one-unit increase in the inflation rate raises the probability of revolution by 6 percentage points and increases the duration of revolution by 0.7 to 1.0 years.

⁴Similar estimates are also obtained by many others, such as Cooley and Hansen (1989), Dotsey and Ireland (1996), Lagos and Wright (2005), and Henriksen and Kydland (2010) in different models.

To properly assess the welfare costs of inflation, it is desirable that a theoretical model takes the liquidity function of money and the precautionary motives of money demand into account and matches at least two styled facts: (i) the interest elasticity of money demand (as suggested by Bailey, 1956),⁵ and (ii) the cross-household distribution of money holdings. The first criterion allows the model to capture the opportunity cost of holding non-interest-bearing cash, whereas the second ensures that it captures the idiosyncratic liquidity risk or the heterogeneous "adverse liquidity effects" of inflation on the population.⁶

This paper constructs such a model by generalizing Bewley's (1980, 1983) precautionary money demand model to a dynamic stochastic general-equilibrium framework.⁷ The key feature distinguishing Bewley's model from the related literature, such as the cash-in-advance (CIA) model of Lucas (1980) and the (S,s) inventory-theoretic model of Baumol (1952) and Tobin (1956), is that money is held solely as a store of value, completely symmetric to any other asset, and is not imposed from outside as the means of payments. Agents can choose whether to hold money depending on the costs and benefits.⁸ By freeing money from its role of medium of exchange, Bewley's approach allows us to focus on the function of money as a pure form of liquidity, so that the welfare implications of the liquidity-preference theory of money demand can be investigated in isolation. Beyond Bewley (1980, 1983), my generalized model is analytically tractable; hence, it greatly simplifies the computation of dynamic stochastic general equilibrium in environments with capital accumulation and nontrivial distributions of cash balances, thus facilitating welfare and business-cycle analysis. Analytical tractability makes the mechanisms of the model transparent.

The major findings of the paper include the following: (i) In sharp contrast to standard monetary models, the generalized Bewley model is able to produce enough variability in velocity relative to output to match the data; in particular, it can explain the negative correlation of velocity with real balances in the short run and its positive correlation with inflation in the long run. (ii) Transitory lump-sum money injections can have positive real effects on aggregate activities despite flexible prices. (iii) Persistent money growth is very costly: When the model is calibrated to match not only the interest elasticity of aggregate money demand but also the cross-household distribution of money holdings in the data, the implied welfare cost of moderate inflation is astonishingly large,

⁵Bailey (1956) first showed that the welfare cost of inflation arising from the inefficiencies of carrying out transactions with means of payment that do not pay interest can be measured by the integral under the demand for money.

⁶An "adverse liquidity effect" of inflation refers to the effect on consumption from the loss of the liquidity value of money due to inflation, which destroys the self-insurance function of money and subjects people to idiosyncratic shocks.

⁷General-equilibrium analysis is important. Cooley and Hansen (1989) emphasize the general-equilibrium effect of inflation on output through substituting leisure for consumption in the face of a positive inflation, which causes labor supply and output to decline. However, because these authors assume that money is held only for transaction purpose, the welfare cost of inflation is still small despite the general-equilibrium effects of inflation on output, about 0.4% of GDP with 10% inflation.

⁸In contrast, in the models of Lucas (1980), Baumol (1952), and Tobin (1956), money is held both as a store of value (for precautionary reasons) and as a medium of exchange (for transaction purposes).

at least 8% of consumption under 10% annual inflation.

Since holding money is both beneficial (providing liquidity) and costly (forgoing interest payment), agents opt to hold different amounts of cash depending on income levels and consumption needs. As a result, a key property of the model is an endogenously determined distribution of money holdings across households, with a strictly positive fraction of households being cash-constrained. Hence, lump-sum money injections have an immediate positive impact on consumption for liquidity-constrained agents, but not for agents with idle cash balances. Consequently, aggregate price does not increase with aggregate money supply one-for-one, so transitory monetary shocks are expansionary to aggregate output (even without open market operations), the velocity of money is countercyclical, and the aggregate price appears "sticky."

However, with anticipated inflation, permanent money growth reduces welfare significantly for several reasons: (i) Precautionary money demand induces agents to hold excessive amount of cash to avoid liquidity constraints, raising the inflation tax on the population. (ii) Liquidity-constrained agents suffer disproportionately more from inflation tax because they are subject to idiosyncratic risks without self-insurance; thus, for the same reduction in real wealth, inflation reduces their expected utility more than it does for liquidity-rich agents. (iii) The size of the liquidity-constrained population rises rapidly with inflation, leading to an increased portion of the population unable to smooth consumption against idiosyncratic shocks. This factor can dramatically raise social welfare costs along the extensive margin. (iv) Agents opt to switch from "cash" goods (consumption) to "credit" goods (leisure), thereby reducing labor supply and aggregate output.⁹ These effects interact and compound each other, leading to large welfare losses.

The Bailey triangle is a poor measure of the welfare costs of inflation because it fails to capture the insurance function of money (as noted by Imrohoroglu, 1992). At a higher inflation rate, not only does the opportunity cost of holding money increase (which is the Bailey triangle), but the crucial benefit of holding money also diminishes. In particular, when demand for money declines, the portion of the liquidity-constrained population rises; consequently, the welfare cost of inflation increases sharply due to the loss of self-insurance for an increasingly larger proportion of the population.¹⁰ This result is reminiscent of the analysis in Aiyagari (1994) where he shows that the welfare cost of the loss of self-insurance in an incomplete-market economy is equivalent to a 14% reduction in consumption even though his calibrated model matches only one third of the income and wealth inequalities in the data.

This paper is also related to the work of Alvarez, Atkeson, and Edmond (2008). Both papers are based on an inventory-theoretic approach with heterogeneous money demand and can explain

⁹Money facilitates consumption by providing liquidity, but consuming leisure does not require liquidity.

¹⁰This important extensive margin is not fully captured by Imrohoroglu (1992).

the short-run dynamic behavior of velocity and sticky aggregate prices under transitory monetary shocks. However, my approach differs from theirs in several aspects. First, their model is based on the Baumol-Tobin inventory-theoretic framework where money is not only a store of value but also a means of payment (similar to CIA models).¹¹ Second, the distribution of money holdings is exogenously given in their model; hence, the portion of population with the need for cash withdrawals is fixed and cannot respond to monetary policy.¹² Third, because agents are exogenously and periodically segregated from the banking system and the CIA constraint always binds, the expansionary real effects of monetary shocks cannot be achieved through lump-sum money injections in their model. For these reasons, the implications of the welfare costs of inflation in their model may also be very different from those in this paper. For example, Attanasio, Guiso, and Jappelli (2002) estimate the welfare costs of inflation based on a simple Baumol-Tobin model and find the cost to be less than 0.1% of consumption under 10% inflation. The main reason is that this segment of the literature has relied exclusively on Bailey’s triangle (or the interest elasticity of money demand) to measure the welfare costs of inflation. Hence, despite having heterogeneous money holdings across households, such research is not able to obtain significantly larger estimates of the cost of inflation than those in representative-agent models.

Bewley’s (1980) model has been studied in the recent literature, but the main body of this literature focuses on an endowment economy. For example, Imrohoroglu (1992) and Akyol (2004) study the welfare costs of inflation in the Bewley model. To the best of my knowledge, Imrohoroglu (1992) may be the first one to recognize that the welfare costs of inflation in a Bewley economy are larger than that suggested by the Bailey triangle. However, like Bewley (1980), this segment of the literature is based on an endowment economy without production and capital and these models are not analytically tractable. In addition, Imrohoroglu and Akyol do not calibrate their models to match the actual distribution of money holdings in the data. To facilitate numerical computations, such models typically assume that the portion of the liquidity-constrained population is exogenously given rather than endogenously determined (e.g., by assuming a binary-point distribution of idiosyncratic shocks). Consequently, the distribution of money demand does not shift in reaction to monetary policies, strongly dampening the adverse liquidity effects of inflation on welfare along the extensive margin.¹³ Ragot (2009), on the other hand, uses a general-equilibrium version of Bewley’s model with segregated markets (similar to Alvarez, Atkeson, and Edmond, 2008) to explain the joint distribution of money demand, consumption, and financial assets. He shows

¹¹For the more recent literature based on the Baumol-Tobin model, see Alvarez, Atkeson and Kehoe (1999), Bai (2005), Chiu (2007), and Khan and Thomas (2007), among others.

¹²Chiu (2007) and Khan and Thomas (2007) relaxed this assumption by having an endogenously determined number of trips for cash withdraws.

¹³This literature tends to find higher welfare costs of inflation, but the absolute magnitude is still small. For example, Imrohoroglu (1992) shows the welfare cost of 10% inflation is slightly above 1% of consumption.

that standard models in which money serves only as a medium of exchange are inconsistent with the empirical distributions of these variables, whereas a general-equilibrium version of the Bewley model in which money is held as a store of value can better explain the empirical distributions. Nonetheless, Ragot (2009) also uses a numerical approach to solve the model and he does not study the welfare implications of inflation.

Kehoe, Levine, and Woodford (1992) study the welfare effects of inflation in a Bewley model with aggregate uncertainty. They show that lump-sum nominal transfers can redistribute wealth from cash-rich agents to cash-poor agents, because the latter receive disproportionately more transfers than the former and thereby benefit from inflation. Consequently, inflation may improve social welfare. However, this positive effect on social welfare is quite small and requires extreme parameter values in their model, and such an effect exists in my model if money injection is purely transitory (i.e., without changing the steady-state stock of money). The Friedman rule is Pareto optimal in my model with anticipated inflation.

The rest of this paper is organized as follows: Section 2 presents the benchmark model on the household side and shows how to solve for individuals' decision rules of money demand and consumption analytically. It reveals some of the basic properties of a monetary model based on liquidity preference. Section 3 uses the model to match the empirical distributions of wealth, money demand, and consumption across U.S. households and the interest elasticity of aggregate money demand over time. Section 4 closes the model with firms and studies the impulse responses of the model to transitory monetary shocks in general equilibrium. Section 5 discusses calibrations and studies the welfare costs of inflation. Section 6 concludes the paper with remarks for future research.

2 The Model

2.1 Households

The model is a stochastic general-equilibrium version of the model of Bewley (1980, 1983).¹⁴ This section studies a partial-equilibrium version of the model without firms. To highlight the liquidity value of money, the model features money as the only asset that can be adjusted quickly and costlessly to buffer idiosyncratic shocks from income or wealth at any moment. Interest-bearing nonmonetary assets (such as capital) can be accumulated to support consumption, but are not as useful (or liquid) as money in buffering idiosyncratic shocks. This setup captures the characteristics of the lack of well-developed financial markets and risk-sharing arrangements in developing countries.

¹⁴In contrast to Bewley (1980, 1983), money does not earn interest in my model; hence the insatiability problem discussed by Bewley does not arise. Consequently, a monetary equilibrium always exists under the Friedman rule as long as the support of the distribution of shocks is bounded.

I make the model analytically tractable by introducing two important features: (i) I allow endogenous labor supply with quasi-linear preferences (as in Lagos and Wright, 2005), and (ii) I replace idiosyncratic labor income shocks typically assumed in the literature (e.g., Imrohoroglu, 1989, 1992; Aiyagari, 1994; Huggett, 1993, 1997) by other types of shocks. Even with quasi-linear preferences, the model is not analytically tractable if labor income is directly subject to idiosyncratic shocks. There are two ways to overcome this difficulty. One is to place idiosyncratic shocks in preferences (i.e., to the marginal utility of consumption as in Lucas, 1980), and the other is to place them on net wealth. This paper takes the second approach.¹⁵ Both alternatives yield similar results for the questions addressed in this paper. This is reassuring because it suggests that the source of uninsurable idiosyncratic shocks does not matter for my results. Having the idiosyncratic uncertainty originating from preferences or household wealth provides an additional advantage, which enables the model to match the distribution of money demand and consumption in the data more easily than labor-income shocks. As shown by Aiyagari (1994) and Krusell and Smith (1998), idiosyncratic labor income shocks are not able to generate a sufficient degree of inequality to match the data, especially when such shocks are transitory. Preference or wealth shocks can do a better job in this regard because by assumption they are not fully self-insurable even when such shocks are *i.i.d.*

Time is discrete. There is a continuum of households indexed by $i \in [0, 1]$. Each household is subject to an idiosyncratic shock, $\theta_t(i)$, to its net wealth. $\theta_t(i)$ has the distribution $F(\theta) \equiv \Pr[\theta(i) \leq \theta]$ with support $[0, \bar{\theta}]$. A household chooses consumption $c_t(i)$, labor supply $n_t(i)$, a nonmonetary asset $s_t(i)$ that pays the real rate of return $r_t > 0$, and nominal balance $m_t(i)$ to maximize lifetime utilities, taking as given the aggregate real wage (w_t), the market real interest rate (r_t), the aggregate price (P_t), and the nominal lump-sum transfers (τ_t). Although money is dominated in the expected rate of return by nonmonetary assets, it is more liquid than other assets (such as capital) as a store of value (— the definition of liquidity to be specified below). Hence, by providing liquidity to facilitate consumption demand, money can coexist with interest-bearing assets.

To capture the liquidity role of money, assume that the decisions for labor supply ($n_t(i)$) and holdings for interest-bearing assets ($s_t(i)$) in each period t must be made before observing the idiosyncratic shock $\theta_t(i)$ in that period, and the decisions, once made, cannot be changed for the rest of the period (i.e., these markets are closed for households afterward until the beginning of the next period). Thus, if there is an urge to consume or an unexpected change in wealth during period t after labor supply and investment decisions are made, money stock is the only asset that

¹⁵See the earlier version of this paper, Wen (2009a), for analysis under the first approach.

can be adjusted to smooth consumption.¹⁶ Borrowing of liquidity (money) from other households is not allowed (i.e., $m(i) \geq 0$).¹⁷ These assumptions imply that households may find it optimal to carry money as inventories (self-insurance device) to cope with income uncertainty (as in Bewley, 1980); even though money is not required as a medium of exchange, it can be used to exchange for consumption goods. As in the standard literature, any aggregate uncertainty is resolved at the beginning of each period before any decisions are made and is orthogonal to idiosyncratic uncertainty.

An alternative way of formulating the above information structure for decision-making is to divide each period into two subperiods, with labor supply and nonmonetary-asset investment determined in the first subperiod, the rest of the variables (consumption and money holdings) determined in the second subperiod, and the idiosyncratic shocks $\theta_t(i)$ realized only in the beginning of the second subperiod. Yet another alternative specification of the model is to have two islands, with $n_t(i)$ and $s_t(i)$ determined in island 1 and $c_t(i)$ and $m_t(i)$ determined in island 2 simultaneously by two spatially separated household members (e.g., a worker and a shopper), but only the shopper — who determines consumption and money balances in island 2 — can observe $\theta_t(i)$ in period t . Both members can observe aggregate shocks and the history of decisions up to period t . At the end of each period the two members reunite and share everything perfectly (e.g., income, wealth, and information) and separate again in the beginning of the next period.

These information structures amount to creating a necessary friction for the existence of money and capturing the feature that labor income and nonmonetary assets are not as useful as money to respond to the random liquidity needs of households. In reality, especially in developing countries, it is costly to exchange labor and real assets for consumption goods in spot markets (e.g., due to search frictions). Even in developed countries, government bonds are rarely held as a major form of liquid assets by low-income households.

With the setup and the information structure in mind, household i 's problem is to solve¹⁸

$$\max_{\{c,m\}} E_0 \left\{ \max_{\{n,s\}} \tilde{E}_0 \left\{ \sum_{t=0}^{\beta} \beta^t [\log c_t(i) - an_t(i)] \right\} \right\} \quad (1)$$

¹⁶This timing friction is what we need to generate a positive liquidity value of money over other assets in equilibrium. This type of timing friction is also assumed by Aiyagari and Williamson (2000) and Akyol (2004) in endowment economies with random income shocks. It is also akin to the transaction costs approach of Aiyagari and Gertler (1991), Chatterjee and Corbae (1992), Greenwood and Williamson (1989), and Ragot (2009).

¹⁷See Wen (2009a) for an extension of the model to allowing for borrowing and lending through financial intermediation.

¹⁸The consumption utility function can be more general without losing analytical tractability. For example, the model can be solved as easily if $u(c) = \frac{c(i)^{1-\sigma}}{1-\sigma}$.

subject to

$$c_t(i) + \frac{m_t(i)}{P_t} = [\varepsilon + \theta_t(i)] x_t(i) \quad (2)$$

$$m_t(i) \geq 0, \quad (3)$$

where

$$x_t(i) \equiv \frac{m_{t-1}(i) + \tau_t}{P_t} + w_t n_t(i) + (1 + r_t) s_{t-1}(i) - s_t(i) \quad (4)$$

defines the net wealth of the household, which includes last-period real balances, labor income and capital gains, and subtracts net investment in interest-bearing assets. The expectation operator \tilde{E}_t denotes expectation conditional on the information set of the first subperiod in period t (which excludes $\theta_t(i)$), and the operator E_t denotes expectation based on information of the second subperiod (which includes $\theta_t(i)$). Without loss of generality, assume $a = 1$.¹⁹

Note that the net wealth on the right-hand side of equation (2) has a multiplier $\varepsilon + \theta_t(i)$, where $\varepsilon \in (0, 1)$ is a constant, and $\theta_t(i)$ is an idiosyncratic shock to net wealth. This implies that an ε fraction of the net wealth is never subject to idiosyncratic shocks. This parameter is important for allowing the model to match the joint distribution of money demand and consumption in the data. The expected value (mean) of $\theta(i)$ is normalized to $E\theta(i) = 1 - \varepsilon$, so that the average value of the multiplier $\theta + \varepsilon(i)$ equals 1, which implies that idiosyncratic shocks do not cause distortions to the resource constraint on average or at the aggregate level. The idiosyncratic wealth shock $\theta_t(i)$ is similar to an idiosyncratic tax shock and it implies that nature redistributes a portion of the net wealth randomly across households. This shock is also important for allowing the model to match the joint distributions of money demand and consumption in the data. This shock gives rise to incentives for households to hold money as a self-insurance device.²⁰ Unlike transitory labor income shocks, wealth shocks are only partially insurable through savings even though the shocks are *i.i.d.* This allows the model to match the empirical inequalities of money demand easily.

Since money is not required as a medium of exchange, choosing $m_t(i) = 0$ for all t is always an equilibrium. In what follows, we focus on monetary equilibria where money is accepted as a store of value. Denoting $\{\lambda(i), v(i)\}$ as the Lagrangian multipliers for constraints (2) and (3), respectively,

¹⁹To make money completely symmetric to interest-bearing assets, we can also impose borrowing constraints on non-monetary assets. However, as long as the real rate of return to such assets exceeds one (— which is the case when r_t is endogenous), borrowing constraints do not bind for these assets in the steady state. Hence, such constraints are ignored in this paper.

²⁰The multiplicative assumption of individual wealth shocks implies that the degree of inequality in wealth distribution does not shrink but remains stable as the economy grows. This implication is consistent with empirical evidence (see, e.g., Wolff, 1998). The sensitivity of the results to this assumption is analyzed in Wen (2009a), where I show that the results remain robust when the idiosyncratic shocks are from preferences instead of wealth.

the first-order conditions for $\{c(i), n(i), s(i), m(i)\}$ are given, respectively, by

$$\frac{1}{c_t(i)} = \lambda_t(i) \quad (5)$$

$$1 = w_t \tilde{E}_t \{[\varepsilon + \theta_t(i)] \lambda_t(i)\} \quad (6)$$

$$\tilde{E}_t \{[\varepsilon + \theta_t(i)] \lambda_t(i)\} = \beta E_t \tilde{E}_{t+1} \{(1 + r_{t+1}) [\varepsilon + \theta_{t+1}(i)] \lambda_{t+1}(i)\} \quad (7)$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t \left\{ \frac{[\varepsilon + \theta_{t+1}(i)] \lambda_{t+1}(i)}{P_{t+1}} \right\} + v_t(i), \quad (8)$$

where equations (6) and (7) reflect that decisions for labor supply $n_t(i)$ and asset investment $s_t(i)$ must be made before the idiosyncratic wealth shocks (and hence the value of $\lambda_t(i)$) are realized. Using equation (6), by the law of iterated expectations and the orthogonality assumption of aggregate and idiosyncratic shocks, equations (7) and (8) can be rewritten, respectively, as

$$\frac{1}{w_t} = \beta E_t (1 + r_{t+1}) \frac{1}{w_{t+1}} \quad (9)$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t \frac{1}{P_{t+1} w_{t+1}} + v_t(i), \quad (10)$$

respectively, where $\frac{1}{w}$ pertains to the expected marginal utility of consumption in terms of labor. Note that equation (9) is the standard Euler equation for asset accumulation and it implies $1 = \beta(1 + r)$ in the absence of aggregate uncertainty (i.e., in the steady state). Thus, idiosyncratic shocks have no effects on the decisions of nonmonetary asset demand in the model, which simplifies the analysis.

The decision rules for an individual's consumption and money demand are characterized by a cutoff strategy, taking as given the aggregate environment. Assuming interior solutions for labor supply and nonmonetary asset holdings and in anticipation that the cutoff, θ_t^* , is independent of i , we consider two possible cases below.

Case A. $\theta_t(i) \geq \theta_t^*$. In this case, the net wealth level is high. It is hence optimal to hold money as inventories to prevent possible liquidity constraints in the future. So $m_t(i) \geq 0$, $v_t(i) = 0$, and the shadow value of good $\lambda_t(i) = \beta E_t \frac{P_t}{w_{t+1} P_{t+1}}$. Equation (5) implies that consumption is given by $c_t(i) = \left[\beta E_t \frac{P_t}{w_{t+1} P_{t+1}} \right]^{-1}$. The budget identity (2) then implies $\frac{m_t(i)}{P_t} = [\varepsilon + \theta_t(i)] x_t(i) - \left[\beta E_t \frac{P_t}{w_{t+1} P_{t+1}} \right]^{-1}$. The requirement $m_t(i) \geq 0$ then implies

$$\varepsilon + \theta_t(i) \geq \frac{1}{x_t(i)} \left[\beta E_t \frac{P_t}{w_{t+1} P_{t+1}} \right]^{-1} \equiv \varepsilon + \theta_t^*, \quad (11)$$

which defines the cutoff θ_t^* .

Case B. $\theta_t(i) < \theta_t^*$. In this case, the net wealth level is low. It is then optimal to spend all money in hand to smooth consumption, so $v_t(i) > 0$ and $m_t(i) = 0$. By the resource constraint (2), we have $c_t(i) = [\varepsilon + \theta_t(i)] x_t(i)$, which by equation (11) implies $c(i) = \frac{\varepsilon + \theta_t(i)}{\varepsilon + \theta_t^*} \left[\beta E_t \frac{P_t}{w_{t+1} P_{t+1}} \right]^{-1}$. Equation (5) then implies that the shadow value is given by $\lambda_t(i) = \frac{\varepsilon + \theta_t^*}{\varepsilon + \theta_t(i)} \left[\beta E_t \frac{P_t}{w_{t+1} P_{t+1}} \right]$. Since $\theta(i) < \theta^*$, equation (10) confirms that $v_t(i) > 0$, provided that $P_t < \infty$. Notice that the shadow value of goods, $\lambda(i)$, is higher under case B than under case A because of a tighter budget constraint under case B.

The above analyses imply that the expected shadow value of goods, $\tilde{E}_t \{[\varepsilon + \theta_t(i)] \lambda_t(i)\}$, and hence the optimal cutoff value, θ_t^* , is determined by the following asset-pricing equation for money (based on equation 6):

$$\frac{1}{w_t} = \left[\beta E_t \frac{P_t}{P_{t+1} w_{t+1}} \right] R(\theta_t^*), \quad (12)$$

where

$$R(\theta_t^*) \equiv \left[\int_{\theta(i) < \theta_t^*} [\varepsilon + \theta_t^*] dF(\theta) + \int_{\theta(i) \geq \theta_t^*} [\varepsilon + \theta_t(i)] dF(\theta) \right] \quad (13)$$

measures the (shadow) rate of return to money (or the liquidity premium). The LHS of equation (12) is the opportunity cost of holding one unit of real balances as inventory (as opposed to holding one more unit of real asset). The RHS is the expected gains by holding money, which take two possible values: The second term on the RHS is simply the discounted next-period utility cost of inventory ($\beta E_t \frac{\varepsilon + \theta_t(i)}{P_{t+1} w_{t+1}}$) in the case of high income ($\theta(i) \geq \theta^*$), which has probability $\int_{\theta(i) \geq \theta^*} dF(\theta)$. The first term on the RHS is the marginal utility of consumption ($\frac{\varepsilon + \theta(i)}{c_t(i)} = \beta E_t \frac{\varepsilon + \theta_t^*}{P_{t+1} w_{t+1}}$) in the case of low income ($\theta(i) < \theta^*$), which has probability $\int_{\theta(i) < \theta^*} dF(\theta)$. The optimal cutoff θ^* is chosen so that the marginal cost equals the expected marginal gains. Hence, the rate of return to investing in money (or the liquidity premium) is determined by $R(\theta^*)$.

Notice that $R(\theta_t^*) > 1$ (recall $E\theta(i) = 1 - \varepsilon$), which implies that the option value of one dollar exceeds 1 because it provides liquidity in the case of a low income shock. This is why money has positive value in equilibrium despite the fact that its real rate of return is negative and dominated by interest-bearing assets. The optimal level of cash reserve (money demand) is always such that the probability of stockout (i.e., being liquidity constrained) is strictly positive (i.e., $F(\theta_t^*) > 0$) unless the cost of holding money is zero. This inventory-theoretic formula of the rate of return to liquidity is akin to that derived by Wen (2008) in an inventory model based on the stockout-

avoidance motive. Also note that aggregate shocks will affect the distribution of money holdings across households by affecting the cutoff θ_t^* .

Most importantly, equation (12) implies that the cutoff θ_t^* is independent of i . This property facilitates aggregation and makes the model analytically tractable. Consequently, numerical solution methods (such as the method of Krusell and Smith, 1998) are not needed to solve the model's general equilibrium and aggregate dynamics.

By equation (11), the net wealth is given by $x_t(i) = \frac{1}{\varepsilon + \theta_t^*} \left[\beta E_t \frac{P_t}{w_{t+1} P_{t+1}} \right]^{-1}$, which is also independent of i . Hence, hereafter we drop the index i from x_t .²¹ The intuition for x_t being independent of i is that (i) x_t is determined before the realization of $\theta_t(i)$ and all households face the same distribution of idiosyncratic shocks when making labor supply and capital accumulation decisions, and (ii) the quasi-linear preference structure implies that labor supply can be adjusted elastically to meet any target level of wealth x_t in anticipation of the possible wealth shocks. Hence, in the beginning of every period agents opt to adjust labor supply so that the target wealth and probability of liquidity constraint in this period are the same across all states of nature, with the target wealth depending only on aggregate states of the economy. This property makes x_t and the cutoff θ_t^* independent of individual history.

Since the Euler equation for interest-bearing assets is given by $\frac{1}{w_t} = \beta E_t (1 + r_{t+1}) \frac{1}{w_{t+1}}$, ignoring the covariance terms and comparing with equation (12), we have $R(\theta_t^*) = \frac{P_{t+1}}{P_t} (1 + r_{t+1})$. This suggests that the equilibrium rate of return to money is positively related to the Fisherian form of nominal interest rate on nonmonetary assets. This asset-pricing implication for the value of money is similar to that discussed by Svensson (1985) in a model with an occasional binding CIA constraint.

Using equation (11), the decision rules of household i are summarized by

$$c_t(i) = \min \left\{ 1, \frac{\varepsilon + \theta_t(i)}{\varepsilon + \theta_t^*} \right\} \times [\varepsilon + \theta_t^*] x_t, \quad (14)$$

$$\frac{m_t(i)}{P_t} = \max \left\{ \frac{\theta_t(i) - \theta_t^*}{\varepsilon + \theta_t^*}, 0 \right\} \times [\varepsilon + \theta_t^*] x_t, \quad (15)$$

$$[\varepsilon + \theta_t^*] x_t = w_t R(\theta_t^*). \quad (16)$$

The marginal propensity to consume out of target wealth is $\min \left\{ 1, \frac{\varepsilon + \theta_t(i)}{\varepsilon + \theta_t^*} \right\}$, which is less than 1 in the case of a low wealth shock ($\theta_t(i) < \theta_t^*$). These decision rules indicate that consumption is lower

²¹However, the effective wealth level depends on i because it is given by $[\varepsilon + \theta_t(i)] x_t$.

when the liquidity constraint (3) binds. That is, when $\theta(i) < \theta^*$, we have $c(i) = [\varepsilon + \theta_t(i)]x_t < [\varepsilon + \theta_t^*]x_t$ and $\frac{m(i)}{P} = 0$.

Hence, if inflation reduces real wealth x_t (holding θ_t^* constant), then both cash-rich and cash-poor agents are worse off in terms of consumption level, since average consumption is lower for all agents. However, since the variance of $\theta(i)$ remains the same, a lower real wealth implies that liquidity-constrained agents suffer disproportionately more from inflation than do cash-rich agents in terms of welfare (because idiosyncratic shocks affect only cash-poor agents). More importantly, if inflation increases the cutoff θ_t^* (holding x_t constant), then the probability of facing a binding liquidity constraint rises, leading to a larger proportion of the population without self-insurance, thereby reducing social welfare significantly along the extensive margin. This last channel is completely missing in the existing literature.

2.2 Partial Equilibrium Analysis

Aggregation. Denoting $C \equiv \int c(i)di$, $M \equiv \int m(i)di$, $S = \int s(i)di$, $N = \int (n(i)di$, and $X = \int x(i)di$, and integrating the household decision rules over i by the law of large numbers, the aggregate variables are given by

$$C_t = D(\theta_t^*)x_t, \quad (17)$$

$$\frac{M_t}{P_t} = H(\theta_t^*)x_t, \quad (18)$$

$$\frac{M_{t-1} + \tau_t}{P_t} + (1 + r_t)S_{t-1} - S_t + wN_t = x_t = \frac{w_t R(\theta_t^*)}{\varepsilon + \theta_t^*}, \quad (19)$$

where

$$D(\theta^*) \equiv \varepsilon + \int_{\theta(i) < \theta^*} \theta_t(i) dF(\theta) + \int_{\theta(i) \geq \theta^*} \theta_t^* dF(\theta), \quad (20)$$

$$H(\theta^*) \equiv \int_{\theta(i) \geq \theta^*} [\theta_t(i) - \theta_t^*] dF(\theta), \quad (21)$$

and these two functions satisfy $D(\theta^*) + H(\theta^*) = 1$ and $D(\theta^*) + R(\theta^*) = 1 + \varepsilon + \theta^*$. Thus, $D(\theta_t^*) \in (0, 1)$ is the aggregate marginal propensity to consume from the target wealth, and $H(\theta_t^*)$ is the marginal propensity to hold money. Because the cutoff θ_t^* is time varying, these marginal propensities are also time varying and depend on monetary policies.

Monetary Policy. We consider two types (regimes) of monetary policies. In the short-run dynamic analysis, money supply shocks are completely transitory without affecting the steady-state stock of money; namely, the lump-sum transfers follow the law of motion

$$\tau_t = \rho\tau_{t-1} + \bar{M}\varepsilon_t, \quad (22)$$

$$M_t = \bar{M} + \tau_t, \quad (23)$$

where $\rho \in [0, 1]$ and \bar{M} is the steady-state money supply. This policy implies the percentage deviation of money stock follows an $AR(1)$ process, $\frac{M_t - \bar{M}}{\bar{M}} = \rho \frac{M_{t-1} - \bar{M}}{\bar{M}} + \varepsilon_t$. Under this policy regime, the steady-state inflation rate is zero, $\pi = 0$.²²

In the long-run (steady-state) analysis, money supply has a permanent growth component with,

$$\tau_t = \bar{\mu} M_{t-1} \quad (24)$$

where $\bar{\mu}$ is a constant growth rate of money supply.

The Quantity Theory. The aggregate relationship between consumption (equation 17) and money demand (equation 18) implies the "quantity" equation,

$$P_t C_t = M_t V_t, \quad (25)$$

where $V_t \equiv \frac{D(\theta_t^*)}{H(\theta_t^*)}$ measures the aggregate consumption velocity of money.²³ A high velocity implies a low demand for real balances relative to consumption. Given the support of θ as $[0, \bar{\theta}]$ and the mean as $E\theta = 1 - \varepsilon$, by the definition for the functions D and H , it is easy to see that the domain of velocity is $\left[\frac{\varepsilon}{1-\varepsilon}, \infty\right]$, which has no finite upper bound, in sharp contrast to CIA models where velocity is typically bounded above by 1. An infinite velocity means that either the value of money ($\frac{1}{P}$) is zero or nominal money demand (M) is zero.

Steady-State Analysis. A steady state is defined as the situation without aggregate uncertainty. Hence, in a steady state all real variables are constant over time. The steady-state cutoff θ^* is determined by the relation

$$R(\theta^*) = \frac{1 + \pi}{\beta}, \quad (26)$$

where $\pi \equiv \frac{P_t - P_{t-1}}{P_{t-1}}$ is the steady-state rate of inflation. Hence, the cutoff θ^* is constant for a given level of inflation. The quantity relation (25) implies $\frac{P_t}{P_{t-1}} = \frac{M_t}{M_{t-1}} = \bar{\mu}$ in the steady state, so the steady-state inflation rate is the same as the growth rate of money.

Since by equation (26) the return to liquidity R must increase with π , the cutoff θ^* must also increase with π (because $\frac{\partial R(\theta^*)}{\partial \theta^*} = F(\theta^*) > 0$). This means that when inflation rises, the required rate of return to liquidity must also increase accordingly to induce people to hold money.

²²This monetary policy is similar to the U.S. Federal Reserve Bank's "Quantitative Easing" programs that inject liquidity into the economy during financial crisis but withdraw completely the injected money out of the economy slowly afterwards. Hence, the long-run stock of money is not affected under this type of policies.

²³Alternatively, we can also measure the velocity of money by aggregate income, $PY = M\tilde{V}$, where $\tilde{V} \equiv V\frac{Y}{C}$ is the income velocity of money.

However, because the cost of holding money increases with π , agents opt to hold less money so that the probability of stockout ($F(\theta^*)$) rises, which reinforces a rise in the equilibrium shadow rate of return (i.e., the liquidity premium), so that $\frac{\partial^2 R}{\partial(\theta^*)^2} > 0$. By definitions (20) and (21), we have $\frac{\partial D}{\partial \theta^*} = 1 - F(\theta^*) > 0$ and $\frac{\partial H}{\partial \theta^*} = F(\theta^*) - 1 < 0$. Also, since the target wealth is given by $x(\theta^*) = w \frac{R(\theta^*)}{\varepsilon + \theta^*}$, we have $\frac{\partial x}{\partial \theta^*} = w \frac{F[\varepsilon + \theta^*] - R}{[\varepsilon + \theta^*]^2} = w \frac{-\int_{\theta > \theta^*} [\varepsilon + \theta] dF(\theta)}{[\varepsilon + \theta^*]^2} < 0$, so the target wealth decreases with π . Therefore, a higher rate of inflation has two types of effects on welfare: the intensive margin and the extensive margin. On the intensive margin, $\frac{\partial x}{\partial \pi} < 0$, so lower wealth leads to lower consumption for all agents. In addition, liquidity-constrained agents suffer disproportionately more because they do not have self-insurance yet face the same variance of idiosyncratic shocks (σ_θ^2) while having a lower wealth level. This second aspect of the intensive margin is emphasized by Imrohorglu (1992) and Akyol (2004). On the extensive margin, $\frac{\partial \theta^*}{\partial \pi} > 0$, thus a larger portion of the population will become liquidity constrained and subject to idiosyncratic shocks. This extensive margin will be shown to be an important force to affect social welfare but has not been fully appreciated by the existing literature.

Under the Friedman rule, $1 + \pi = \beta$, we must have $R = 1$ and $\theta^* = 0$ according to equation (13), and consequently, $D(\theta^*) = \varepsilon$ and $H(\theta^*) = 1 - \varepsilon$ according to equations (20) and (21). Hence, the demand for money does not become infinity under the Friedman rule. This is in contrast to (but consistent with) Bewley's (1983) analysis because money does not earn interest in my model. Hence, a monetary equilibrium with positive prices always exists under the Friedman rule in my model. In this case, we have $c(i) = w$ for all i — that is, consumption is perfectly smoothed across all states under the Friedman rule.

However, since θ^* is bounded above by $\bar{\theta}$, there must exist a maximum rate of inflation π_{\max} under which the highest liquidity premium is given by $R(\bar{\theta}) = \varepsilon + \bar{\theta} = \frac{1 + \pi_{\max}}{\beta}$. At this maximum inflation rate $\pi_{\max} = \beta(\varepsilon + \bar{\theta}) - 1$, we have $D(\bar{\theta}) = 1$ and $H(\bar{\theta}) = 0$. That is, the optimal demand for real balances becomes zero: $\frac{M}{P} = 0$. When the cost of holding money is so high, agents opt not to use money as the store of value and the velocity becomes infinity: $V = \frac{D}{H} = \infty$. The steady-state velocity is an increasing function of inflation because money demand drops faster than consumption as the inflation tax rises: $\frac{\partial V}{\partial \theta^*} = \frac{(1-F)}{H^2} > 0$. This long-run implication is consistent with empirical data. For example, Chiu (2007) has found using cross-country data that countries with higher average inflation also tend to have significantly higher levels of velocity and argued that such an implication cannot be deduced from the Baumol-Tobin model with an exogenously segmented asset market.²⁴

²⁴Also see the empirical analysis in Lucas (2000).

Notice that positive consumption can always be supported in equilibrium without the use of money: No agents will hold money if they anticipate others do not. For example, consider the situation where the value of money is zero, $\frac{1}{P} = 0$. In this case, equation (12) is still valid with both sides of the equation equal zero, thus equation (26) is no longer necessary. Equation (18) implies that $H(\theta^*) = 0$ and $\theta^* = \bar{\theta}$, so that money demand is always zero for all households. Equation (17) implies that consumption is strictly positive because $D(\bar{\theta}) = 1 > 0$. Hence, as in the overlapping-generations model of Samuelson (1958), this model permits multiple equilibria: a monetary equilibrium when the inflation rate is sufficiently low and a nonmonetary equilibrium regardless of the inflation rate.

When money is no longer held as a store of value because of sufficiently high inflation ($\pi \geq \pi_{\max}$), we have $c(i) = [\varepsilon + \theta(i)]w$, so consumption is completely uninsured and unsmoothed. However, the average (or aggregate) consumption is still $C = w$, the same as that under the Friedman rule. This suggests the potential danger of measuring the welfare cost of inflation or the business cycle based on aggregate variables.

3 Matching Money Demand

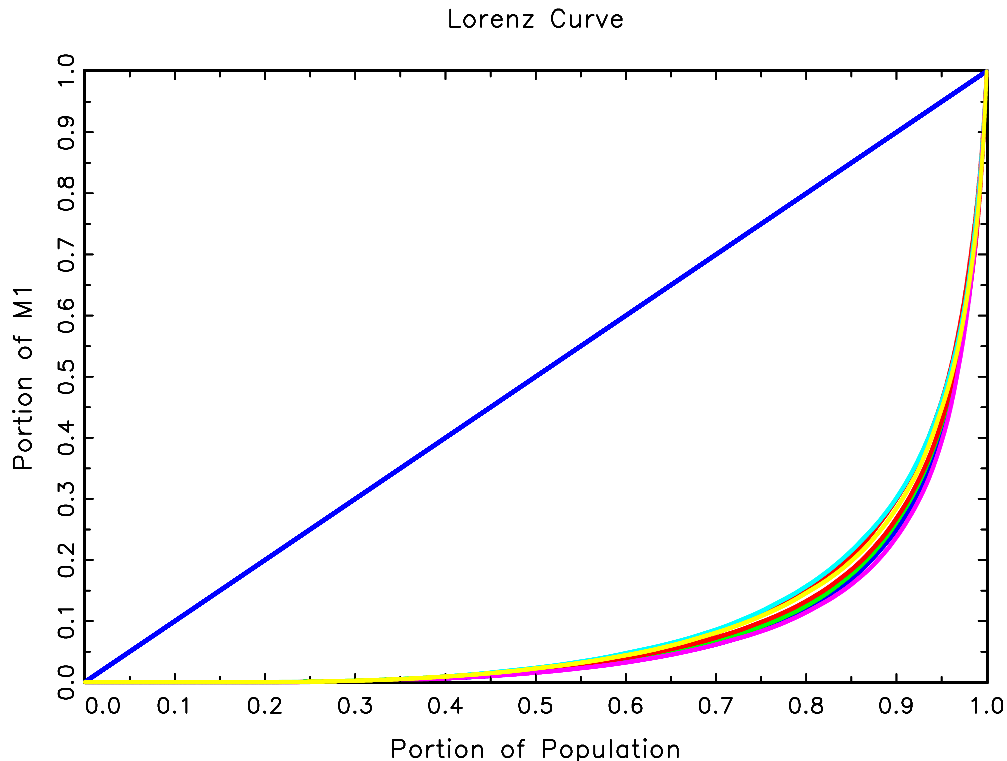


Figure 1. Distribution of Money Demand in the United States (1989-2007)

Figure 1 plots the Lorenz curve of money demand by American households based on the Survey of Consumer Finance for the years of 1989, 1992, 1995, 1998, 2001, 2004, and 2007.²⁵ Each survey covers about 4000 households and is conducted every three years. The Lorenz curve shows the portion of total aggregate money balances held by different fractions of the population. The 45 degree line on the diagonal indicates complete equality. The figure shows that the amount of money held by households is highly unequally distributed, even though the top 1% of the richest households have been excluded from the sample. The implied Gini coefficients for these different years are, respectively, 0.81, 0.84, 0.83, 0.81, 0.83, 0.84, and 0.82, with an average of 0.82.²⁶ The degree of inequality is extremely large and has not declined over the past 20 years. For example, in 2007, about 22% of the population holds essentially no money (less than or equal to \$10 in their checking accounts), 50% of the population holds less than 3% of the aggregate money balances, and the richest 10% of the population holds more than 75% of the total liquid assets.

To capture these features of the distribution of money demand, we assume that the idiosyncratic shocks $\theta(i)$ follow a generalized power distribution,

$$F(\theta) = \frac{(\theta + \varepsilon)^\sigma - \varepsilon^\sigma}{(\bar{\theta} + \varepsilon)^\sigma - \varepsilon^\sigma}, \quad (27)$$

with $\sigma > 0$ and the support $\theta \in [0, \bar{\theta}]$, where $\varepsilon \in [0, 1]$ is the wealth multiplier in equation (2).

The mean of this distribution is $E\theta = \frac{\sigma}{1+\sigma} \frac{(\bar{\theta} + \varepsilon)^{1+\sigma} - \varepsilon^{1+\sigma}}{(\bar{\theta} + \varepsilon)^\sigma - \varepsilon^\sigma} - \varepsilon$. The requirement, $E\theta = 1 - \varepsilon$, implies

$\frac{(\bar{\theta} + \varepsilon)^{1+\sigma} - \varepsilon^{1+\sigma}}{(\bar{\theta} + \varepsilon)^\sigma - \varepsilon^\sigma} = \frac{1+\sigma}{\sigma}$, which determines the upper bound $\bar{\theta}$ of the distribution. This distribution

includes the Uniform distribution as a special case when $\sigma = 1$, in which case we have $\bar{\theta} = 2(1 - \varepsilon)$

and $F(\theta) = \frac{\theta}{2(1-\varepsilon)}$. In the special case where $\varepsilon = 0$, we have $F(\theta) = \left(\frac{\theta}{\bar{\theta}}\right)^\sigma$ with $\bar{\theta} = \frac{1+\sigma}{\sigma}$. On the

other hand, when $\varepsilon = 1$, the distribution degenerates to the Dirac delta function with entire mass at $\theta = 0$, regardless the value of σ .²⁷ In this case, the model reduces to a representative-agent

model without idiosyncratic risks where the equilibrium value of money is $\frac{1}{P} = 0$, and wealth is

equally distributed across agents with Gini coefficient equal to 0. Hence, the distribution is quite general and it covers a variety of interesting cases by changing the parameter values of $\{\sigma, \varepsilon\}$.

With the generalized power distribution function, we have

$$R(\theta^*) = (\theta^* + \varepsilon) \frac{(\theta^* + \varepsilon)^\sigma - \varepsilon^\sigma}{(\bar{\theta} + \varepsilon)^\sigma - \varepsilon^\sigma} + \frac{\sigma}{1 + \sigma} \frac{(\bar{\theta} + \varepsilon)^{1+\sigma} - (\theta^* + \varepsilon)^{1+\sigma}}{(\bar{\theta} + \varepsilon)^\sigma - \varepsilon^\sigma} \quad (28)$$

²⁵Money demand is defined as cash and checking accounts, consistent with Ragot (2009).

²⁶The average Gini coefficient without excluding the top 1% richest individuals is 0.92.

²⁷The Dirac delta function is a degenerate distribution that has the value zero everywhere except at $\theta = 0$, where its value is infinitely large in such a way that its total integral is 1.

$$D(\theta^*) = 1 + \varepsilon + \theta^* - R(\theta^*) \quad (29)$$

$$H(\theta^*) = 1 - D(\theta^*). \quad (30)$$

The cutoff can be solved using the relation $R(\theta^*) = \frac{1+\pi}{\beta}$. The solution is unique because $R(\theta^*)$ is monotone in the interval $\theta^* \in [0, \bar{\theta}]$, $\frac{\partial R(\theta^*)}{\partial \theta^*} = F(\theta^*) > 0$. The Lorenz curves in the model can be computed as follows. At any level of wealth θ , the cumulative population density is $F(\theta)$, and the cumulative densities for wealth (F_w), money balances (F_m), and consumption (F_c) as a fraction of aggregate wealth, aggregate money balances, and aggregate consumption are given, respectively, by

$$F_w = \frac{1}{E(\varepsilon + \theta)} \int_0^\theta (\varepsilon + \theta) dF = \frac{\sigma}{1 + \sigma} \frac{(\varepsilon + \theta)^{1+\sigma} - \varepsilon^{1+\sigma}}{(\varepsilon + \bar{\theta})^\sigma - \varepsilon^\sigma} \quad (31)$$

$$\begin{aligned} F_m &= \frac{1}{H(\theta^*)} \int_0^\theta \max\{\theta - \theta^*, 0\} dF \\ &= \frac{1}{H(\theta^*)} \begin{cases} 0 & \text{if } \theta < \theta^* \\ \left[\frac{\sigma}{1+\sigma} \frac{(\varepsilon+\theta)^{1+\sigma} - (\varepsilon+\theta^*)^{1+\sigma}}{(\varepsilon+\bar{\theta})^\sigma - (\varepsilon)^\sigma} - (\varepsilon + \theta^*) \frac{(\varepsilon+\theta)^\sigma - (\varepsilon+\theta^*)^\sigma}{(\varepsilon+\bar{\theta})^\sigma - (\varepsilon)^\sigma} \right] & \text{if } \theta \geq \theta^* \end{cases} \end{aligned} \quad (32)$$

$$\begin{aligned} F_c &= \frac{1}{D(\theta^*)} \int_0^\theta \min\{\varepsilon + \theta, \varepsilon + \theta^*\} dF \\ &= \frac{1}{D(\theta^*)} \begin{cases} \left[\frac{\sigma}{1+\sigma} \frac{(\varepsilon+\theta)^{1+\sigma} - \varepsilon^{1+\sigma}}{(\varepsilon+\bar{\theta})^\sigma - \varepsilon^\sigma} \right] & \text{if } \theta < \theta^* \\ \left[\frac{\sigma}{1+\sigma} \frac{(\varepsilon+\theta^*)^{1+\sigma} - \varepsilon^{1+\sigma}}{(\varepsilon+\bar{\theta})^\sigma - \varepsilon^\sigma} \right] + \left[(\varepsilon + \theta^*) \frac{(\varepsilon+\theta)^\sigma - (\varepsilon+\theta^*)^\sigma}{(\varepsilon+\bar{\theta})^\sigma - \varepsilon^\sigma} \right] & \text{if } \theta \geq \theta^* \end{cases} \end{aligned} \quad (33)$$

Notice that these distributions must satisfy the relation, $D(\theta^*) F_c + H(\theta^*) F_m = E(\varepsilon + \theta) F_w$.

Suppose we set the time period to one year, $\beta = 0.96$, the annual inflation rate $\pi = 2\%$, and calibrate the parameters $\{\sigma, \varepsilon\}$ in the power distribution function such that the implied Gini coefficient of the distribution of money demand in the model is 0.8. These values are consistent with $\sigma = 0.01$ and $\varepsilon = 0.0015$.²⁸ Under these parameter values, the implied Gini coefficient for the distribution of net wealth, $[\varepsilon + \theta(i)] x$, is 0.76 and that for consumption is 0.39 in the model.²⁹ The consumption Gini for the United States is around 0.3, smaller than predicted by the model. This

²⁸There is no problem for the model to generate a higher Gini (say, above 0.9) by reducing the values of $\{\sigma, \varepsilon\}$ further.

²⁹Since the nonmonetary wealth $s(i)$ can only be determined explicitly at the aggregate level in the model, it is not possible to measure the distribution of gross household wealth, $\frac{m(i)}{P} + (1+r)s(i)$. This is why we report only the distribution of net wealth in the paper.

may be because households in the United States have other means to smooth consumption besides holding money.³⁰ However, these predictions are qualitatively consistent with the U.S. data — the distribution of money demand is far closer to that of wealth than to consumption. This is the consequence of holding money as an asset instead of a means of payment, so money serves mainly as a buffer stock to smooth consumption against wealth (income) shocks. The better money can serve as a store of value to smooth consumption, the closer is the distribution of money demand to that of wealth than to consumption. The predicted Lorenz curves are graphed in Figure 2.

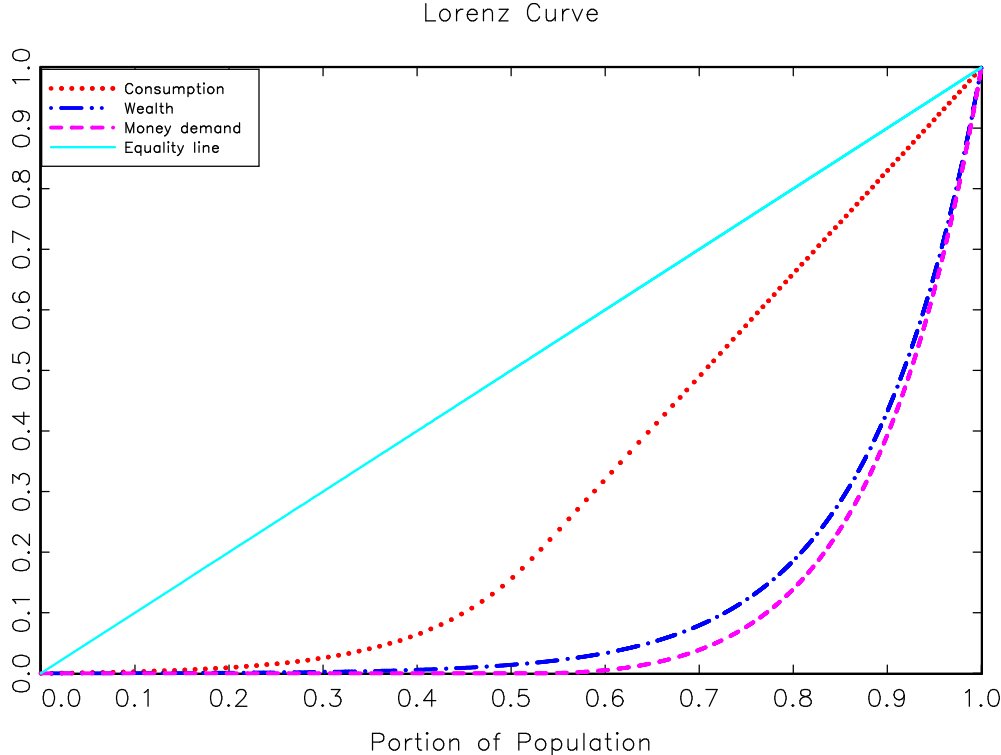


Figure 2. Predicted Distributions of Consumption, Money Demand, and Wealth

Figure 2 indicates that the liquidity-constrained (cash-poor) agents are less able to smooth consumption, so the consumption curve and the wealth curve are close to each other toward the left of the figure, whereas the cash-rich agents are better able to smooth consumption using money, so the consumption curve and wealth curve lie far apart toward the right of the figure. Hence, the Lorenz curve for consumption is not symmetric. In particular, the consumption level is constant across agents for the richest 45% of the population because they are not liquidity constrained; therefore, the consumption Lorenz curve becomes increasingly like a straight line toward the right side of the graph.

³⁰Ragot (2009) reports that for the United States, the financial-wealth Gini is around 0.8, the money demand Gini is around 0.83, and the consumption Gini is 0.28.

The calibrated model is also able to rationalize the empirical "money demand" curve estimated by Lucas (2000). Using historical data for GDP, money stock (M1), and the nominal interest rate, Lucas (2000) showed that the ratio of M1 to nominal GDP is downward sloping against the nominal interest rate. Lucas interpreted this downward relationship as a "money demand" curve and argued that it can be rationalized by the Sidrauski (1967) model of MIU. Lucas estimated that the empirical money demand curve can be best captured by a power function of the form

$$\frac{M}{PY} = Ar^{-\eta}, \quad (34)$$

where A is a scale parameter, r the nominal interest rate, and η the interest elasticity of money demand. He showed that $\eta = 0.5$ gives the best fit. Because the money demand defined by Lucas is identical to the inverted velocity, a downward-sloping money demand curve is the same as an upward-sloping velocity curve (namely, velocity is positively related to nominal interest rate or inflation). Similar to Lucas, the money demand curve implied by the benchmark model of this paper takes the form

$$\frac{M}{PY} = A \frac{H(\theta^*)}{D(\theta^*)}, \quad (35)$$

where A is a scale parameter, the functions $\{H, D\}$ are defined by equations (21) and (20), and the cutoff θ^* is a function of the nominal interest rate implied by equation (26). Figure 3 shows a close fit of the theoretical model to the U.S. data.³¹

The model is able to match the empirical aggregate money demand curve because velocity in the model is highly sensitive to inflation, especially near the Friedman rule. When the parameters of the model (especially ε and σ) are calibrated to match the empirical distribution of money demand, the implied velocity of money is close to zero near the Friedman rule but increases rapidly toward infinity as inflation rises. This is the consequence of optimal behaviors of the households: When the idiosyncratic risk is large (i.e., when both ε and σ are small), households opt to hold excessive amounts of liquidity as self-insurance against idiosyncratic shocks if the cost of doing so is small (i.e., near the Friedman rule). However, large cash reserves also hurt the households when inflation rises; so they deplete cash holdings rapidly when anticipated inflation increases, raising the portion of the liquidity-constrained population significantly. As a result, aggregate money demand declines and the velocity of money increases sharply. This also suggests that the marginal welfare cost of

³¹The circles in Figure 3 show plots of annual time series of a short-term nominal interest rate (the commercial paper rate) against the ratio of M1 to nominal GDP, for the United States for the period 1892–1997. The data are from the online Historical Statistics of the United States–Millennium Edition. The solid line with crosses is the model's prediction calibrated at annual frequency with $\beta = 0.96$ and $\delta = 0.1$. The other parameters remain the same: $\alpha = 0.3$, $\varepsilon = 0.0015$, and $\sigma = 0.02$. The nominal interest rate in the model is defined as $\frac{1+\pi}{\beta}$. The scale parameter is set to $A = 0.04$.

inflation may be extremely large near the Friedman rule because of the high inflation elasticities of money demand and velocity (to be shown below).

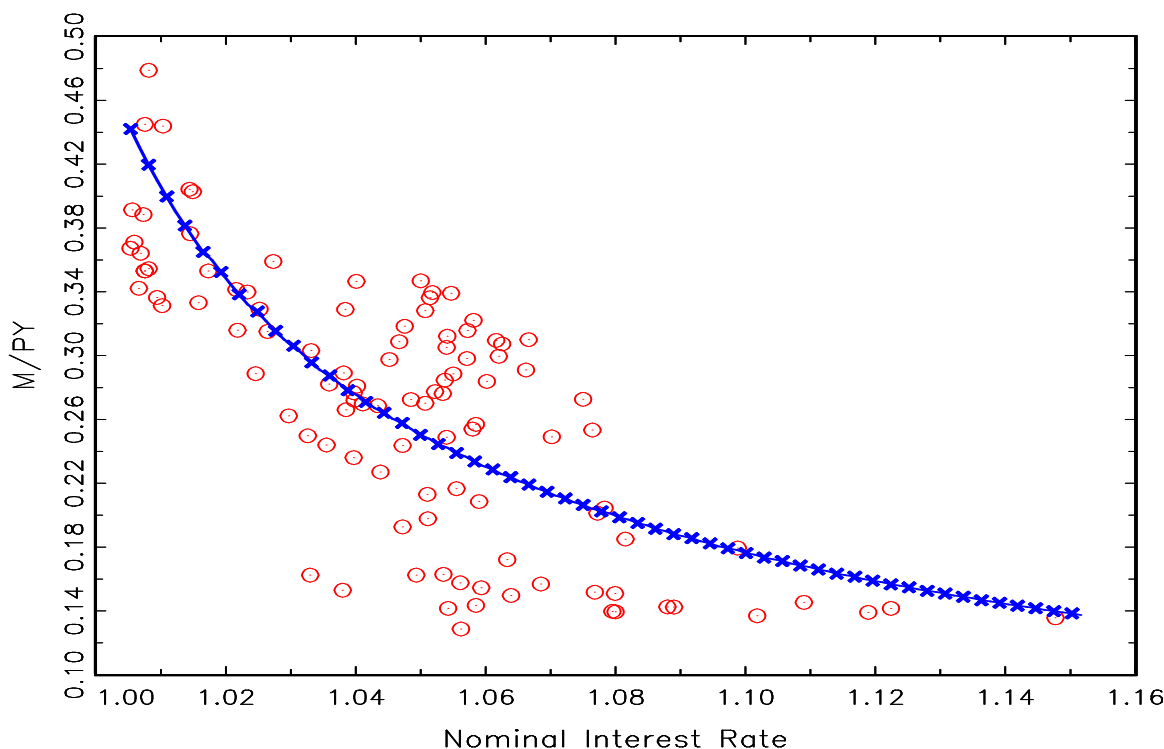


Figure 3. Aggregate Money Demand Curve in the Model (\times) and Data (o).

4 General-Equilibrium Analysis

The model of money demand outlined above can be easily embedded in a standard real business cycle (RBC) model for general-equilibrium analysis, as in Cooley and Hansen (1989). As pointed out by Cooley and Hansen (1989), a general-equilibrium analysis is essential for obtaining the correct measures of the welfare costs of inflation because it takes into account the trade-off between consumption and leisure under inflation tax.

Assume that capital is the only nonmonetary asset and is accumulated according to $K_{t+1} - (1 - \delta)K_t = I_t$, where I is gross aggregate investment with δ the rate of capital depreciation; the production technology is given by $Y_t = A_t K_t^\alpha N_t^{1-\alpha}$, where A_t denotes TFP. Under perfect competition and assuming that firms can rent capital from a competitive rental market, factor prices are determined by marginal products, $r_t + \delta = \alpha \frac{Y_t}{K_t}$ and $w_t = (1 - \alpha) \frac{Y_t}{N_t}$. The market clearing conditions for the capital market, labor market, and money market are given, respectively, by $S_t = K_{t+1}$, $\int n_t(i) = N_t$, and $M_t = \bar{M}_t = \bar{M}_{t-1} + \tau_t$, where \bar{M}_t denotes aggregate money supply

in period t . Notice that equations (17), (18), and (19) with money market clearing ($M = M_{-1} + \tau$) imply the aggregate goods-market clearing condition,

$$C_t + K_{t+1} - (1 - \delta)K_t = Y_t. \quad (36)$$

A general equilibrium is defined as the sequence $\{C_t, Y_t, N_t, K_{t+1}, M_t, P_t, w_t, r_t, \theta_t^*\}$, such that all households maximize utilities subject to their resource and borrowing constraints, firms maximize profits, all markets clear, the law of large numbers holds, and the set of standard transversality conditions is satisfied.³² The equations needed to solve for the general equilibrium are (9), (12), (17), (18), (36), the production function, firms' first-order conditions with respect to $\{K, N\}$, and the law of motion for money, $M = M_{-1} + \tau$. It is straightforward to confirm by the eigenvalue method that the aggregate model has a unique saddle path near the steady state, which is unique. Because the steady state is unique and the system is saddle stable, the distribution of money demand converges to a unique time-invariant distribution in the long run. The aggregate dynamics of the model can be solved by standard methods in the representative-agent RBC literature, such as log-linearizing the aggregate model around the steady state and then applying the method of Blanchard and Kahn (1980) to find the stationary equilibrium saddle path as in King, Plosser, and Rebelo (1988). This is the method we will use in the following analysis.

4.1 Steady-State Allocation

In the steady state, the capital-to-output and consumption-to-output ratios are given by $\frac{K}{Y} = \frac{\beta\alpha}{1-\beta(1-\delta)}$ and $\frac{C}{Y} = 1 - \frac{\delta\beta\alpha}{1-\beta(1-\delta)}$, respectively, which are the same as in standard representative-agent RBC models without money.³³ Since $r + \delta = \alpha\frac{Y}{K}$ and $w = (1 - \alpha)\frac{Y}{N}$, the factor prices are given by $r = \frac{1}{\beta} - 1$ and $w = (1 - \alpha)\left(\frac{\beta\alpha}{1-\beta(1-\delta)}\right)^{\frac{\alpha}{1-\alpha}}$, respectively. Hence, the existence of money in this model does not alter the steady-state saving rate, the great ratios, the utilization rate, the real wage, and the real interest rate in the neoclassical growth model, in contrast to standard CIA models. However, the levels of income, consumption, employment, the capital stock, and hence welfare will be affected by money supply. These levels are given recursively by

$$C = \frac{wR(\theta^*)}{\varepsilon + \theta^*}D(\theta^*), \quad Y = \frac{1 - \beta(1 - \delta)}{1 - \beta(1 - \delta) - \delta\beta\alpha}C, \quad K = \frac{\beta\alpha}{1 - \beta(1 - \delta)}Y, \quad N = \frac{1 - \alpha}{w}Y, \quad (37)$$

which are affected by monetary policy only through its influence on the distribution of money demand via the cutoff $\theta^*(\pi)$.

³²Such transversality conditions include $\lim_{t \rightarrow \infty} \beta^t \frac{K_{t+1}}{w_{t+1}} = 0$ and $\lim_{t \rightarrow \infty} \beta^t \frac{M_t}{P_{t+1}} = 0$, where $\frac{1}{w}$ is the shadow value of capital and $\frac{1}{P}$ is the value of money.

³³That is, the model achieves the modified golden rule.

4.2 Business Cycle Implications

For short-run business-cycle analysis, we follow the existing literature by calibrating the model to quarterly frequency. We set $\beta = 0.99$, $\delta = 0.025$, and $\alpha = 0.3$. We keep $\sigma = 0.02$ and $\varepsilon = 0.0015$ in accordance with Figure 2. The impulse responses of velocity and the aggregate price to a 1% transitory increase in the money stock under the first policy regime, $\frac{M_t - \bar{M}}{M} = \rho \frac{M_{t-1} - \bar{M}}{M} + \varepsilon_t$, where $\rho = 0.9$, are shown in Figure 4.

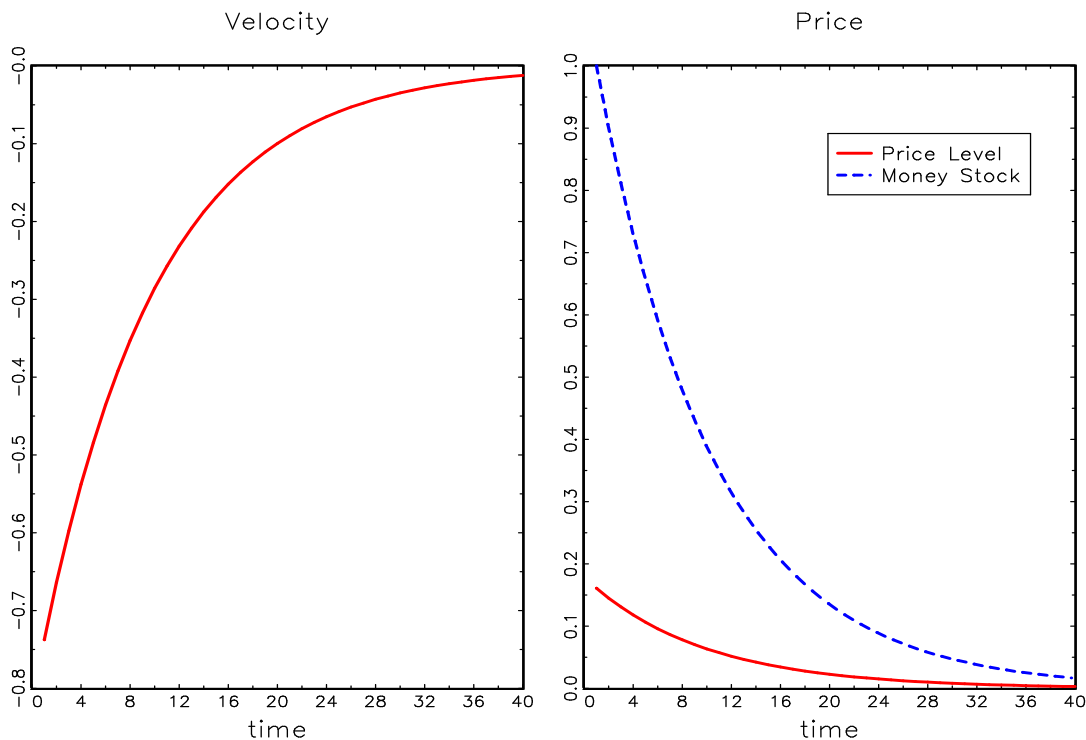


Figure 4. Impulse Responses to a 1% Money Injection.

Clearly, the velocity of money decreases nearly one-for-one with the money injection. The aggregate price level is "sticky" — it increases by less than 0.2 percent, far less than the 1 percent increase of money stock (see the right-hand window in Figure 1). Such a "sluggish" response of aggregate price to money is also noted by Alvarez, Atkeson, and Edmond (2008) in a Baumol-Tobin inventory-theoretic model of money demand. Thus, velocity and real money balance move in the opposite directions at the business-cycle frequency. This negative relationship is a stylized business-cycle fact documented by Alvarez, Atkeson, and Edmond (2008).

Because only a portion of the population is liquidity constrained and only the constrained agents will increase consumption when nominal income is higher, the aggregate price level will not rise proportionately to the monetary increase. Also, since the money injection is transitory (i.e., the aggregate money stock will return to its steady-state level in the long run), the expected inflation

rate, $E_t \frac{P_{t+1}}{P_t}$, falls and the real cost of holding money is lowered. This encourages all agents to increase money demand to reduce the probability of borrowing constraint. Consequently, aggregate real balances rise more than aggregate consumption and the measured velocity of money decreases. Associated with the sluggish response of the aggregate price are some real expansionary effects of money: In addition to the rise in aggregate consumption level, other variables such as aggregate output, labor supply, and capital investment also increase after the monetary shock. Because agents opt to maintain a target level of real wealth to provide just enough liquidity to balance the cost and benefit of holding money, labor supply must increase to replenish real income when the price level rises. If money injections are serially correlated, then investment will also increase to help sustain future wealth level by enhancing labor productivity.³⁴ These real expansionary effects of lump-sum injections benefit cash-poor agents and may thus improve social welfare, which is reminiscent of the redistributive effects of a lump-sum monetary injection discussed by Kehoe, Levine, and Woodford (1992).

An alternative way of understanding the movements in velocity and price is through equation (12), $1 = \left[\beta E_t \frac{P_t w_t}{P_{t+1} w_{t+1}} \right] R(\theta^*)$. Suppose the real wage is constant under a transitory money injection. As long as the expected inverse of inflation $E_t \frac{P_t}{P_{t+1}}$ rises, the rate of return to liquidity R must fall. Hence, the cutoff θ^* must decrease because $\frac{\partial R}{\partial \theta^*} > 0$. Since the function $\frac{D(\theta^*)}{H(\theta^*)}$ will decrease, velocity must fall, offsetting the impact of the money injection on aggregate price.³⁵

5 Welfare Costs of Inflation

Permanent changes in the money stock are no longer expansionary because of the anticipated permanently higher cost of holding money under rational expectations. When $E_t \frac{P_t}{P_{t+1}}$ declines under anticipated inflation, the arguments in the previous section on transitory monetary shocks are reversed. The rate of return to liquidity must increase to compensate for the cost of holding money. Hence, the cutoff θ^* must increase and demand for real balances must fall (recall $\frac{\partial H}{\partial \theta^*} < 0$). In particular, when the expected inflation rate is higher than the critical value π_{\max} , money will cease to be accepted as a store of value, optimal money demand goes to zero, and the velocity of money becomes infinity. Thus inflation has an "adverse liquidity effect" on the economy: It reduces the purchasing power of nominal balances, destroys the liquidity value of money, and therefore raises the number of liquidity-constrained agents. When liquidity constraint binds, the insurance function

³⁴The response of investment is negative if money injection is *i.i.d* because monetary shocks are like aggregate demand shocks, so they promote consumption and crowd out investment. But if such shocks have certain degree of persistence, agents will opt to save to enhance future consumption.

³⁵It can be shown that $\frac{\partial D(\theta^*)}{\partial \theta^*} = 1 - F(\theta^*) > 0$, $\frac{\partial H(\theta^*)}{\partial \theta^*} = F(\theta^*) > 0$, and $\frac{\partial}{\partial \theta^*} \left(\frac{D(\theta^*)}{H(\theta^*)} \right) = \frac{D'H - H'D}{H^2} = \frac{1}{H^2} \{ [1 - F]H - FD \} = \frac{1}{H^2} \{ H - \theta^* F \} < 0$.

of money disappears and consumption can no longer be smoothed against idiosyncratic shocks.

To put things in perspective, consider the household's optimal program in equation (1) without money — that is, money does not exist. In this case, agents have no insurance device to smooth consumption against idiosyncratic shocks because labor income and capital investment are both predetermined before the realizations of $\theta_t(i)$. However, it can be shown easily that the aggregate variables in this economy have identical equilibrium paths to a representative-agent counterpart economy where $\theta_t(i) = 1 - \varepsilon$ for all i and t . To see this, the first-order conditions of the household in the money-less economy are given by

$$\frac{1}{c_t(i)} = \lambda_t(i) \quad (38)$$

$$1 = w_t \tilde{E}_t \{ [\varepsilon + \theta_t(i)] \lambda_t(i) \} \quad (39)$$

$$\frac{1}{w_t} = \beta E_t (1 + r_{t+1}) \frac{1}{w_{t+1}} \quad (40)$$

$$c_t(i) = [\varepsilon + \theta_t(i)] [(1 + r_t) s_{t-1}(i) + w_t n_t(i) - s_t(i)], \quad (41)$$

where the last equation is the budget constraint. Since net wealth $[(1 + r_t) s_{t-1}(i) + w_t n_t(i) - s_t(i)]$ is chosen before observing $\theta_t(i)$, it is independent of $\theta_t(i)$. Hence, the budget identity (41) implies that the aggregate (or average) consumption is given by $C_t = (1 + r_t) S_{t-1} + w_t N_t - S_t$ (recall $E(\varepsilon + \theta) = 1$). Equation (39) then implies $\frac{1}{w_t} = \tilde{E}_t \frac{\varepsilon + \theta_t(i)}{c_t(i)} = \tilde{E}_t \frac{1}{x_t(i)} = \frac{1}{C_t}$. Thus, the aggregate allocation of this economy is identical to that without idiosyncratic uncertainty (i.e., when $\theta(i) = 1 - \varepsilon$ for all i , we have $\lambda_t(i) = \frac{1}{w_t}$ and $c_t(i) = w_t = C_t$). However, that these two economies are identical in aggregate allocation does not at all imply that the welfare is the same across the two economies because, with idiosyncratic uncertainty, agents are completely without consumption insurance in one economy, whereas they are completely free of idiosyncratic uncertainty in the other economy. This contrast helps to see the liquidity value of money and the implied welfare cost of inflation.

5.1 Measures of Welfare Costs

We measure the welfare costs of inflation as the state-independent percentage increase (λ) in consumption that would make each individual indifferent in terms of expected utilities between having a positive inflation and the Friedman rule, $\pi = \beta - 1$. By the law of large numbers, the aggregate (or average) utility of the population (with equal weights) is the same as the expected utility of an individual, namely, $\int \log c(i) di = \int \log c(i) dF$. Thus our welfare measure also corresponds to a

social planner's measure. That is, the value of λ is determined by the equation,

$$\int \log(1 + \lambda) c(i) di - \int n(i) di = \int \log \tilde{c}(i) di - \int \tilde{n}(i) di, \quad (42)$$

where $\tilde{c}(i)$ and $\tilde{n}(i)$ denote, respectively, the optimal consumption and labor supply of agent i under the Friedman rule. Notice that $\tilde{c}(i) = \tilde{C}$ is constant across agents under the Friedman rule. Solving for λ gives

$$\log(1 + \lambda) = \log \tilde{C} - \int \log c(i) dF + N - \tilde{N}, \quad (43)$$

where \tilde{N} is the aggregate labor supply under the Friedman rule.

Using the decision rules in equations (14) and (16), we have $c(i) = \min\{\varepsilon + \theta(i), \varepsilon + \theta^*\} \frac{wR(\theta^*)}{\varepsilon + \theta^*}$. Since aggregate consumption is given by $C = \frac{D(\theta^*)R(\theta^*)}{\varepsilon + \theta^*} w = \frac{DR}{\varepsilon + \theta^*} (1 - \alpha) \frac{Y}{N}$, we have $N = \frac{DR}{\varepsilon + \theta^*} (1 - \alpha) \frac{Y}{C}$, where the output-to-consumption ratio, $\frac{Y}{C} = \frac{1 - \beta(1 - \delta)}{1 - \beta(1 - \delta) - \delta\beta\alpha}$, is independent of inflation. Under the Friedman rule, we have $R(\theta^*) = 1$, $\theta^* = 0$, and $D(\theta^*) = \varepsilon$; hence, $\tilde{c}(i) = w$ and $\tilde{N} = (1 - \alpha) \frac{Y}{C}$. That is, consumers are perfectly insured under Friedman rule. Therefore, the above equation can be rewritten as

$$\log(1 + \lambda) = - \int \log[\min\{\varepsilon + \theta(i), \varepsilon + \theta^*\}] dF - \log \frac{R(\theta^*)}{\varepsilon + \theta^*} + \left(\frac{D(\theta^*)R(\theta^*)}{\varepsilon + \theta^*} - 1 \right) (1 - \alpha) \frac{Y}{C}, \quad (44)$$

where the last term is negative and reflects welfare gains due to rises in leisure in the face of positive inflation (Cooley and Hansen, 1989).

With the generalized power distribution and integrating by parts, the first term in equation (44) is given by

$$\begin{aligned} & \int_{\theta < \theta^*} \log(\varepsilon + \theta) dF + \int_{\theta \geq \theta^*} \log(\varepsilon + \theta^*) dF \\ &= F(\theta^*) \log(\varepsilon + \theta^*) - \int_{\theta < \theta^*} \left[\frac{F(\theta)}{\varepsilon + \theta} \right] d\theta + [1 - F(\theta^*)] \log(\varepsilon + \theta^*) \\ &= -\frac{1}{\sigma} F(\theta^*) + \log(\varepsilon + \theta^*) + \frac{\varepsilon^\sigma [\log(\varepsilon + \theta^*) - \log \varepsilon]}{(\varepsilon + \bar{\theta})^\sigma - \varepsilon^\sigma}. \end{aligned} \quad (45)$$

Thus, the welfare cost is

$$\log(1 + \lambda) = \frac{1}{\sigma} F(\theta^*) - \frac{\varepsilon^\sigma [\log(\varepsilon + \theta^*) - \log \varepsilon]}{(\varepsilon + \bar{\theta})^\sigma - \varepsilon^\sigma} - \log R(\theta^*) + \left(\frac{D(\theta^*)R(\theta^*)}{\varepsilon + \theta^*} - 1 \right) (1 - \alpha) \frac{Y}{C}. \quad (46)$$

Notice that as $\sigma \rightarrow 0$, the variance of the idiosyncratic shocks approaches infinity and the Gini coefficient of wealth distribution approaches 1, in which case the welfare cost of inflation also approaches infinity.

As a comparison, suppose we ignore idiosyncratic risk and (incorrectly) measure the welfare cost by using average consumption ($C = \int c(i)di$) as implied by the representative-agent literature, we would obtain

$$\log(1 + \lambda^o) = \log \tilde{C} - \log C + N - \tilde{N} \quad (47)$$

as the welfare cost. In this case, when $\pi \geq \pi_{\max}$, we have $C = \tilde{C} = w$, $N = \tilde{N} = (1 - \alpha) \frac{Y}{C}$, and $Y = \tilde{Y}$. That is, the above incorrect measure of welfare cost would be zero under hyper-inflation. This shows that, even if the aggregate economy under hyper-inflation may look identical to that under the Friedman rule, the actual welfare can differ dramatically between the two economies. This difference in welfare arises precisely because inflation has destroyed the liquidity value of money and made agents unable to self-insure against idiosyncratic shocks. It is this aspect of the welfare costs of inflation that is missed by the Bailey triangle.

On the other hand, as $\sigma \rightarrow \infty$, the distribution of $\theta(i)$ becomes degenerate with the entire mass locating at $\theta(i) = \theta^* = \bar{\theta} = 1 - \varepsilon$, so $R(\theta^*) = 1$, $D(\theta^*) = 1$, $H(\theta^*) = 0$, and $\log(1 + \lambda) = 0$ according to equation (46). That is, without idiosyncratic risk, there is no welfare gain by holding money; hence, the welfare cost of inflation is zero under either measures.

5.2 Calibration

Since our model features money as the only financial asset providing self-insurance against idiosyncratic shocks, it is more suitable for describing developing countries than for developed countries because households in developed countries can smooth consumption also by borrowing or trading nonmonetary assets in spot markets. The data in Table 1 (panel A) show that, in developing countries, the share of cash and bank deposits accounts for more than 90% of total financial wealth, whereas this number is much smaller in developed countries. Hence, we first calibrate the model based on developing-country data. One problem of this approach is that data for distribution of money demand in developing countries are rarely available in the literature. As an alternative, we calibrate the model to match the Gini coefficients of consumption or health care expenditures in developing countries based on data provided in Makinen et al. (2007). This is a reasonable alternative because at the end it is the idiosyncratic movements in consumption that matter for welfare.

The bottom panel (panel B) in Table 1 shows the Gini coefficients for consumption expenditure and health care expenditure in several developing countries. The average consumption Gini across those countries is 0.43 and the average health care Gini is 0.4; these values are both significantly

larger than the consumption Gini (0.28) in the United States. However, considering that cross-household dispersion of consumption expenditure may overstate idiosyncratic risk for individuals over time, to be conservative we assume the variance of household consumption over time is only one third of that (dispersion) across households in developing countries; that is, we calibrate the model to generate a consumption Gini of 0.133, instead of $0.4 \sim 0.43$.³⁶

We set the parameters $\{\beta, \delta\}$ to quarterly frequency with $\beta = 0.99$ and $\delta = 0.025$, so the annual inflation rate is 4π . If $\pi = 0.5\%$, a combination of $\sigma = 0.025$ and $\varepsilon = 0.02$ would yield a consumption Gini of 0.133 in the model. We will use these parameter values as a benchmark to assess the welfare costs of inflation.

Table 1. Household Portfolio and Expenditure Inequality*

A. Composition of Household Financial Wealth							
	France	Germany	Italy	Japan	Spain	UK	US
Liquid assets	33%	46%	40%	59%	40%	21%	15%
Shares & equities	32%	24%	39%	34%	38%	25%	39%
Other	35%	30%	21%	7%	22%	57%	46%
	China	India					
Liquid assets	90%	92%					
Shares & equities	6%	5%					
Other	4%	3%					
B. Expenditure Inequality for Developing Countries							
	Burkina Faso	Guatemala	Kazakhstan	Kyrgyzstan	Paraguay	South Africa	Thailand
Consumption Gini	0.43	0.39	0.37	0.45	0.47	0.54	0.39
	Burkina Faso	Guatemala	Kazakhstan	South Africa	Paraguay	Zambia	Thailand
Health care Gini	0.43	0.42	NA	0.67	0.18	0.32	0.38

*Data source: For composition of financial wealth, see Davies et al. (2006, Tables 3 and 4). Data for China are based on Yi and Song (2008). For the consumption and health care expenditure, see Makinen et al. (2007).

The welfare implications of inflation are graphed in Figure 5. The upper-left panel in Figure 5 shows the correct measure of welfare cost (λ). It is monotonically increasing with inflation. Hence, the Friedman rule is clearly optimal.³⁷ The maximum welfare cost is reached at the maximum

³⁶This assumption seems reasonable in light of the findings of Guvenen and Smith (2009), who estimate that in the United States the measured income risk of individual households is about one third of the cross-household dispersion in income.

³⁷This result is in sharp contrast to Aiyagari (2005) where he shows that taxing the rate of return to capital is optimal in an incomplete-market economy with uninsurable risk. In contrast, here it is shown that taxing the rate of return to money is not optimal. The intuition for the difference is as follows. When agents use fixed capital as a liquid asset to buffer idiosyncratic shocks, precautionary saving motives lead to over-accumulation of aggregate capital in equilibrium. As a result, the marginal product of (i.e., rate of return to) capital is too low compared

inflation rate $\pi_{\max} = 531\%$, where $\lambda = 1.86$. Beyond this point agents stop holding money as a store of value, so the welfare cost remains constant at 1.86 for $\pi \geq \pi_{\max}$. The upper-right panel shows the incorrect measure of welfare cost based on average consumption (λ^o). This measure is not monotonic; it equals zero at two extreme points: the point of the Friedman rule and the point where $\pi = \pi_{\max}$. In the first case, individual consumption level is the highest because there is no cost to holding money, so agents are perfectly insured against idiosyncratic risk. In the latter case, individual consumption becomes homogeneous again and is back to the maximum level when money is no longer held as a store of value. Without money, inflation no longer has any adverse liquidity effects on consumption, so λ^o remains at zero beyond π_{\max} .³⁸

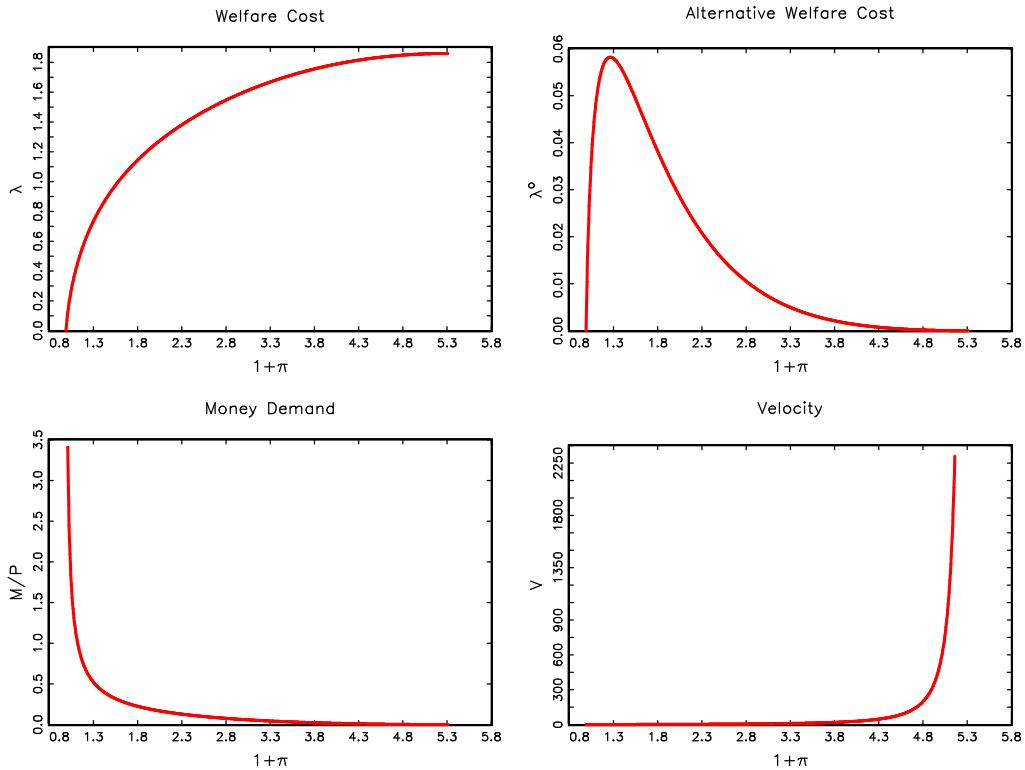


Figure 5. Welfare Costs, Money Demand, and Velocity

The bottom-left panel shows the level of aggregate money demand ($\frac{M}{P}$), which monotonically decreases with inflation. At the maximum inflation rate $\pi_{\max} = 531\%$, the demand for real balances

with the modified golden rule. Hence, taxing the rate of return to capital will lower the aggregate capital stock and raise the equilibrium interest rate to the golden-rule level. The higher interest rate will generate a positive income effect on consumption that is more than enough to compensate for the loss in an individual's welfare due to the reduction in the buffer stock. A key here is that individuals take the real interest rate as externally given but their collective actions determine the level of the real interest rate in equilibrium. In my monetary model, the situation is different because the shadow rate of return to money ($R(\theta^*)$) is determined by the probability of liquidity constraint, which is determined internally by an agent's optimal choice of money holdings. Also, money does not affect capital accumulation in my model. Hence, taxing the rate of return to money through inflation will reduce welfare because there are no externalities in the rate of return to money as in the case of Aiyagari (1995).

³⁸When $\pi \geq \pi_{\max}$, we have $R(\bar{\theta}) = \varepsilon + \bar{\theta}$, $D(\bar{\theta}) = 1$, and $H(\bar{\theta}) = 0$, so $C = \frac{DR}{\varepsilon + \bar{\theta}} w = w$.

becomes zero, and the velocity of money becomes infinity (bottom-right panel).³⁹ The velocity of money is the lowest under the Friedman rule, because people opt to hoard as much money as they can when its rate of return equals the inverse of the time discounting factor. These implications for money demand and velocity are very different from standard CIA models, which imply an upper bound of unity on velocity and a strictly positive lower bound on money demand, because agents under the CIA constraint must hold money even with an infinite rate of inflation. In the real world, people often stop accepting domestic currency as the means of payment when the inflation rate is too high, consistent with the model.

Consider the welfare cost at the inflation rate $\pi_{\max} = 531\%$, where $\lambda = 1.86$: Although each individual's average consumption is the same as that under the Friedman rule, the social welfare differs dramatically. When money is no longer held because inflation is too high to bear, individuals face completely uninsurable idiosyncratic uncertainty across states of nature, and they are willing to reduce consumption by more than 65% to eliminate such uncertainty. In other words, people's consumption or income level must be more than 286% larger (since $1 + \lambda = 2.86$) in order to accept a 531% quarterly inflation rate.

Even at moderate inflation rates, the welfare cost is astonishingly large. For example, when $\pi = 1.5\%$ (i.e., 10 percentage points above the Friedman rule in terms of annualized inflation rate), individuals' consumption level must increase by 14.5% to make them indifferent; and when $\pi = 2.5\%$ (i.e., a 10% annual inflation rate above zero), consumption must be increase by 12% to make individuals indifferent. Since most of the welfare loss happens near the Friedman rule, we can also follow Lucas (2000) by computing the value of λ when annualized inflation increases from 4% to 14%. The result is $\lambda = 10\%$. Remember, most of the welfare losses come from the cash-poor households.

On the other hand, if we use the incorrect measure λ_0 , then the welfare cost of 10% annualized inflation is only 2% of GDP compared with zero inflation, and it is only 1.3% of GDP when inflation rises from 4% to 14% a year. These numbers are similar (in the order of magnitude) to those estimated by Lucas (2000). This indicates again that it is highly misleading to use aggregate consumption or representative-agent models to measure the welfare costs of inflation.

In a heterogeneous-agent economy with incomplete markets, the larger the variance of idiosyncratic shocks, the stronger the precautionary motive for holding money. This raises the interest elasticity of money demand because agents with larger nominal balances incur a disproportionately higher inflation tax. More importantly, higher inflation shifts the mass of the distribution of money demand toward zero balances by reducing cash holdings across agents, resulting in a larger portion of the population without self-insurance against idiosyncratic shocks. This shift of the distribution

³⁹The graph only shows velocity for $\pi < \pi_{\max}$.

of money demand in response to inflation is most critical in generating the large welfare cost. To show the importance of this extensive margin, Figure 6 graphs the relationship between inflation and the probability of a binding liquidity constraint (or the proportion of the liquidity-constrained population). It shows that, as inflation rises, the portion of the population holding zero balances increases rapidly. For example, when annual inflation increases from 2% to 10%, an additional 9% of the entire population is left without cash (thus without self-insurance), raising the total number of liquidity-constrained agents to about 37% of the population. When holding money is so costly, the demand for real balances is so low so that the probability of a binding liquidity constraint is very high. This amounts to large welfare losses.

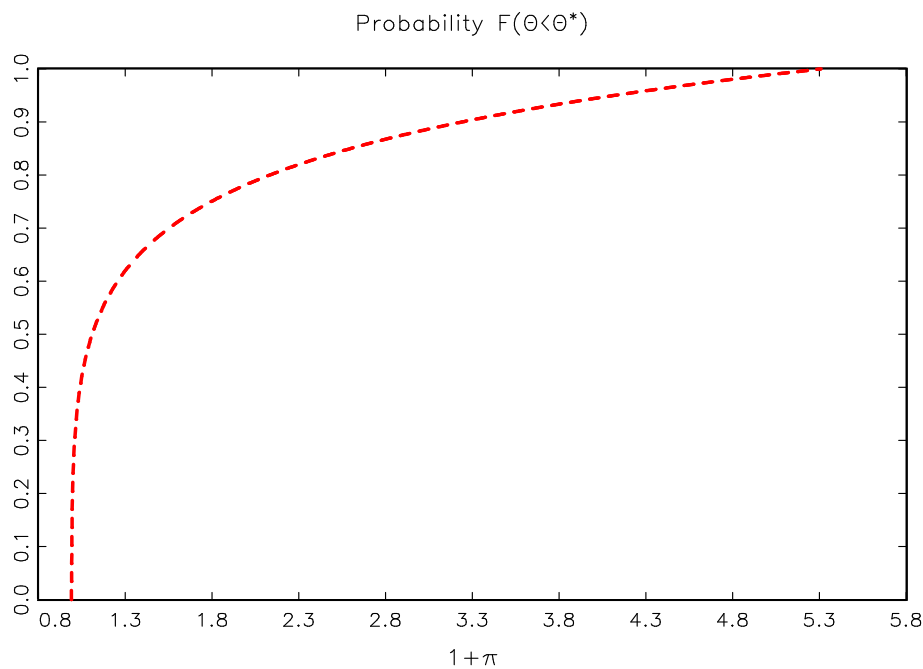


Figure 6. Portion of Borrowing-Constrained Population

5.3 Alternative Calibration

Since money demand data are available in the United States, an alternative calibration is to choose the parameters $\{\sigma, \varepsilon\}$ so that the model matches exactly the fraction of liquidity-constrained population in the United States. According to the Survey of Consumer Finances, the portion of households having zero balances in checking accounts is 19.3% of the population based on surveys in the years between 1989 and 2007 (with standard deviation of 1.3%), the portion of households having less than \$10 in checking accounts is 20% (with standard deviation of 1.4%), and that having

less than \$20 is 20.6% (with standard deviation 1.3%).⁴⁰ Households with little balances in their checking accounts also tend to have very little balances in other types of accounts, such as saving accounts and money-market accounts. Hence, if we define a household with less than \$10 in checking accounts as liquidity constrained, the data suggest that 20% of the U.S. population is liquidity constrained. This estimate is consistent with other independent empirical studies. For example, Hall and Mishkin (1982) use the Panel Study of Income Dynamics and find that 20% of American families are liquidity constrained. Mariger (1986) uses a life-cycle model to estimate this fraction to be 19.4%. Hubbard and Judd (1986) simulate a model with a constraint on net worth and find that about 19.0% of United States consumers are liquidity constrained. Jappelli (1990) uses information on individuals whose request for credit has been rejected by financial intermediaries and estimates through a Tobit model that 19.0% of families are liquidity constrained. Therefore, the emerging consensus points to a fraction of approximately 20% of the population to be liquidity constrained.

To generate a 20% fraction of liquidity-constrained population in the model under the parameter value of $\beta = 0.99$ and the annual inflation rate of 2%, it requires $\sigma = 0.27$ and $\varepsilon = 0.05$. These parameter values imply that the Gini of money demand in the model is only 0.57 and that of consumption is only 0.08, which is $\frac{2}{7}$ of the actual consumption Gini in the United States and looks a reasonable measure of the individual consumption risk across time for the average U.S. households. With these parameter values, the welfare cost of 10% annualized inflation is 8.2% of consumption.

6 Conclusion and Remarks for Further Research

This paper provides a tractable general-equilibrium model of money demand. The model is designed to isolate the store-of-value function of money from its other functions and is applied to study the dynamics of velocity, the monetary business cycle, and especially the welfare costs of inflation in an environment where financial markets are incomplete and interest-bearing assets are either unavailable or highly illiquid for households to smooth consumption against idiosyncratic shocks that are not fully insurable. The most important findings include (i) the aggregate velocity of money can be extremely volatile and is countercyclical to money supply in the short run but procyclical to inflation in the long run, as in the data; (ii) transitory monetary shocks have positive real effects despite their lump-sum nature and flexible prices; and (iii) the welfare costs of inflation can be astonishingly high — at least 8% of consumption under 10% annual inflation.

Two simplifying strategies allow the generalized Bewley (1980) model to be analytically tractable. First, the idiosyncratic shocks come from wealth rather than from labor income. Second and more

⁴⁰On the other hand, the portion of households with balances greater than \$3,000 in checking accounts is larger than 32% (with standard deviation of 2.2%).

importantly, the utility function is linear in leisure. These simplifying strategies make the expected marginal utility of an individual's consumption and the cutoff value for target wealth independent of idiosyncratic shocks and individual histories. With these properties, closed-form decision rules for individuals' consumption and money demand can be derived explicitly and aggregation becomes easy. After aggregation by the law of large numbers, the variables form a system of nonlinear dynamic equations as in a representative-agent RBC model. The stability of this nonlinear system can be checked by traditional eigenvalue method. Hence, traditional solution methods available in the RBC literature can be applied to solving the model's equilibrium saddle path, given the distribution of the idiosyncratic shocks. The impulse response functions to aggregate shocks and the second (or higher) moments of the model can then be computed in a tractable way.

These simplifying strategies have some costs. First, the *i.i.d.* assumption of idiosyncratic shocks rules out any persistence in the distribution of wealth and money demand. However, adopting the approach in Wang and Wen (2010) with discrete distribution may overcome this problem. Another cost is that the elasticity of labor supply is not a free parameter. This imposes some limits on the model's ability to study labor supply behavior and labor market dynamics. Nonetheless, the payoff of the simplifying assumptions is significant: They not only make the model analytically tractable with closed-form solutions for individuals' decision rules, but also reduce the computational costs to the level of solving a representative-agent RBC model despite nontrivial distribution of idiosyncratic shocks and endogenous production with capital. Because of this, the model may prove very useful in applied work.

However, there may be at least two potential objections to the large welfare cost of inflation found in this paper. First, the model features uninsurable idiosyncratic risk and money is the only liquid asset to help self-insure against such risk. This setup rules out other types of insurance devices and, especially, does not take into account the role of banking in mitigating the idiosyncratic risk through borrowing and lending. Second, it is a common belief of the existing literature that inflation benefits debtors by redistributing the burden of inflation toward creditors. For these reasons, the welfare costs of inflation may be overstated in the paper.

To address some of these concerns, Wen (2009a) extends the general-equilibrium Bewley model to a setting with "narrow banking," where cash-rich agents can deposit their idle cash into a bank, and cash-constrained agents can borrow from the bank by paying nominal interest. My analysis therein shows that (i) financial intermediation can dramatically improve social welfare; (ii) the expansionary real effects of transitory money injections can be greatly magnified by the credit channel of money supply through financial intermediation; (iii) however, the welfare cost of inflation remains astonishingly high despite the possibility of borrowing and lending (i.e., risk sharing) at market-determined interest rates for credit. The intuition is as follows. Financial intermediation amplifies the real effects of transitory monetary shocks because the injected liquidity

can be reallocated from cash-rich agents to cash-poor agents through the banking system via borrowing and lending. This generates a significant and persistent liquidity effect on the nominal and real interest rates of loans. However, permanent money growth is still highly costly.⁴¹ Even though financial intermediation promotes risk sharing and alleviates liquidity constraints, it cannot completely eliminate idiosyncratic risk because of positive nominal interest rate. That is, the banking system cannot undo the adverse liquidity effects of inflation, not only because inflation reduces money demand (thus diminishing the insurance function of money) but also because it increases the costs of borrowing — the nominal interest of loans rises with inflation more than one-for-one.⁴² Hence, inflation always generates an "adverse liquidity effect" on the economy regardless of financial intermediation, and it is precisely the higher interest costs of loans that may make debtors worse off (instead of better off) in the face of positive inflation (offsetting the redistributive effects).

This paper does not consider the welfare costs of inflation in situations where money is required as a medium of exchange in addition to being a store of value. Wen (2009b) studies a general-equilibrium version of the Lucas (1980) model where money is held both as a store of value and as a means of payment. There I show that the welfare costs of inflation, especially hyperinflation, are larger than those in this paper (other things equal), because in CIA models, agents are not able to get rid of money even when holding cash becomes extremely costly.

This research provides only a rough measure of the welfare costs of inflation by taking into account the precautionary motives of money demand. Although this is an important first step toward more accurate measures of the welfare costs of inflation, an obvious shortcoming of the current model is that it implies that the cross-household dispersion of money demand is the same as its cross-time variation. Hence, if the model is calibrated to match the Gini coefficients of cross-household distributions, it tends to overstate the over-time variation; and when it is calibrated to match the over-time variation, it tends to understate the cross-household dispersion. Since both margins are important for properly assessing the heterogeneous liquidity effects of inflation, a critical line of future research requires (i) better panel data for consumption and money demand, especially for developing countries, so that both the cross-household dispersion and the cross-time variation of money demand can be estimated; and (ii) a richer model that is able to match both the across-section and the cross-time properties of the panel data. Currently, panel data with a long enough time dimension is extremely difficult to find, especially for developing countries.⁴³ As

⁴¹That is, given inflation rate, financial intermediation improves welfare significantly; but given financial intermediation, inflation is still very costly.

⁴²Wen (2009a) shows that the nominal interest rate of credit in the money market can differ fundamentally from the nominal interest rate of nonmonetary assets with incomplete financial markets.

⁴³The Panel Study of Income Dynamics (PSID) may be one of the best panel data source available to date, but even this data archive does not track exactly the same households for their income and consumption over a sufficiently long period of time, let alone information on money demand.

panel and longitudinal data with better qualities develop, the estimated welfare costs of inflation in this paper can be further refined.

References

- [1] Aiyagari, R., 1994, Uninsured idiosyncratic risk and aggregate saving, *Quarterly Journal of Economics* 109(3), 659-684.
- [2] Aiyagari, R. and M. Gertler, 1991, Asset returns with transactions costs and uninsured individual risk, *Journal of Monetary Economics* 27(3), 311-332.
- [3] Aiyagari, S.R. and S.D. Williamson, 2000, Money and dynamic credit arrangements under private information, *Journal of Economic Theory* 91(2), 248-279.
- [4] Akyol, A., 2004, Optimal monetary policy in an economy with incomplete markets and idiosyncratic risk, *Journal of Monetary Economics* 51(6), 1245-1269.
- [5] Alvarez, F., A. Atkeson and C. Edmond, 2008, Sluggish responses of prices and inflation to monetary shocks in an inventory model of money demand, *Quarterly Journal of Economics* 124(3), 911-967.
- [6] Alvarez, F., A. Atkeson, and P. Kehoe, 1999, Money and interest rates with endogenously segmented markets, NBER Working Paper w7060.
- [7] Andolfatto, D. and E. Nosal, 2003. A theory of money and banking. Federal Reserve Bank of Cleveland, Working Paper 03-10.
- [8] Attanasio, O., L. Guiso, and T. Jappelli, 2002, The demand for money, financial innovation, and the welfare cost of inflation: An analysis with Household Data, *Journal of Political Economy* 110(2), 317-351.
- [9] Bai, Jinhui, 2005, Stationary monetary equilibrium in a Baumol-Tobin exchange economy: Theory and computation, Georgetown University Working Paper.
- [10] Bailey, M., 1956, The welfare cost of inflationary finance, *Journal of Political Economy* 64, 93-110.
- [11] Baumol, W. J., 1952, The transactions demand for cash: An inventory theoretic approach, *Quarterly Journal of Economics* 66(4), 545-556.
- [12] Bewley, T., 1980, The optimum quantity of money, in *Models of Monetary Economics*, ed. by John Kareken and Neil Wallace. Minneapolis, Minnesota: Federal Reserve Bank, 1980.
- [13] Bewley, T., 1983, A difficulty with the optimum quantity of money, *Econometrica* 51(5), 1485-1504.

- [14] Blanchard, O. and C. Kahn, 1980, The solution of linear difference models under rational expectations, *Econometrica* 48(5), 1305-1311.
- [15] Cartwright, P., C. Delorme, and N. Wood, 1985, The by-product theory of revolution: Some empirical evidence, *Public Choice* 46, 265-274.
- [16] Chatterjee, S. and D. Corbae, 1992, Endogenous market participation and the general equilibrium value of money, *The Journal of Political Economy* 100(3), 615-646.
- [17] Chiu, J., 2007, Endogenously segmented asset market in an Inventory theoretic model of money demand, Bank of Canada Working Paper 2007-46.
- [18] Cooley, T. and G. Hansen, 1989, The inflation tax in a real business cycle model, *American Economic Review* 79(4), 733-748.
- [19] Davies, J., S. Sandstrom, A. Shorrocks, and E. Wolpin, 2006, The World Distribution of Household Wealth, University of Western Ontario Working Paper.
- [20] Dotsey, M. and P. Ireland, 1996, The welfare cost of inflation in general equilibrium, *Journal of Monetary Economics* 37, 29-47.
- [21] Guvenen, F. and A. Smith, 2009, Inferring labor income risk from economic choices: An indirect inference approach, Working Paper.
- [22] Gross, D. and N. Souleles, 2002, Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data, *The Quarterly Journal of Economics* 117 (1), 149-185.
- [23] Hall, R. and F. Mishkin, 1982, The sensitivity of consumption to transitory income: Estimates from panel data on households, *Econometrica* 50, 461-81.
- [24] Heathcote, J., K. Storesletten and G. Violante, 2008, Quantitative Macroeconomics with Heterogeneous Households, Federal Reserve Bank of Minneapolis Research Department Staff Report September 2008.
- [25] Henriksen, E. and F. Kydland, 2010, Endogenous money, inflation, and welfare, *Review of Economic Dynamics* 13, 470-486.
- [26] Hubbard, R. and K. Judd, 1986, Liquidity constraints, fiscal policy, and consumption, *Brookings Papers on Economic Activity* 1, 1-50.
- [27] Jaffee, D. M., and T. Russell, "Imperfect

- [28] Huggett, M., 1993, The risk-free rate in heterogeneous-agent incomplete-insurance economies, *Journal of Economic Dynamics and Control* 17(5-6), 953-969.
- [29] Huggett, M., 1997, The one-sector growth model with idiosyncratic shocks: Steady states and dynamics, *Journal of Monetary Economics* 39(3), 385-403.
- [30] Imrohoroglu, A., 1989, Cost of business cycles with indivisibilities and liquidity constraints, *Journal of Political Economy* 97(6), 1364-1383.
- [31] Imrohoroglu, A., 1992, The welfare cost of inflation under imperfect insurance, *Journal of Economic Dynamics and Control* 16(1), 79-91.
- [32] Jappelli, T., 1990, Who is credit constrained in the U. S. economy? *Quarterly Journal of Economics* 105(1), 219-234.
- [33] Kehoe, T., D. Levine, and M. Woodford, 1992, The optimum quantity of money revisited, in *Economic Analysis of Markets and Games*, ed. by P. Dasgupta, D. Gale, O. Hart, and E. Maskin, Cambridge, Mass.: M.I.T. Press.
- [34] Khan, A. and J. Thomas, 2007, Inflation and interest rates with endogenous market segmentation, Federal Reserve Bank of Philadelphia working paper 07-1.
- [35] King, R., C. Plosser, and S. Rebelo, 1988, Production, growth, and business cycles: I. The basic neoclassical model, *Journal of Monetary Economics*, 21(2-3), 195-232.
- [36] Krusell, P. and A. Smith, 1998, Income and wealth heterogeneity in the macroeconomy, *Journal of Political Economy* 106(5), 867-96.
- [37] Lagos, R. and R. Wright, 2005, A unified framework for monetary theory and policy analysis, *Journal of Political Economy* 113(3), 463-484.
- [38] Looney, R., 1985, The Inflationary process in prerevolutionary Iran, *The Journal of Developing Areas* 19, 329-350.
- [39] Lucas, R.E. Jr., 1980, Equilibrium in a pure currency economy, *Economic Inquiry* 18(2), 203-220.
- [40] Lucas, R.E. Jr., 2000, Inflation and welfare, *Econometrica* 68(2), 247-274.
- [41] Makinen, M. et al., 2007, Inequalities in health care use and expenditures: Empirical data from eight developing countries, *Bulletin of the World Health Organization* 78, 55-65.

- [42] Mariger, R., 1986, *Consumption Behavior and the Effects of Fiscal Policies*, Cambridge: Harvard University Press.
- [43] Ragot, X., 2009, The case for a financial approach to money demand, Working Paper, Banque de France.
- [44] Samuelson, P., 1958. An exact consumption-loan model of interest with or without the social contrivance of money, *Journal of Political Economy* 66(6), 467-482.
- [45] Sidrauski, M., 1967, Rational choice and patterns of growth in a monetary economy, *American Economic Review* 57(2), 534-544.
- [46] Svensson, L., 1985, Money and asset prices in a cash-in-advance economy, *Journal of Political Economy* 93(5), 919-944.
- [47] Tobin, J., 1956, The interest-elasticity of the transactions demand for cash, *Review of Economics and Statistics* 38(3), 241-247.
- [48] Townsend, R., 1995, Consumption insurance: An evaluation of risk-bearing systems in low-income economies, *Journal of Economic Perspectives* 9(3), 83-102.
- [49] Wang, P.F. and Y. Wen, 2009, Inventory Accelerator in General Equilibrium, Federal Reserve Bank of St. Louis Working Paper 2009-010A.
- [50] Wen, Y., 2008, Input and output inventory dynamics, Federal Reserve Bank of St. Louis Working Paper 2008-008B.
- [51] Wen, Y., 2009a, Liquidity and welfare in a heterogeneous-agent economy, Federal Reserve Bank of St. Louis Working Paper 2009-019B.
- [52] Wen, Y., 2009b, When does heterogeneity matter? Federal Reserve Bank of St. Louis Working Paper 2009-024A.
- [53] Wolff, E., 1998, Recent Trends in the Size Distribution of Household Wealth, *The Journal of Economic Perspectives* 12(3), 131-150.
- [54] Yi, G. and W. Song, 2008, The evolution of China's financial assets structure: 1991-2007, *Economic Research* 2008(8), 4-15 (in Chinese).