Liquidity and Welfare in a Heterogeneous-Agent Economy

Authors	Yi Wen
Working Paper Number	2009-019B
Revision Date	May 2009
Citable Link	https://doi.org/10.20955/wp.2009.019
Suggested Citation	Wen, Y., 2009; Liquidity and Welfare in a Heterogeneous-Agent Economy, Federal Reserve Bank of St. Louis Working Paper 2009-019. URL https://doi.org/10.20955/wp.2009.019

Liquidity and Welfare in a Heterogeneous-Agent Economy*

Yi Wen
Federal Reserve Bank of St. Louis
& Tsinghua University (Beijing)

This Version: April 29, 2009

Abstract

This paper reconsiders the welfare costs of inflation and the welfare gains from financial intermediation in a heterogeneous-agent economy where money is held as a store of value (as in Bewley, 1980). Because of heterogeneous liquidity demand, transitory lump-sum money injections can have persistent expansionary effects despite flexible prices, and such effects can be greatly amplified by the banking system through the credit channel. However, permanent money growth can be extremely costly: With log utility functions, consumers are willing to reduce consumption by 15% (or more) to avoid a 10% annual inflation. For the same reason, financial intermediation can significantly improve welfare: The welfare costs of a collapse of the banking system is estimated as about $10-68\%$ of aggregate output. These welfare implications differ dramatically from those of the existing literature.

Keywords: Interest Rate, Liquidity Preference, Liquidity Trap, Banking, Money Demand, Velocity, Welfare Costs of Inflation.

JEL codes: E12, E13, E31, E32, E41, E43, E51.

*I thank Pengfei Wang for discussions on issues related to this project, Qing Liu and Steve Williamson for comments, and Luke Shimek for research assistance. The usual disclaimer applies. Correspondence: Yi Wen, Research Department, Federal Reserve Bank of St. Louis, St. Louis, MO, 63104. Phone: 314-444-8559. Fax: 314-444-8731. Email: yi.wen@stls.frb.org.

1 Introduction

This paper reconsiders the business-cycle and welfare effects of money by generalizing Bewley's (1980) precautionary money demand model to a dynamic stochastic general equilibrium framework. The model may serve as an alternative to the heterogeneous-agent cash-in-advance (CIA) model of Lucas (1980) and the (S,s) inventory-theoretic model of Baumol (1952) and Tobin (1956). The key feature distinguishing Bewley's model from the literature is that money is held solely as a store of value, completely symmetric to any other asset, and is not imposed from outside nor required as the means of payments. Agents can choose whether to hold money, depending on the costs and benefits.

By freeing money from its role of medium of exchange, Bewley's approach allows one to focus on the role of money as a pure form of liquidity so that the liquidity-preference theory of money can be investigated more thoroughly in isolation. Beyond Bewley (1980), my generalized model is analytically tractable; hence, it greatly simplifies the computation of dynamic stochastic general equilibrium in environments with heterogeneous agents and capital accumulation, which facilitates general-equilibrium analyses of banking and credit through the use of money.[1] The model is applied to studying some important issues in monetary literature, including (i) the business-cycle dynamics of velocity, (ii) the welfare costs of inflation, (iii) the determination of the nominal interest rate in the money market, and (iv) the welfare costs of a collapse of the banking system.

The major findings of the paper include: (i) The model is able to produce enough variability in velocity relative to GDP 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 significant positive effects on aggregate activities despite flexible prices; and such real effects can be greatly magnified by the credit channel of money supply through financial intermediation. (iii) The nominal interest rate of loans is fundamentally different from the rates of returns to non-monetary assets (such as capital) and does not obey the Fisherian relationship. (iv) The welfare cost of moderate inflation can be around 15% of consumption, which is several orders larger than that estimated by Lucas (2000). (v) Financial intermediation improves welfare and the gains are in the range of $10-68\%$ of aggregate output under moderate inflation, suggesting that active government policies to prevent a banking collapse are desirable in the case of a financial crisis.

The key property of the model is an endogenously determined distribution of money holdings

[1] The analytical tractability of my model is achieved by assuming quasi-linear preferences as in Lagos and Wright (2005).

under heterogeneous liquidity preferences. Hence, lump-sum money injections have an immediate impact on consumption for liquidity-constrained agents but not for agents with idle cash balances. Consequently, transitory monetary shocks are expansionary (even without open market operations), the velocity of money is countercyclical, and aggregate price appears "sticky". Financial intermediation amplifies these real effects because the injected liquidity can be reallocated from cash-rich agents to cash-poor agents through borrowing and lending in the banking system. This generates a significant and persistent liquidity effect on the nominal and real interest rates of loans. With anticipated inflation, permanent money growth reduces welfare significantly because of three reasons: (i) Agents with large money holdings suffer disproportionately more from inflation tax than agents with less real balances. (ii) Anticipated inflation reduces money demand; hence, the fraction of the population with a binding liquidity constraint increases with inflation. This generates additional welfare costs along the extensive margin. (iii) Agents opt to switch from "cash" goods (consumption) to "credit" goods (leisure), thereby reducing labor supply and aggregate output.[2]

This paper is closely related to the work of Alvarez, Atkeson, and Edmond (2008). Both papers are based on an inventory-theoretic approach and can explain the short-run dynamic behavior of velocity and aggregate prices under 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). Second, the distribution of money holdings is exogenously given in their model; hence, the fraction of population with the need for cash withdrawals is fixed and cannot respond to monetary policy. 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.[3]

Bewley's model has been studied extensively in the literature. But the main body of this literature focuses on endowment economy. For example, Imrohoroglu (1992) and Akyol (2004) study the welfare costs of inflation in the Bewley model. Like Bewley (1980), their models are based on an endowment economy without capital accumulation and are not analytically tractable. More importantly, these authors do not address some of the issues considered in this paper, such as the dynamics of velocity, the monetary business cycle, the liquidity trap, the determination of the nominal interest rate in the credit market, and the welfare gains of financial intermediation. Also, my model captures much larger welfare costs of inflation than implied by this literature.

In what follows, Section 2 presents a benchmark model without banking. It reveals some of the basic properties of a monetary model based on liquidity preference or the precautionary motive.

[2] Money facilitates consumption by providing liquidity. However, consuming leisure does not require liquidity.

[3] For Baumol-Tobin models with endogenously segmented markets, see, Alvarez, Atkeson and Kehoe (1999), Chiu (2007), Khan and Thomas (2007), and King and Thomas (2007), among others.

Section 3 extends the benchmark model to study narrow banking. The nominal interest rate of loans is determined and the existence of liquidity traps is proven. Section 4 concludes the paper with remarks for future research.

2 The Benchmark Model

2.1 Individual Households

The benchmark model is a stochastic general-equilibrium version of Bewley (1980, 1983).[4] Although money is dominated in the expected rate of return, it is more liquid than non-monetary assets as stores of value. Hence, by providing liquidity to facilitate consumption demand, money can coexist with interest-bearing assets.[5]

There is a continuum of households indexed by $i \in [0, 1]$. As in Lucas (1980), each household is subject to an idiosyncratic preference shock to the marginal utility of consumption, $\theta(i)$, which has the distribution $F(\theta) \equiv \Pr[\theta(i) \leq \theta]$ with support $[\theta_l, \theta_h]$. Leisure enters the utility function linearly as in Lagos and Wright (2005).[6] A household chooses consumption $c(i)$, labor supply $n(i)$, a non-monetary asset $s(i)$ that pays the real rate of return $r > 0$, and nominal balance $m(i)$ to maximize lifetime utility.

To capture the liquidity role of money, assume that the decisions for labor supply and investment on interest-bearing assets (such as capital) must be made before observing the idiosyncratic preference shock $\theta(i)$ in each period. Thus, if there is an urge to consume in period t, money stock is the asset that can be adjusted most quickly to buffer the random preference shock.[7] Borrowing of liquidity (money) from other households is not allowed in the basic model.[8] These assumptions imply that households may find it optimal to carry money as inventories to cope with demand uncertainty, even though money is not essential for exchange. As in the standard literature, any aggregate uncertainty is resolved at the beginning of each period and is orthogonal to idiosyncratic

[4] In 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.

[5] Liquidity in this paper is defined as the easiness to exchange for goods. Keynes wrote that an asset is more liquid than another "if it is more certainly realisable at short notice without loss" (Keynes, 1930, Vol. II, p. 67).

[6] The linearity assumption simplifies the model by making the distribution of wealth degenerate. However, unlike Lagos and Wright (2005), the distribution of money holdings in my model is not degenerate. This setup also makes the results regarding the welfare costs of inflation comparable to the CIA model of Cooley and Hansen (1989) based on indivisible labor.

[7] This timing friction is what we need to generate a positive liquidity value of money over other assets in equilibrium. It is a realistic friction in the sense that people may need to make consumption decisions on daily or hourly bases due to biological needs, but do not have to make investment and working decisions as frequently. This type of timing friction is also assumed by Aiyagari and Williamson (2000) and Akyol (2004) in endowment economies with random income shocks. The timing friction is also akin to the transaction costs approach of Aiyagari and Gertler (1991), Chatterjee and Corbae (1992), and Greenwood and Williamson (1989).

[8] To make money completely symmetric to interest-bearing assets, we can also impose borrowing constraints on non-monetary assets. However, as long as the discounted real rate of return for such assets exceeds one, borrowing constraints do not bind for these assets in the steady state. Hence, such constraints are ignored.

uncertainty. Household i solves[9]

$$\max E_0 \sum_{t=0}^{\beta} \beta^t \left\{\theta(i) \log c(i) - an(i)\right\}$$

subject to

$$c_t(i) + \frac{m_t(i)}{P_t} + s_t(i) = \frac{m_{t-1}(i) + \tau_t}{P_t} + (1 + r_t)s_{t-1}(i) + w_t n_t(i) \tag{1}$$

$$m_t(i) \geq 0, \tag{2}$$

where P denotes aggregate price, w the real wage, and τ a lump-sum per-capita nominal transfer. Without loss of generality, assume $a = 1$. Denoting $\{\lambda(i), \pi(i)\}$ as the Lagrangian multipliers for constraints (1) and (2), respectively, the first-order conditions with respect to $\{c(i), n(i), s(i), m(i)\}$ are given, respectively, by

$$\frac{\theta(i)}{c(i)} = \lambda(i) \tag{3}$$

$$1 = w_t E_t^i \lambda(i) \tag{4}$$

$$E_t^i \lambda_t(i) = \beta E_t (1 + r_{t+1}) \lambda_{t+1}(i) \tag{5}$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t \frac{\lambda_{t+1}(i)}{P_{t+1}} + \pi_t(i), \tag{6}$$

where the expectation operator E^i denotes expectations conditional on the information set of time t excluding $\theta_t(i)$. Hence, equations (4) and (5) reflect the fact that labor supply $n_t(i)$ and asset investment $s_t(i)$ must be made before the idiosyncratic taste shocks (and hence the value of $\lambda_t(i)$) are realized. By the law of iterated expectations and the orthogonality assumption of aggregate and idiosyncratic shocks, equation (5) and (6) can be written as

$$\frac{1}{w_t} = \beta E_t \left(1 + r_{t+1}\right) \frac{1}{w_{t+1}} \tag{7}$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t \frac{1}{P_{t+1} w_{t+1}} + \pi_t(i), \tag{8}$$

where $\frac{1}{w}$ is the marginal utility of consumption in terms of labor.

The decision rule for an individual's money demand is characterized by a cutoff strategy, taking as given the aggregate environment. Consider two possible cases:

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

Case A. $\theta_t(i) \leq \theta_t^*$. In this case the urge to consume is low. It is hence optimal to hoard money as inventories so as to prevent possible liquidity constraints in the future. So $m_t(i) \geq 0$, $\pi_t(i) = 0$ and the shadow value of good $\lambda_t(i) = \beta E_t \frac{P_t}{w_{t+1}P_{t+1}}$. Equation (3) implies that consumption is given by $c(i) = \theta(i)\left[\beta E_t \frac{P_t}{w_{t+1}P_{t+1}}\right]^{-1}$. Defining

$$x(i) \equiv \frac{m_{t-1}(i) + \tau_t}{P_t} + (1 + r_t)s_{t-1}(i) - s_t(i) + wn_t(i) \tag{9}$$

as real wealth net of asset investment, the budget identity (1) then implies $\frac{m_t(i)}{P_t} = x_t(i) - \theta(i)\left[\beta E_t \frac{P_t}{w_{t+1}P_{t+1}}\right]^{-1}$. The requirement $m_t(i) \geq 0$ then implies

$$\theta(i) \leq \left[\beta E_t \frac{P_t}{w_{t+1}P_{t+1}}\right] x_t(i) \equiv \theta_t^*, \tag{10}$$

which defines the cutoff θ^*. Notice that the cutoff is independent of i because wealth $x(i)$ is determined before the realization of $\theta_t(i)$ and all households face the same distribution of idiosyncratic shocks. This property simplifies the computation of the general equilibrium of the model tremendously.

Case B. $\theta_t(i) > \theta_t^*$. In this case the urge to consume is high. It is then optimal to spend all money in hand, so $\pi_t(i) > 0$ and $m_t(i) = 0$. By the resource constraint (1), we have $c_t(i) = x_t(i)$, which by equation (10) implies $c(i) = \theta_t^* \left[\beta E_t \frac{P_t}{w_{t+1}P_{t+1}}\right]^{-1}$. Equation (3) then implies that the shadow value is given by $\lambda_t(i) = \frac{\theta_t(i)}{\theta_t^*}\left[\beta E_t \frac{P_t}{w_{t+1}P_{t+1}}\right]$. Since $\theta(i) > \theta^*$, equation (8) implies $\pi_t(i) > 0$. Notice that the shadow value of goods, $\lambda(i)$, is higher under case B than under case A.

The above analyses imply that the expected shadow value of goods, $E^i\lambda(i)$, and hence the optimal cutoff value θ^*, is determined by the following asset-pricing equation for money based on (4):

$$\frac{1}{P_t w_t} = \left[\beta E_t \frac{1}{P_{t+1}w_{t+1}}\right] R(\theta_t^*), \tag{11}$$

where

$$R(\theta_t^*) \equiv \left[\int_{\theta(i)\leq\theta^*} dF(\theta) + \int_{\theta(i)>\theta^*} \frac{\theta(i)}{\theta^*} dF(\theta)\right] \tag{12}$$

measures the (shadow) rate of return to liquidity. The left-hand side of equation (11) is the utility cost of holding one unit of real balances as inventory. The right-hand side is the expected gain by holding money, which takes two possible values. The first is simply the discounted next-period utility cost of inventory ($\beta E_t \left[P_{t+1}w_{t+1}\right]^{-1}$) in the case of low demand ($\theta \leq \theta^*$), which has probability $\int_{\theta(i)\leq\theta^*} dF(\theta)$. The second is the marginal utility of consumption ($\frac{\theta_t(i)}{\theta_t^*}\beta E_t \left[P_{t+1}w_{t+1}\right]^{-1}$) in the case of high demand ($\theta > \theta^*$), 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 investment in money (liquidity) is determined by $R(\theta^*)$. Notice that $R(\theta^*) > 1$ as long as $\theta^* < \theta_h$ and that aggregate shocks will affect the distribution of money holdings across households by affecting the cutoff θ_t^*. The fact $R > 1$ implies that the option value of one dollar exceeds one because it provides liquidity in the case of the urge to consume. The optimal level of cash reserve (money demand) is always such that the probability of stockout (being liquidity constrained) is strictly positive unless the cost of holding money is zero. This inventory-theoretic formula of the rate of return to liquidity is derived by Wen (2008) in an inventory model based on the stockout-avoidance motive.

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, 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 non-monetary 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. However, as will become clear in the next section, this shadow asset-rate of return to money is not the same as the equilibrium interest rate in the credit market.

The cutoff strategy implies that the optimal level of wealth in period t is determined by a "target" given by $x_t(i) = \theta_t^* \left[\beta E_t \frac{P_t}{w_{t+1}P_{t+1}}\right]^{-1}$, which specifies that wealth (real money balances plus labor income net of asset investment) is set to a target level depending on the cutoff (distribution of demand shocks) and the expected future utility. This implies that labor supply will adjust so that the wealth level meets its target. Utilizing equation (11), the decision rules of household i are summarized by

$$c_t(i) = w_t R(\theta_t^*) \min\{\theta(i), \theta_t^*\}, \tag{13}$$

$$\frac{m_t(i)}{P_t} = w_t R(\theta_t^*) \max\{\theta_t^* - \theta(i), 0\}, \tag{14}$$

$$x_t(i) = w_t R(\theta_t^*)\theta_t^*. \tag{15}$$

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 and by the law of large numbers, the aggregate variables are given by

$$C_t = w_t R(\theta_t^*) D(\theta_t^*), \tag{16}$$

$$\frac{M_t}{P_t} = w_t R(\theta_t^*) H(\theta_t^*), \tag{17}$$

$$\frac{M_{t-1} + \tau_t}{P_t} + (1 + r_t) S_{t-1} - S_t + w N_t = w_t R(\theta_t^*) \theta^*, \tag{18}$$

where

$$D(\theta^*) \equiv \int_{\theta(i) \leq \theta^*} \theta(i) dF(\theta) + \int_{\theta(i) > \theta^*} \theta^* dF(\theta), \tag{19}$$

$$H(\theta^*) \equiv \int_{\theta(i) \leq \theta^*} [\theta^* - \theta(i)] \, dF(\theta), \tag{20}$$

and these two functions satisfy $D(\theta^*) + H(\theta^*) = \theta^*$.

Monetary Policy. We consider two types (regimes) of monetary policies. In the short-run dynamic analysis, money supply shocks are purely transitory without affecting the steady-state stock of money,

$$\tau_t = \rho \tau_{t-1} + \bar{M} \varepsilon_t, \tag{21}$$

$$M_t = \bar{M} + \tau_t, \tag{22}$$

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, $\pi = 0$.

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

$$\tau_t = (\mu_t - 1) M_{t-1} \tag{23}$$

$$\log \mu_t = \rho \log \mu_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(\bar{\varepsilon}, \sigma^2), \tag{24}$$

where μ_t is the gross growth rate of money with mean $\bar{\mu}$ and $\bar{\varepsilon} = (1 - \rho) \log \bar{\mu}$ is the mean of the innovation ε.

The Quantity Theory. The aggregate relationship between consumption (16) and money demand (17) implies the "Quantity" equation,

$$P_t C_t = M_t V_t, \tag{25}$$

where $V_t \equiv \frac{D(\theta_t^*)}{H(\theta_t^*)}$ measures the aggregate consumption-velocity of money.[10] A high velocity implies a low demand for real balances relative to consumption. Given the support of θ as $[\theta_l, \theta_h]$ and the mean as $E\theta$, by the definition for the functions D and H, it is easy to see that the domain of velocity is $\left[\frac{E\theta}{\theta_h - E\theta}, \infty\right]$, which has no finite upper bound, in sharp contrast to CIA models. An infinite velocity means that either the value of money $(\frac{1}{P})$ is zero or nominal money demand (M) is zero. On the contrary, a zero velocity (in the case $\theta_h = \infty$) implies the demand for real balances is infinity (because consumption demand can never be zero given our utility function).

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

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

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 $\frac{P_t}{P_{t-1}} = \frac{M_t}{M_{t-1}}$ in the steady state, so the steady-state inflation rate is the same as the growth rate of money.

Since by (26) the return to liquidity R must increase with π, the cutoff θ^* must decrease with π (because $\frac{\partial R(\theta^*)}{\partial \theta^*} < 0$). What this says is that when inflation rises, the required rate of return to liquidity must also increase accordingly in order to induce people to hold money. However, because the cost of holding money increases with π, agents opt to hold less money so that the probability of stockout $(1 - F(\theta^*))$ rises, which forces a rise in the equilibrium shadow rate of return. By definitions (19) and (20), we have $\frac{\partial D}{\partial \theta^*} = 1 - F(\theta^*) > 0$ and $\frac{\partial H}{\partial \theta^*} = F(\theta^*) > 0$, so both functions decrease with π. Therefore, a high rate of inflation has two offsetting effects on aggregate consumption and money demand. On the one hand, they both increase because of a higher equilibrium rate of return to liquidity R; on the other hand, they both decrease because fewer households choose to hold money when the cost of doing so increases. Since the second effect dominates, inflation is welfare reducing (see below for general-equilibrium analysis on welfare issues).

Under the Friedman rule, $1 + \pi = \beta$, we have $R = 1$ and $\theta^* = \theta_h$ according to (12), and $D(\theta^*) = E(\theta)$ and $H(\theta^*) = \theta_h - E(\theta)$ according to (19) and (20). Hence, as long as θ_h is finite, the demand for money does not become infinity under the Friedman rule. This is consistent with Bewley's (1983) analysis because money does not earn interest in my model. Hence, a monetary equilibrium with positive prices always exists around the Friedman rule in my model if the support of the idiosyncratic shocks is bounded above.

[10] Alternatively, 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, since θ^* is bounded below by θ_l, there must exist a maximum rate of inflation π_h such that the highest rate of return to liquidity is given by $R(\theta_l) = \frac{1+\pi_h}{\beta}$. At this inflation rate π_h, we have $D(\theta_l) = \theta_l$ and $H(\theta_l) = 0$. That is, the optimal demand for real balances becomes zero: $\frac{M}{P} = wRH = 0$. 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: $\frac{\partial V}{\partial \theta^*} = \frac{1}{H^2}\{H - \theta^* F\} < 0$. 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.[11] Such an implication cannot be deduced from the Baumol-Tobin model with exogenously segmented asset market (see, e.g., Chiu, 2007).

Notice that positive consumption can always be supported in equilibrium without the use of money. This is so because 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 (11) is valid. Equation (17) implies that $H(\theta^*) = 0$ and $\theta^* = \theta_l$, so that money demand is zero. Equation (16) implies that consumption is strictly positive because $D(\theta_l) = \theta_l > 0$.[12]

2.3 General Equilibrium Analysis

The model of money demand outlined above can be easily embedded into a standard real business cycle (RBC) model. For example, assume that capital is the only non-monetary asset and is accumulated according to $K_{t+1} - (1-\delta_t) K_t = I_t$, where I is gross aggregate investment and δ_t a time-varying rate of depreciation; the production technology is given by $Y_t = A_t(e_t K_t)^\alpha N_t^{1-\alpha}$, where A denotes TFP, e the capacity utilization rate, which is related to the capital depreciation rate according to the relation proposed by Greenwood, Hercowitz, and Huffman (1988), $\delta_t = \frac{1}{1+\omega} e_t^{1+\omega}$.[13] Under perfect competition, factor prices are determined by marginal products, $r_t + \delta_t = \alpha \frac{Y_t}{K_t}$ and $w_t = (1-\alpha)\frac{Y_t}{N_t}$. Optimal capacity utilization implies $\alpha \frac{Y_t}{e_t} = e_t^\omega K_t$. Market clearing implies $S_t = K_{t+1}$, $\int n_t(i) = N_t$, 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 (16), (17), and (18) with money market clearing ($M = M_{-1} + \tau$) implies the aggregate goods-market clearing condition,

$$C_t + K_{t+1} - (1-\delta)K_t = Y_t. \tag{27}$$

[11] Also see Liu, Wang, and Wright (2008) and Lucas (2000). Based on U.S. time-series data, Lucas shows that the inverse of the velocity is negatively related to inflation.

[12] I thank Pengfei Wang for pointing this out to me.

[13] Capacity utilization amplifies the responses of output to shocks.

A general equilibrium is defined as the sequence $\{C_t, Y_t, N_t, e_t, \delta_t, K_{t+1}, M_t, P_t, w_t, r_t, \theta_t^*\}$, such that all households maximize utility 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.[14] The equations needed to solve for the general equilibrium are (7), (11), (16), (17), (27), the production function, the capital depreciation function, firms' first-order conditions with respect to $\{e, K, N\}$, and the law of motion for money, $M = M_{-1} + \tau$. The aggregate model has a unique steady state. The aggregate dynamics of the model can be solved by log-linearizing the aggregate model around the steady state and then applying the method of Blanchard and Kahn (1980) to find the stationary saddle path as in King, Plosser, and Rebelo (1988).

Steady-State Allocation. In the steady state, the capital-output and consumption-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 RBC models without money. 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, and the real factor prices in the neoclassical growth model, in contrast to CIA models. However, the levels of income, consumption, employment, and capital stock will be affected by money. These levels are given by

$$C = wR(\theta^*)D(\theta^*), \qquad Y = \frac{1-\beta(1-\delta)}{1-\beta(1-\delta)-\delta\beta\alpha}C, \qquad K = \frac{\beta\alpha}{1-\beta(1-\delta)}Y, \qquad N = \frac{1-\alpha}{w}Y. \tag{28}$$

Calibration and Impulse Responses. To facilitate quantitative analysis, we assume the idiosyncratic shocks $\theta(i)$ follow the Pareto distribution, $F(\theta) = 1 - \theta^{-\sigma}$, with $\sigma > 1$ and the support $\theta \in (1, \infty)$. Since the support is not bounded above, monetary equilibrium with a strictly positive price level $P > 0$ does not exist under the Friedman rule. Hence, our analysis in this part of the paper treats the Friedman rule as a limiting case.[15] With the Pareto distribution, we have

$$R(\theta_t^*) = 1 + \frac{1}{\sigma-1}\theta^{*-\sigma}, \tag{29}$$

$$D(\theta^*) = \frac{\sigma}{\sigma-1} - \frac{1}{\sigma-1}\theta^{*1-\sigma}, \tag{30}$$

$$H(\theta^*) = \theta^* - \frac{\sigma}{\sigma-1} + \frac{1}{\sigma-1}\theta^{*1-\sigma}. \tag{31}$$

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

[15] With the Pareto distribution, as $1+\pi$ approaches β, the demand for real balances approaches infinity. Since in equilibrium money demand must equal money supply (which is finite), this implies that the price level must approach zero (or the value of money must approach infinity).

Following the standard RBC literature, we set the time period to a quarter of a year, and $\beta = 0.99$, $\omega = 0.4$ (implying $\delta = 0.025$), and $\alpha = 0.3$. We choose a degree of heterogeneity by setting the shape parameter $\sigma = 1.5$.[16] The impulse responses of the model to a 1% transitory increase in the money stock under the first policy regime, $\frac{M_t - \bar{M}}{\bar{M}} = \rho \frac{M_{t-1} - \bar{M}}{\bar{M}} + \varepsilon_t$, where $\rho = 0.9$, are shown in Figure 1.

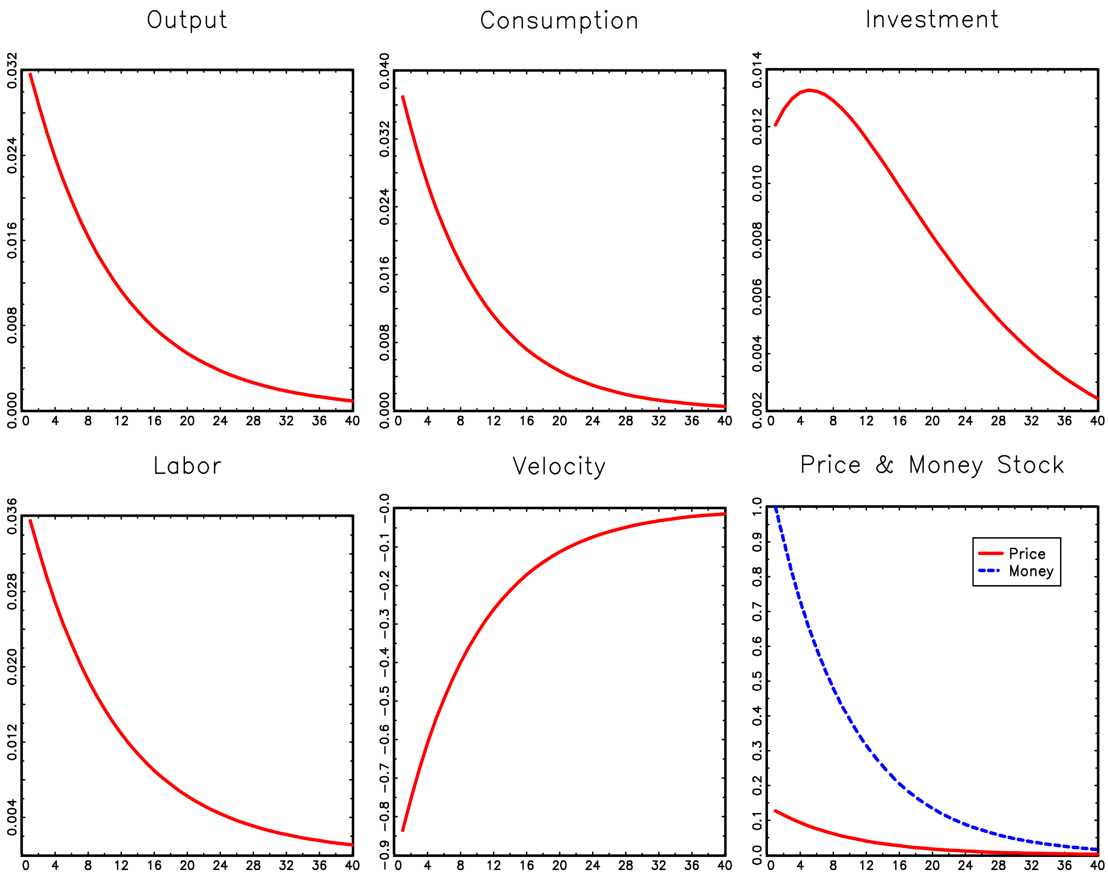

Figure 1. Impulse Responses to 1% Money Injection.

Clearly, money is expansionary: Output, consumption, investment, and labor all increase, albeit by a relatively small amount.[17] The velocity of money decreases. The aggregate price level is "sticky" – it increases by less than 0.15 percent, far less than the one-percent increase of money stock (see the window at the bottom right corner in Figure 1). Such a "sluggish" response of

[16]The variance of the Pareto distribution is a decreasing function of σ. The empirical literature based on distributions of income and wealth typically finds $\sigma \in (1.1, 3.5)$ or centered around $1.5 \sim 2.5$ (see, e.g., Wolff, 1996; Fermi, 1998; Levy and Levy, 2003; Clementi and Gallegati, 2005; and Nirei and Souma, 2007). Hence, $\sigma = 1.5$ is within the empirical estimates. However, if we allow the degree of risk aversion to be larger than 1 in the utility function, say let $u(c) = \theta(i) \frac{c(i)^{1-\gamma}}{1-\gamma}$ with $\gamma = 2$, then we can allow higher values of σ to yield similar results. The intuition is that risk aversion enhances the degree of heterogeneity because heterogeneity does not matter if individuals are risk neutral.

[17]As will be shown in the next section, financial intermediation can significantly amplify the real effects of lump-sum monetary injections.

aggregate price to money is also noted by Alvarez et al. (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 et al. (2008).

Transitory changes in the stock of money have real effects because only a fraction of the population are liquidity constrained and only the constrained agents will increase consumption when nominal income is higher. Consequently, the aggregate price level will not rise proportionately to the monetary increase. In addition, since agents opt to maintain a target level of real income-wealth so as 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. 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 so as to reduce the probability of being borrowing-constrained. Consequently, aggregate real balances rise more than aggregate consumption and the measured velocity of money decreases. Lastly, if the temporary money injection has a certain degree of persistence, then investment will also increase so as to help maintain future income-wealth on target by enhancing the productivity of labor.

An alternative way to understanding the movements in velocity and price is through equation (11), $1 = \left[\beta E_t \frac{P_t w_t}{P_{t+1} w_{t+1}}\right] R(\theta_t^*)$. 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 increase because $\frac{\partial R}{\partial \theta^*} < 0$. Since the function $D(\theta^*)$ will increase by less than $H(\theta^*)$, velocity must fall, offsetting the impact of the money injection on aggregate price.[18]

Welfare Costs of Long-Run Inflation. However, 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 under transitory monetary shocks are reversed. The rate of return to liquidity investment must increase to compensate for the cost of holding money. Hence, the cutoff θ^* must decrease and demand for real balances must fall. In particular, when the expected inflation rate is high enough above the critical value π_h, money will cease to be accepted as a store of value, optimal money demand goes to zero, and the velocity of money becomes infinity. In fact, inflation has two opposing effects on welfare: First, it increases the rate of return to money (R) by increasing

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

the probability of stockout and thus improves individual welfare for those who are not liquidity constrained. Second, it reduces the purchasing power of nominal balances and therefore raises the number of liquidity-constrained agents. Although the two effects work against each other, the second effect dominates and the Friedman rule (as a limiting case) is thus optimal.

We use two different measures for the welfare costs of inflation. The first is based on aggregate utility, $U = \int \theta(i) \log c(i) - \int n(i)$, and the second is based on aggregate consumption $C = \int c(i)di$. Since in the steady state aggregate output, employment, and capital stock are proportional to consumption with the coefficient independent of the inflation rate, the consumption-based measure is identical to measures based on income or employment. However, because the value of money becomes zero at high enough inflation rates, the relationship between inflation and the consumption-based measure of welfare costs is not monotonic and consequently not the same as the utility-based measure. Agents are able to avoid the inflation tax by reducing money holdings when inflation is too high. This implies that the aggregate consumption level is U-shaped and becomes identical to the Pareto-optimal level (under the Friedman rule) once the value of money reduces to zero. However, the utility level does not always increase with consumption because it requires a high cost of leisure to sustain a high level of consumption.

For example, with the Pareto distribution, we have $R(\theta) = \frac{1+\pi}{\beta} = 1 + \frac{1}{\sigma-1}\theta^{*-\sigma}$, or

$$\theta^* = \left[\frac{\beta}{1+\pi-\beta} \frac{1}{\sigma-1} \right]^{\frac{1}{\sigma}}. \tag{32}$$

Since the support of the distribution is $(1, \infty)$, an interior solution for the cutoff requires $\theta^* > 1$, which by (32) implies $1 + \pi < \beta \frac{\sigma}{\sigma-1}$. If this condition is violated, then no agents will hold money because $\Pr[\theta(i) \leq \theta^*] = 0$ if $\theta^* \leq 1$. Hence, the maximum rate of inflation to support a monetary equilibrium is $1 + \pi_h = \beta \frac{\sigma}{\sigma-1}$. For example, if the time interval is a year, $\beta = 0.95$, and $\sigma = 1.5$, then the maximum annual rate of inflation to support a monetary equilibrium is $\pi_h = 185\%$. If $\sigma = 1.1$, then $\pi_h = 945\%$.

Equation (16) implies the steady-state aggregate consumption is given by

$$C = w \left(\frac{1+\pi}{\beta} \right) \left[\frac{\sigma}{\sigma-1} - \frac{1}{\sigma-1} \left(\frac{1+\pi-\beta}{\beta}(\sigma-1) \right)^{\frac{\sigma-1}{\sigma}} \right]. \tag{33}$$

This function is U shaped in the interval $\beta \leq 1 + \pi \leq \beta \frac{\sigma}{\sigma-1}$ and has one maximum of $C^* = w \frac{\sigma}{\sigma-1}$ under the Friedman rule $1 + \pi = \beta$. The other maximum is attained at the upper bound where $1 + \pi_h = \beta \frac{\sigma}{\sigma-1}$. At this value of inflation we have $\theta^* = 1$ and the consumption level becomes identical to C^*. Hence, if we use the consumption-based measure, welfare can be increasing with inflation for a certain range of π. For example, if we follow Cooley and Hansen (1989) by measuring the welfare costs of inflation as the consumption ratio

$$\lambda = \frac{C^*}{C}, \tag{34}$$

then the welfare costs can be decreasing with inflation under high enough inflation rates. But the consumption-based measure of welfare costs ignores the cost of leisure. In order to maintain a high consumption level under high inflation rates, labor supply also has to be high, which reduces welfare if leisure cost is taken into consideration.

The utility-based measure, on the other hand, does not have the non-monotonic problem. The aggregate utility level is given by aggregating the individual's utilities by the law of large numbers, which gives

$$U = \frac{\sigma}{\sigma - 1} \log\left(w \frac{1+\pi}{\beta}\right) + \frac{\sigma}{(\sigma-1)^2}\left[1 - \left(\frac{1+\pi-\beta}{\beta}(\sigma - 1)\right)^{\frac{\sigma-1}{\sigma}}\right] - N(\pi). \tag{35}$$

This function can be shown to be monotonically decreasing in π. The Pareto-optimal level of aggregate utility is given by

$$U^* = \frac{\sigma}{\sigma-1}\log(w) + \frac{\sigma}{(\sigma-1)^2} - N^*, \tag{36}$$

where $N^* = (1-\alpha)\frac{1-\beta(1-\delta)}{1-\beta(1-\delta)-\delta\beta\alpha}\frac{\sigma}{\sigma-1}$ is the Pareto-optimal employment under the Friedman rule.

The effects of inflation on the model economy are graphed in Figure 2. The upper left window shows the utility difference, $U^* - U$, as the measure of welfare cost.[19] It is monotonically increasing with inflation. The upper right window shows the consumption-based welfare cost. It is not monotonic. Under this second measure, the welfare cost of inflation is zero at the two extreme points: the point of the Friedman rule and the point where velocity is infinity. In the first case, the welfare is the highest because there is no cost to holding money. In the latter case, no agent holds money anyway; hence, inflation at and beyond π_h has no adverse effects on aggregate consumption. The maximum cost is more than 20% of aggregate consumption when the inflation rate is about 10% a quarter or about 45% a year. The lower left window shows money demand as a function of inflation; it is downward sloping. Optimal money demand is infinity under the Friedman rule because of the special feature of the Pareto distribution (not because of the reasons provided by Bewley, 1983). It reduces to zero when $\pi_h = \beta\frac{\sigma}{\sigma-1} - 1$. The behavior of velocity is shown in the lower right window. The velocity is zero when money demand is infinity, and it becomes infinity

[19] We use the utility difference because the utility level may be zero or negative.

when money demand is zero at the upper bound of inflation π_h. These implications for money demand and velocity are very different from canonical 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, it is often observed that people refuse to accept domestic currency as the means of payment when the inflation rate is too high, long before it becomes infinity.

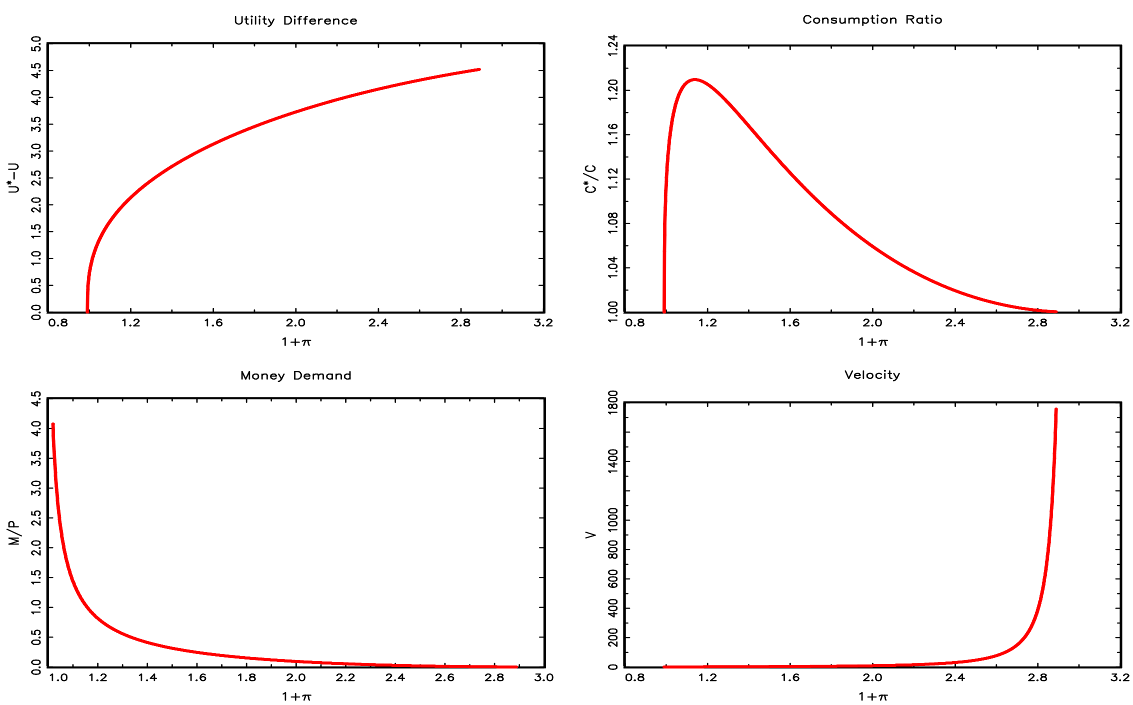

Figure 2. Welfare Cost and Velocity ($\sigma = 1.5$)**.**

In a heterogeneous-agent economy, the welfare costs, regardless how they are measured, are potentially much higher than those in a representative-agent model because of uneven distribution of money holdings. The larger the variance of θ, the stronger the precautionary motive for holding money. Agents with higher nominal balances suffer disproportionately larger welfare losses than agents with smaller balances because of the concavity of the utility function. In addition, higher inflation induces all agents to hold less money, increasing the probability of being borrowing-constrained. Hence, the larger the degree of heterogeneity (given the degree of risk aversion), the higher is the welfare cost of inflation. For example, when $\sigma = 3.5$, the consumption-based welfare cost for zero inflation is about 1% of aggregate consumption, and for 3% quarterly inflation it is about 1.8% of consumption. On the other hand, when $\sigma = 1.5$ (as in Figure 2), the consumption-based welfare cost for zero inflation is about 11.8% of consumption and for 3% quarterly inflation it is about 17% of consumption. These are extraordinary numbers. Also, the value of money becomes zero at a much lower rate of inflation when the degree of idiosyncratic risk is higher. These findings suggest that heterogeneity has great implications for welfare and the business cycle, and such implications have not been fully appreciated by the literature.[20]

These high levels of welfare costs cannot possibly be generated from representative-agent models unless extreme degrees of risk aversion in the representative agent's utility function are assumed. Based on such findings, it is safe to say that the welfare costs of inflation may have been significantly underestimated by Lucas (2000) and Cooley and Hansen (1989) in representative-agent models.[21]

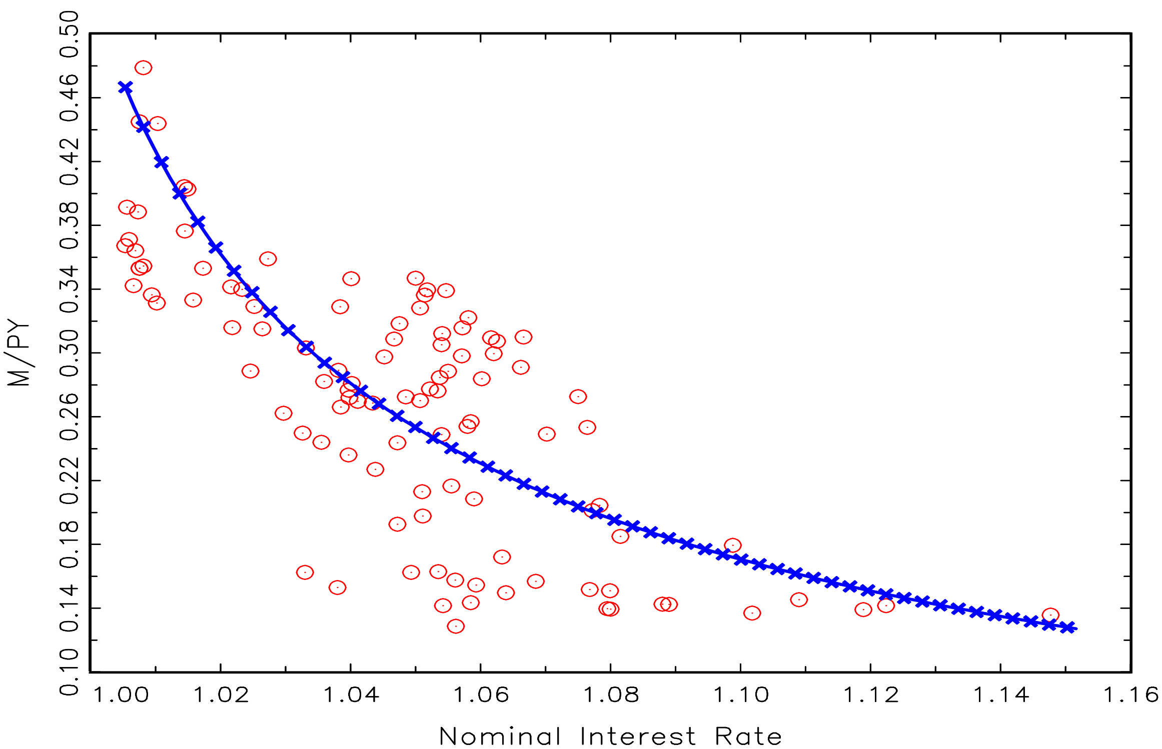

Figure 3. Money Demand Curve in the Model and Data.

The realism of the calibrated model can be tested using empirical data. For example, the model is 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 money-in-the-utility. Lucas estimated that the

[20] An additional factor contributing to the large welfare costs of inflation is that agents switch from "cash" goods (consumption) to "credit" goods (leisure), reducing hours worked and aggregate output. Although money is not a medium of exchange, goods consumption can be facilitated by holding money whereas leisure cannot.

[21] Imrohoroglu (1992) also found that heterogeneous agents in the Bewley model generate higher welfare costs of inflation. His estimate is around $1 - 2\%$ of aggregate consumption, still far smaller than what is obtained in this paper. The reason may be that his model is an endowment economy.

empirical money demand curve can be best captured by a power function of the form,

$$\frac{M}{PY} = Ar^{-\eta}, \tag{37}$$

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. Since the "money demand" defined by Lucas is identical to the inverted velocity, a downward-sloping money demand curve is the same thing as an upward-sloping velocity curve (namely, velocity is positively related to nominal interest rate or inflation). Analogous to Lucas, the money demand curve implied by the benchmark model of this paper takes the form

$$\frac{M}{PY} = A\frac{H(\theta^*)}{D(\theta^*)}, \tag{38}$$

where A is a scale parameter, the functions $\{H, D\}$ are defined by equations (20) and (19), and the cutoff θ^* is a function of the nominal interest rate implied by equation (26). Figure 3 shows a surprisingly close fit of my theoretical model to the U.S. data.[22]

3 Banking, Interest Rates, and the Liquidity Trap

The key friction in the benchmark model is the borrowing constraint on nominal balance. With this constraint, there is an ex post inefficiency since some agents are holding idle balances while others are liquidity constrained. This creates needs for risk sharing, as suggested by Lucas (1980). However, without necessary information- and record-keeping technologies, households cannot lend and borrow among themselves. In this section, we assume that a community bank emerges to resolve the risk-sharing problem by developing the required information technologies. The function of the bank is to accept nominal deposits from households and make nominal loans to those in need. For simplicity, we assume that deposits do not pay interest and all households voluntarily deposit their idle cash into the bank as a safety net. The benefit of making deposits is that bank members are qualified for loans when needed. This provides enough incentives for agents in the community to pull together their cash resources. Assume all deposits are withdrawn at the end of each period (100-percent reserve banking), and all loans are one-period loans that charge the competitive nominal interest rate $1 + \tilde{\imath}$, which is determined by the demand and supply of loans in the community. Any profits earned by the bank are redistributed back to community members as lump-sum transfers.

[22] The 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 taken from the online Historical Statistics of the United States–Millennium Edition. The solid line with star symbols is the model's prediction calibrated at annual frequency with $\beta = 0.96$ and $\delta = 0.1$. The other parameters remain the same; namely, $\alpha = 0.3$ and $\sigma = 1.5$. The nominal interest rate in the model is defined as $\frac{1+\pi}{\beta}$. The scale parameter is set to $A = 0.125$.

Similar banking arrangements have been studied recently by Berentsen, Camera, and Waller (2006) and others. This literature shows that financial intermediation improves welfare.[23] However, these authors study the issue in the framework of Lagos and Wright (2005), which has no capital accumulation and aggregate uncertainty, and they do not analyze the issue of the liquidity trap. In addition, in their model the welfare gains of financial intermediation come solely from the payment of interest on deposits and not from relaxing borrowers' liquidity constraints. In sharp contrast, gains in welfare in this paper derive entirely from relaxing borrowers' liquidity constraints.

The time line of events is as follows: In the beginning of each period, aggregate shocks are realized, each household then makes decisions on labor supply and capital investment, taking as given the initial wealth from last period. After that, idiosyncratic preference shocks are realized, and each household chooses consumption, the amount of nominal balances to be carried over to the next period, and the size of new loans if needed. Given such an environment, it is clear that agents with idle cash will not take a loan in that period and that agents who take loans must be cash constrained. It is also possible for a cash-constrained agent not to take any loans if the urge to consume is not high enough to justify the interest rate on a loan. Hence, in terms of cash balances, there may exist three types of households in each period: depositors, borrowers, and agents with zero deposits and loans.

Household i takes the bank's real profit income (T) and government money transfers (τ) as given, and chooses consumption, capital investment, labor supply, money demand, and credit borrowing (b_t) to solve

$$\max \sum_{t=0}^{\infty} \beta^t \left\{ \theta(i) \log c(i) - n(i) \right\}$$

subject to

$$c_t(i) + k_{t+1}(i) + \frac{m_t(i)}{P_t} + (1+\tilde{\imath}_t)\frac{b_{t-1}(i)}{P_t} \leq (1+r_t)k_t(i) + \frac{m_{t-1}(i)+\tau_t}{P_t} + \frac{b_t(i)}{P_t} + w_t n_t(i) + T_t \quad (39)$$

$$m_t(i) \geq 0 \quad (40)$$

$$b_t(i) \geq 0, \quad (41)$$

where $\tilde{\imath}$ denotes the nominal loan rate and r the rental rate of capital. The non-negativity constraints on nominal balances (m_t) and loans (b_t) capture the idea that households cannot borrow or lend outside the banking system. As in the benchmark model, hours worked and non-monetary asset investment in each period must be determined before the idiosyncratic preference shock $\theta_t(i)$ is realized.

[23] However, Chiu and Meh (2008) show that banking may reduce welfare under moderate inflation rates if there exist transaction costs for using financial intermediation. For alternative approaches to money and banking, see Williamson (1986) and Andolfatto and Nosal (2003).

Denoting $\left\{\lambda, \pi^m, \pi^b\right\}$ as Lagrangian multipliers for constraints (39)-(41), respectively, the first-order conditions with respect to $\{c, k', n, m, b\}$ are given by

$$\frac{\theta(i)}{c(i)} = \lambda(i) \tag{42}$$

$$1 = w_t E^i \lambda_t(i) \tag{43}$$

$$E^i \lambda_t(i) = \beta E_t \left(1 + r_{t+1}\right) \lambda_{t+1}(i) \tag{44}$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t \frac{\lambda_{t+1}(i)}{P_{t+1}} + \pi_t^m(i) \tag{45}$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t (1 + \tilde{\imath}_{t+1}) \frac{\lambda_{t+1}(i)}{P_t} + \pi_t^b(i). \tag{46}$$

Note that there are three possible situations for money demand: (i) If $m_t(i) > 0$, then $b_t(i) = 0$; namely, a household has no incentive to take a loan if it has idle cash in hand. (ii) If $b_t(i) > 0$, then $m_t(i) = 0$; namely, a household will take a loan only if it runs out of cash. (iii) It is possible that a household has no cash in hand but does not want to borrow money from the bank because the interest rate is too high; namely, $m_t(i) = b_t(i) = 0$. Which of the three situations prevails in each period depends on the realized value of the preference shock $\theta_t(i)$. There exist two cutoff values, $\underline{\theta}$ and $\bar{\theta}$ with $\underline{\theta} < \bar{\theta}$. If $\theta(i) < \underline{\theta}$, the urge to consume is low, then $m_t(i) > 0$; if $\theta(i) > \bar{\theta}$, the demand for liquidity is high, then $b_t(i) > 0$; if $\underline{\theta} \leq \theta \leq \overline{\theta}$, then $m_t(i) = b_t(i) = 0$; where $\left\{\underline{\theta}, \overline{\theta}\right\}$ are determined endogenously by the households.

The analyses of each case proceed as follows. By the law of iterated expectations and by the orthogonality condition between the idiosyncratic shocks and aggregate shocks, equations (44)-(46) can be rewritten (by using equation 43) as:

$$\frac{1}{w_t} = \beta E_t \left(1 + r_{t+1}\right) \frac{1}{w_{t+1}} \tag{47}$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t \frac{1}{P_{t+1} w_{t+1}} + \pi_t^m(i) \tag{48}$$

$$\frac{\lambda_t(i)}{P_t} = \beta E_t (1 + \tilde{\imath}_{t+1}) \frac{1}{P_t w_{t+1}} + \pi_t^b(i). \tag{49}$$

Case A. $m_t(i) > 0, b_t(i) = 0, \pi_t^m(i) = 0$. Equation (47) implies $\lambda_t(i) = \beta E_t \frac{P_t}{P_{t+1} w_{t+1}}$; equation (42) implies $c_t(i) = \theta(i) \left[\beta E \frac{P_t}{P_{t+1} w_{t+1}}\right]^{-1}$. Defining the wealth,

$$x_t(i) \equiv \frac{m_{t-1}(i) + \tau_t}{P_t} + w_t n_t(i) + T_t - (1 + \tilde{\imath}_t) \frac{b_{t-1}}{P_t} + \left(1 + r_t\right) k_t(i) - k_{t+1}(i),$$

the budget constraint implies $m_t(i) = x_t(i) - \theta(i)\left[\beta E \frac{P_t}{P_{t+1}w_{t+1}}\right]^{-1} > 0$, which implies

$$\theta(i) < x_t(i)\beta E \frac{aP_t}{P_{t+1}w_{t+1}} \equiv \underline{\theta}_t, \tag{50}$$

which defines the lower cutoff value $\underline{\theta}$.

Case B. $b_t(i) > 0, m_t(i) = 0, \pi_t^b(i) = 0$. Equation (49) implies $\lambda_t(i) = \beta E_t(1+\tilde{\imath}_{t+1})\frac{P_t}{P_{t+1}w_{t+1}}$; equation (42) implies $c_t(i) = \theta(i)\left[\beta E(1+\tilde{\imath}_{t+1})\frac{P_t}{P_{t+1}w_{t+1}}\right]^{-1}$. The budget constraint implies $b_t(i) = -x_t(i) + \theta(i)\left[\beta E(1+\tilde{\imath}_{t+1})\frac{P_t}{P_{t+1}w_{t+1}}\right]^{-1} > 0$, which implies

$$\theta(i) > x_t(i)\beta E(1+\tilde{\imath}_{t+1})\frac{P_t}{P_{t+1}w_{t+1}} \equiv \bar{\theta}_t, \tag{51}$$

which defines the upper cutoff value $\bar{\theta}$. Clearly, $\bar{\theta}_t \geq \underline{\theta}_t$ if and only if $\tilde{\imath}_t \geq 0$ for all t.

Because we have both $x_t(i) \equiv \underline{\theta}_t\left[\beta E\frac{P_t}{P_{t+1}w_{t+1}}\right]^{-1}$ and $x_t(i) \equiv \bar{\theta}_t\left[\beta E(1+\tilde{\imath}_{t+1})\frac{P_t}{P_{t+1}w_{t+1}}\right]^{-1}$, it must be true that

$$\frac{\bar{\theta}_t}{\underline{\theta}_t} = \frac{E(1+\tilde{\imath}_{t+1})\frac{P_t}{P_{t+1}w_{t+1}}}{E\frac{P_t}{P_{t+1}w_{t+1}}}. \tag{52}$$

Notice that the lower and the upper cutoff values depend on aggregate economic conditions and hence both are time varying.

Case C. $m_t(i) = b_t(i) = 0$. The budget constraint implies $c_t(i) = x_t(i) = \underline{\theta}_t\left[\beta E\frac{P_t}{P_{t+1}w_{t+1}}\right]^{-1}$, and equation (42) implies $\lambda_t(i) = \frac{\theta(i)}{\underline{\theta}}\beta E\frac{P_t}{P_{t+1}w_{t+1}}$.

Because the shadow price $\lambda_t(i)$ takes three possible values across the three cases, equation (43) implies

$$1 = \beta E\frac{P_t w_t}{P_{t+1}w_{t+1}}R(\underline{\theta}_t, \bar{\theta}_t), \tag{53}$$

where

$$R_t \equiv \left[\int_{\theta(i)<\underline{\theta}} dF(\theta) + \int_{\underline{\theta}\leq\theta(i)\leq\bar{\theta}} [\theta(i)/\underline{\theta}]\, dF(\theta) + \int_{\theta(i)>\bar{\theta}} [\bar{\theta}/\underline{\theta}]\, dF(\theta)\right] \tag{54}$$

measures the equilibrium rate of return to liquidity in the model of narrow banking. Notice that $R_t \geq 1$ if and only if $\bar{\theta}_t \geq \underline{\theta}_t$. As in the benchmark model, $R = \frac{1+\pi}{\beta}$ in the steady state.

Individuals' decision rules can be summarized by

$$c_t(i) = \begin{cases} \theta(i) w_t R_t & \text{if } \theta(i) < \underline{\theta} \\ \underline{\theta} w_t R_t & \text{if } \underline{\theta} \leq \theta(i) \leq \overline{\theta} \\ \frac{\theta(i)}{\overline{\theta}} \underline{\theta} w_t R_t & \text{if } \theta(i) > \overline{\theta} \end{cases} \tag{55}$$

$$\frac{m_t(i)}{P_t} = \begin{cases} \left[\underline{\theta} - \theta(i)\right] w_t R_t & \text{if } \theta(i) < \underline{\theta} \\ 0 & \text{if } \theta(i) \geq \underline{\theta} \end{cases} \tag{56}$$

$$\frac{b_t(i)}{P_t} = \begin{cases} 0 & \text{if } \theta(i) \leq \overline{\theta} \\ \left[\theta(i) - \overline{\theta}\right] \left(\underline{\theta}/\overline{\theta}\right) w_t R_t & \text{if } \theta(i) > \overline{\theta} \end{cases} \tag{57}$$

$$x_t(i) = \underline{\theta} w_t R_t. \tag{58}$$

Aggregating these decision rules across households gives

$$C_t = w_t R(\underline{\theta}_t, \bar{\theta}_t) D(\underline{\theta}_t, \bar{\theta}_t) \tag{59}$$

$$\frac{M_t}{P_t} = w_t R\left(\underline{\theta}_t, \bar{\theta}_t\right) H\left(\underline{\theta}_t, \bar{\theta}_t\right) \tag{60}$$

$$\frac{B_t}{P_t} = w_t R\left(\underline{\theta}_t, \bar{\theta}_t\right) G\left(\underline{\theta}_t, \bar{\theta}_t\right) \tag{61}$$

$$\frac{M_{t-1} + \tau_t}{P_t} + w_t N_t + T_t - (1 + \tilde{\imath}_t) \frac{B_{t-1}}{P_t} + (1 + r_t) K_t - K_{t+1} = w_t R_t \underline{\theta}, \tag{62}$$

where the functions $\{D, H, G\}$ are defined as

$$D_t \equiv \left[\int_{\theta(i)<\underline{\theta}} \theta(i) dF(\theta) + \int_{\underline{\theta} \leq \theta(i) \leq \overline{\theta}} \underline{\theta} dF(\theta) + \int_{\theta(i)>\overline{\theta}} \theta(i) \left(\underline{\theta}/\overline{\theta}\right) dF(\theta)\right] > 0 \tag{63}$$

$$H_t \equiv \int_{\theta(i)<\underline{\theta}} \left[\underline{\theta} - \theta(i)\right] dF(\theta) > 0 \tag{64}$$

$$G_t \equiv \int_{\theta(i)>\overline{\theta}} \left[\theta(i) - \overline{\theta}\right] \left(\underline{\theta}/\overline{\theta}\right) dF(\theta) > 0 \tag{65}$$

and the three functions satisfy the identity

$$D + H - G = \underline{\theta}. \tag{66}$$

In the credit market, the aggregate supply of credit is M_t and the aggregate demand is $B(\tilde{\imath}_t)$. Note credit demand cannot exceed supply because the loan rate will always rise to clear the market, and the nominal loan rate cannot be negative because people have the option not to deposit. Hence, the credit market-clearing conditions are characterized by the following complementarity conditions:

$$(M_t - B_t)\,\tilde{\imath}_t = 0; \quad M_t \geq B_t, \tilde{\imath}_t \geq 0. \tag{67}$$

That is, the nominal loan rate is zero if liquidity supply exceeds its demand. On the other hand, if credit demand exceeds supply, the nominal interest rate will rise to clear the market. Notice that the bank does not accumulate reserves because all reserves are redistributed back to bank members by the end of each period. The bank's balance sheet is given by

$$\underbrace{M_t}_{\text{deposit}} + \underbrace{(1+\tilde{\imath}_{t+1})B_t}_{\text{loan payment}} \quad \Longrightarrow \quad \underbrace{M_t}_{\text{withdraw}} + \underbrace{B_t}_{\text{new loan}} + \underbrace{T_{t+1}}_{\text{profit income}}, \tag{68}$$

where the left-hand side is total inflow of liquidity in period t and the right-hand side is total outflow of liquidity in period t. That is, in the beginning of period t the bank accepts deposit M_t and makes new loans B_t, and at the end of period t it receives loan payment $(1+\tilde{\imath}_{t+1})\,B_t$ and faces withdrawal of M_t. Any profits are distributed back to households in a lump sum at the end of period t in the amount T_{t+1}, which becomes household income in the beginning of the next period.

3.1 The Liquidity Trap

Suppose the steady-state inflation rate is π and the real wage is w. In a steady state, equations (52), (53), (59), (60), (61), (62), and (67) imply

$$R(\underline{\theta}, \overline{\theta}) = \frac{1+\pi}{\beta} \tag{69}$$

$$1 + \tilde{\imath} = \overline{\theta}/\underline{\theta} \tag{70}$$

$$C = w\frac{1+\pi}{\beta}D(\underline{\theta}, \overline{\theta}) \tag{71}$$

$$\frac{M}{P} = w\frac{1+\pi}{\beta}H(\underline{\theta}) \tag{72}$$

$$\frac{B}{P} = w\frac{1+\pi}{\beta}G(\underline{\theta}, \overline{\theta}) \tag{73}$$

$$(M - B)\tilde{\imath} = 0, \;\; M - B \geq 0, \;\;\; \tilde{\imath} \geq 0, \tag{74}$$

where the functions $\{R, D, H, G\}$ are defined in equations (54), (63), (64), and (65).

Proposition 1 *If $1+\pi \geq \beta$, the above equation system uniquely solves for $\left\{\underline{\theta}, \overline{\theta}\right\}$ with the property $\overline{\theta} \geq E(\theta) \geq \underline{\theta}$.*

Proof. First, by (54), it is clear that $R(\underline{\theta}, \bar{\theta}) = 1$ if and only if $\overline{\theta} = \underline{\theta}$; and $R > 1$ if and only if $\overline{\theta} > \underline{\theta}$. Hence, by (69) we have $1 + \pi \geq \beta$ if and only if $\bar{\theta} \geq \underline{\theta}$. Second, since $\bar{\theta} \geq \underline{\theta}$, we have $\tilde{\imath} \geq 0$ and $M = B$. But $M = B$ is equivalent to $H(\underline{\theta}) = G(\underline{\theta}, \bar{\theta})$ or

$$\int_{\theta(i)<\underline{\theta}} \left[\underline{\theta} - \theta(i)\right] dF(\theta) = \int_{\theta(i)>\overline{\theta}} \left[\theta(i) - \overline{\theta}\right] \left(\underline{\theta}/\overline{\theta}\right) dF(\theta);$$

which, by rearranging, implies

$$\begin{aligned} \left[F(\underline{\theta}) + 1 - F(\overline{\theta})\right] &= \int_{\theta<\underline{\theta}} \frac{\theta(i)}{\underline{\theta}} dF(\theta) + \int_{\theta>\overline{\theta}} \frac{\theta(i)}{\overline{\theta}} dF(\theta) \\ &= \left[F(\underline{\theta}) + 1 - F(\overline{\theta})\right] + \int_{\theta<\underline{\theta}} \frac{\theta(i) - \underline{\theta}}{\underline{\theta}} dF(\theta) + \int_{\theta>\overline{\theta}} \frac{\theta(i) - \overline{\theta}}{\overline{\theta}} dF(\theta) \end{aligned}$$

which implies

$$\int_{\theta<\underline{\theta}} \frac{\underline{\theta} - \theta}{\underline{\theta}} dF(\theta) = \int_{\theta>\overline{\theta}} \frac{\theta - \overline{\theta}}{\overline{\theta}} dF(\theta). \tag{75}$$

When $1 + \pi = \beta$, $\overline{\theta} = \underline{\theta}$, the above equation implies the solution $\overline{\theta} = \underline{\theta} = \int \theta dF(\theta)$ $(= E(\theta))$. When $1 + \pi > \beta$, $\overline{\theta} > \underline{\theta}$, the above equation implies that, for any value of $\underline{\theta} < E(\theta)$, there exists a $\overline{\theta} > E(\theta)$ such that the area measured by the left-hand-side of (75) equals the area measured by the right-hand-side of (75) for any non-degenerated distributions. Plugging this relationship implied by (75), $\overline{\theta}(\underline{\theta})$, into the continuous and single-valued relation $R(\underline{\theta}, \overline{\theta}) = \frac{1+\pi}{\beta}$ uniquely determines the value of $\underline{\theta}$. Given $\underline{\theta}$, $\bar{\theta}$ can then be uniquely determined by (75). ■

Proposition 2 *Suppose the support of the distribution is given by $[\theta_l, \theta_h]$ with $\theta_h > \theta_l > -\infty$. Then there exists a finite upper limit (π_h) of the inflation rate such that, if $\pi = \pi_h$, then the optimal demand for real balances $\frac{M}{P} = 0$; namely, no household is willing to hold cash if inflation is at or above π_h.*

Proof. By (54), we have $\frac{\partial R}{\partial \underline{\theta}} < 0$ and $\frac{\partial R}{\partial \bar{\theta}} > 0$. Hence, given the support of θ, the maxim value of R ($R_{\max}$) is reached either when $\underline{\theta} = \theta_l$ or $\bar{\theta} = \theta_h$ or both. By Proposition 1 and (75), $\underline{\theta} = \theta_l$ if and only if $\bar{\theta} = \theta_h$. Hence, there exists π_h such that R is at its maximum value $R_{\max} = \frac{1+\pi_h}{\beta}$ if and only if $\underline{\theta} = \theta_l$. Then by (64), we have $H(\underline{\theta}) = 0$; and by (72), we have $\frac{M}{P} = 0$. Since there is no credit supply in the banking system, we must also have $\frac{B}{P} = 0$ in equilibrium regardless of the nominal loan rate. ■

Proposition 3 *(The Liquidity Trap) There exists a lower limit on the inflation rate, $\pi_l = \beta - 1$, such that if $\pi < \pi_l$, then $\tilde{\imath}_t = 0$ and $M > B$. Namely, households opt to hoard too much cash such that the aggregate credit supply (deposits) exceeds credit demand even with a zero interest rate on loans.*

Proof. The function $R(\underline{\theta}, \overline{\theta})$ is decreasing in $\underline{\theta}$ and increasing in $\overline{\theta}$. Also, if $\overline{\theta} \geq \underline{\theta}$, then the minimum value of R is given by $R_{\min} = F(\underline{\theta}) + \left[F(\overline{\theta}) - F(\underline{\theta})\right] + \left(\overline{\theta}/\underline{\theta}\right)\left[1 - F(\overline{\theta})\right] \geq 1$; if $\overline{\theta} < \underline{\theta}$, then the maximum value of R is given by $R_{\max} = F(\underline{\theta}) + \left(\overline{\theta}/\underline{\theta}\right)\left[F(\overline{\theta}) - F(\underline{\theta})\right] + \left(\overline{\theta}/\underline{\theta}\right)\left[1 - F(\overline{\theta})\right] < 1$. Hence, $R \lesseqgtr 1$ if and only if $\overline{\theta} \lesseqgtr \underline{\theta}$. It is also true that $R \lesseqgtr 1$ if and only if $\pi \lesseqgtr \beta - 1$. Therefore, when $\pi \leq \beta - 1$, $R \leq 1$, we must also have $\overline{\theta} \leq \underline{\theta}$. However, for equation (75) to hold (i.e., under the condition $M = B$), it must be true that $\underline{\theta} \geq E(\theta) \geq \bar{\theta}$ if $\overline{\theta} \leq \underline{\theta}$. Since $1 + \tilde{\imath} = \overline{\theta}/\underline{\theta}$ and the nominal interest rate cannot be negative, we cannot have $\overline{\theta} < \underline{\theta}$. Therefore, there exists a lower limit $\pi_l = \beta - 1$, at which point $R = 1, \overline{\theta} = \underline{\theta}$, and $\tilde{\imath} = 0$; and whenever $\pi < \pi_l$, we must violate equation (75) so that $H(\underline{\theta}) > G(\underline{\theta}, \bar{\theta} = \underline{\theta})$ (i.e., $M > B$ at the point $\tilde{\imath} = 0$) because $H(\underline{\theta})$ (credit supply) is increasing in $\underline{\theta}$ while $G(\underline{\theta}, \overline{\theta} = \underline{\theta})$ (loan demand at zero interest rate) is decreasing in $\underline{\theta}$ (since at $\bar{\theta} = \underline{\theta}$ we have $G \equiv \int_{\theta(i)>\bar{\theta}} \left[\theta(i) - \overline{\theta}\right] \left(\underline{\theta}/\overline{\theta}\right) dF(\theta) = \int_{\theta(i)>\underline{\theta}} \left[\theta(i) - \underline{\theta}\right] dF(\theta)$). ■

Proposition 4 *Velocity increases with inflation.*

Proof. $V = \frac{D(\underline{\theta}, \overline{\theta})}{H(\underline{\theta})}$, where H increases with $\underline{\theta}$ and D increases with $\left(\overline{\theta} - \underline{\theta}\right)$ if $\overline{\theta} > \underline{\theta}$. We know that $\underline{\theta}$ decreases and $\left(\overline{\theta} - \underline{\theta}\right)$ increases as π increases; hence, V increases with π. ■

In a liquidity trap, money is such an attractive asset to hold that any additional money injection will be hoarded by the private sector, thus increasing liquidity supply (deposits) in the banking system. On the other hand, since the nominal interest rate on loans cannot decrease below zero, the demand for loans will not be further stimulated to absorb the excess supply of liquidity. Hence, monetary policy will cease to be effective in stimulating credit demand and aggregate spending through the credit channel of the banking system.

If banks use deposits to invest in financial assets such as a one-period government bonds, then in equilibrium the nominal interest rate on government bonds must equal the loan rate $1 + \tilde{\imath}_t$. In this case, an open-market operation by the government has the same effects of a lump-sum money injection because we assume households always deposit idle balances into the banking system. Another important feature of the liquidity trap is that money demand does not become infinity at the trap, nor does the demand for loans become zero.[24] This is in contrast to the argument of Grandmont and Laroque (1976).

3.2 Welfare Gains of Financial Intermediation

The model can be closed by adding production and capital accumulation in the same way as in the benchmark model. Namely, on the firm side we have $Y_t = A_t(e_t K_t)^\alpha N_t^{1-\alpha}$, $\delta_t = \frac{1}{1+\omega} e_t^{1+\omega}$, $w_t = (1-\alpha)\frac{Y_t}{N_t}$, $r_t - \delta_t = \alpha \frac{Y_t}{K_t}$, and $\alpha \frac{Y_t}{e_t} = e_t^\omega K_t$. The goods market-clearing condition is given by

$$C_t + K_{t+1} - (1-\delta)K_t = Y_t. \tag{76}$$

As in the benchmark model, we consider two types of monetary policies and the aggregate money stock evolves according to

$$M_t = M_{t-1} + \tau_t, \tag{77}$$

where money injection τ_t follows either policy (21) or policy (23).

The general equilibrium is defined as the path $\left\{C_t, K_{t+1}, N_t, Y_t, e_t, \delta_t, w_t, \tilde{\imath}_t, r_t, P_t, M_t, B_t, \bar{\theta}_t, \underline{\theta}_t, T_t\right\}$, which satisfies equations (47), (52), (53), (59), (60), (61), (62), (67), (76), (77), the production function, the capital depreciation function, and firms' first-order conditions with respect to $\{e_t, K_t, N_t\}$, as well as standard transversality conditions.[25] The model has a unique steady state in which the following relationships hold: $\frac{K}{Y} = \frac{\alpha\beta}{1-\beta(1-\delta)}$, $\frac{C}{Y} = 1 - \frac{\delta\alpha\beta}{1-\beta(1-\delta)}$, $w = (1-\alpha)\left(\frac{\beta\alpha}{1-\beta(1-\delta)}\right)^{\frac{\alpha}{1-\alpha}}$, and $1 + r = \frac{1}{\beta}$.

Consider the Pareto distribution $F(\theta) = 1 - \theta^{-\sigma}$, with support $\theta \in (1, \infty)$ and the shape parameter $\sigma > 1$. The aggregate utility as a function of inflation is given by

$$U^{bank} = \frac{\sigma}{\sigma - 1} \log\left(w \frac{1+\pi}{\beta}\right) + \frac{\sigma}{(\sigma-1)^2}\left[1 + \bar{\theta}^{1-\sigma} - \underline{\theta}^{1-\sigma}\right] - N(\pi). \tag{78}$$

Under the Friedman rule, $\pi = \beta - 1$, we have $\bar{\theta} = \underline{\theta}$. Hence, the aggregate utility level is given by

$$U^* = \frac{\sigma}{\sigma-1}\log(w) + \frac{\sigma}{(\sigma-1)^2} - N^*, \tag{79}$$

where $N^* = (1-\alpha)\frac{1-\beta(1-\delta)}{1-\beta(1-\delta)-\delta\beta\alpha}\frac{\sigma}{\sigma-1}$ is the Pareto-optimal employment under the Friedman rule. Equation (79) looks identical to equation (36) because the optimal allocations are the same under Friedman rule with or without financial intermediation. Comparing equation (78) with equation

[24] With banking, the Friedman rule implies $\underline{\theta} = \bar{\theta} = E\theta$; hence, the money demand function is always finite valued even if the support of θ is unbounded.

[25] The transversality conditions are the same as in the benchmark model.

(35), it is clear that financial intermediation improves welfare mainly because of the extra term $\bar{\theta}^{1-\sigma}$, which reflects the utility gains from additional consumption by relaxing the borrowing constraint through credit lending.

Figure 4 shows the welfare gains of financial intermediation based on the two measures of welfare, one by utility difference ($U^{bank} - U^0$) and the other by consumption ratio (C^{bank}/C^0), where U^{bank} denotes aggregate utility with banking and U^0 the counterpart in the benchmark model. Similar notations apply to aggregate consumption. The top window in Figure 4 shows that financial intermediation improves welfare significantly. This is especially the case for moderate values of inflation. Near the Friedman rule there is little gain from risk sharing because agents can perfectly self-insure against consumption risk when the cost of holding money is zero. For high inflation rates the value of money is low, so redistributing idle cash balances does not significantly improve welfare. Similar results are also obtained by Berentsen, Camera, and Waller (2006) in the Lagos-Wright (2005) framework.

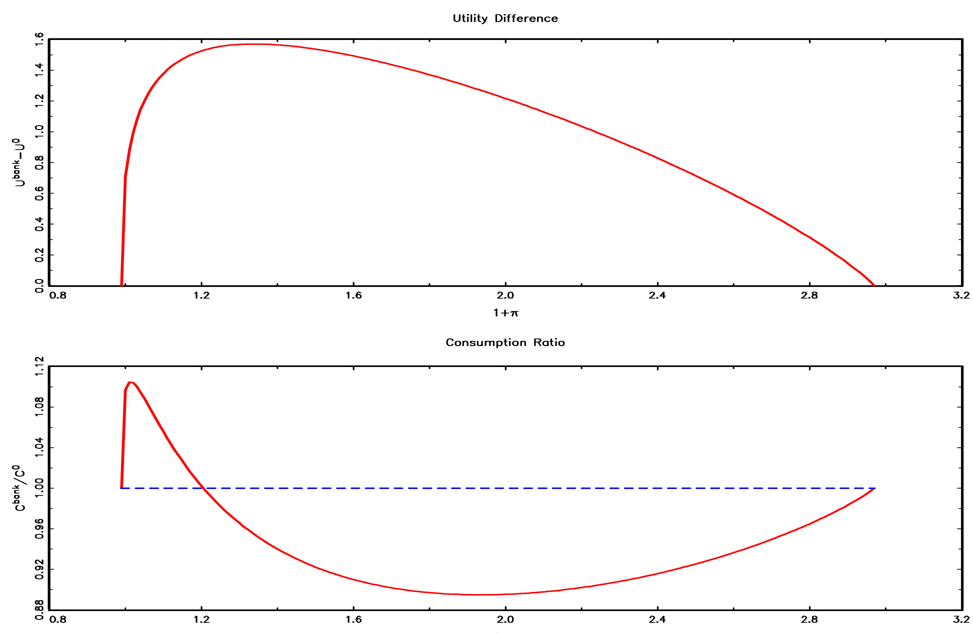

Figure 4. Welfare Gains from Banking.

However, the bottom window in Figure 4 shows that financial intermediation does not necessarily improve welfare if welfare is measured by consumption or the output gap. For high enough inflation rates, consumption and output are lower with financial intermediation than without. The reason: With the possibility of borrowing, agents pay higher interest costs on their debts under higher inflation, which crowds out savings. Hence, in the steady state the levels of output and consumption are lower, everything else equal. Thus, with moderate inflation financial intermediation increases aggregate consumption by relaxing borrowing constraints, and this effect dominates the interest effects on debt. But with high inflation, the interest effects dominate, so the welfare gain is reversed.

Also, the larger the degree of heterogeneity, the higher are the output gains from financial intermediation with moderate inflation. For example, under the current calibration with $\sigma = 1.5$, the welfare gain of banking is about 10 percent of aggregate consumption (or output) when the inflation rate is 2% a quarter (or 8% a year). If $\sigma = 1.1$, which is also consistent with the empirical "money demand" curve estimated by Lucas (2000), the welfare gain becomes 68% of aggregate consumption (or output) when inflation is 2% a quarter, suggesting that the collapse of the banking system can destroy as much as 68% of aggregate GDP for a sufficiently heterogeneous economy.

Interest Rates. The interest rate of credit does not necessarily equal nor comove with the rate of return to liquidity (R) or the rate of yield on non-monetary assets (such as the capital stock). More importantly, the gap depends positively on the rate of inflation. For example, in the current model, the steady-state rate of return to capital (the Fisherian fundamentals) is independent of the inflation rate, $1 + r = \frac{1}{\beta}$, so the nominal rate of return to capital is given by $\frac{1+\pi}{\beta}$, which is the same as the shadow rate of return to liquidity, $R(\theta^*)$. However, the nominal interest rate of loans behaves very differently; it is given by $1 + \tilde{\imath} = \frac{\bar{\theta}(\pi)}{\underline{\theta}(\pi)}$. Although both types of nominal rates increase with inflation, their elasticities differ dramatically. For the nominal asset return this elasticity is one, so the real rate is independent of inflation. But the nominal loan rate has an elasticity larger than one. Suppose we define the real rate of credit by $\frac{1+\tilde{\imath}}{1+\pi}$; the right window in Figure 5 shows that the real rate is still an increasing function of the inflation rate. Because real money demand decreases with inflation (the left window in Figure 5), the relationship between money demand and the real interest rate is negative.

The fact that the real loan rate is increasing with inflation may appear puzzling because it may suggest that the demand for loans is rising faster than the supply of deposits during inflation. However, the demand for real balances actually falls with inflation. Hence, the true reason for the real loan rate to increase with inflation is because deposits (the supply of liquidity) shrink faster than the decline of credit demand. Hence, (real) liquidity reserves in the banking system dry up more quickly than credit demand when inflation increases. The same mechanism explains the liquidity trap. As inflation falls, the supply of liquidity (deposits) rises faster than the demand for loans, pushing down both the nominal and the real interest rates on credit. Under the Friedman rule, the nominal rate becomes zero and the real rate becomes identical to the Fisherian fundamentals, $\frac{1+\tilde{\imath}}{1+\pi} = \frac{1}{\beta}$. At this point there is still positive lending ($B > 0$), but the bank's liquidity reserves exceed the demand for loans despite the zero loan rate. In such a case, money injections will only raise the excess supply of liquidity further without having any impact on the demand for loans because the loan rate cannot be lowered further below zero (otherwise people will be better off by withdrawing deposits and keeping them at home).[26]

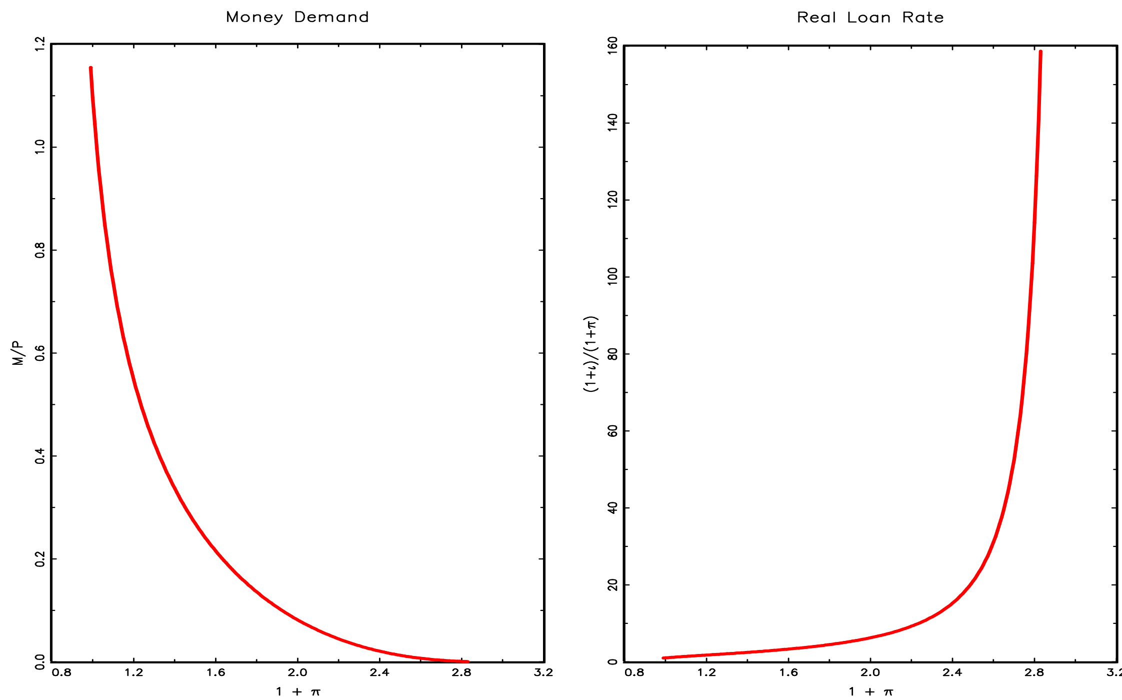

Figure 5. Money Demand and Real Interest Rate.

An important implication of the positive relationship between inflation and the real loan rate is that inflation can be potentially far more costly than realized by the literature – not only lenders but also borrowers may become losers during inflation. The lenders lose their wealth because of the inflation tax, and the borrowers lose their income because of the high real interest payments.

3.3 Short-Run Dynamics and Monetary Business Cycle

The stochastic equilibrium path of the model is solved by log-linear approximation as in the benchmark model. Under the calibrated parameter values, the model has a unique stationary saddle path. The impulse responses of the model to a one-percent transitory increase in the money stock (under policy 21 with $\rho = 0.9$) are graphed in Figure 6. It shows that transitory monetary shocks are expansionary as in the benchmark model, but with magnitudes of the real variables about six times larger. Hence, financial intermediation greatly amplifies the impact of monetary shocks through the credit lending channel. These real effects can be further amplified if the variance of idiosyncratic shocks are increased. For example, if $\sigma = 1.1$, then the real effects will be three times larger; that

[26] Notice that the nominal loan rate is identical to the Lagrangian multiplier on excess supply of liquidity (equation 74). Hence, $\tilde{\imath}$ reflects the shadow value of liquidity. As inflation rises, the excess supply falls; hence, the shadow value of liquidity increases. On the other hand, as inflation decreases, the excess supply of cash rises, driving down the shadow value of liquidity until the zero lower bound, at which point we have $M_t > B_t$, i.e., money supply exceeds loan demand.

is, a one-percent increase in money stock can raise output by about 0.6 percent, consumption by 0.3 percent, and investment by 1.4 percent. The price level remains sluggish, albeit not as "sticky" as in the benchmark model. Also, the nominal interest rate decreases for a prolonged period (see the bottom middle window in Figure 6), capturing the so-called liquidity effect found in the data (see, e.g., Christiano, Eichenbaum, and Evans, 1995).

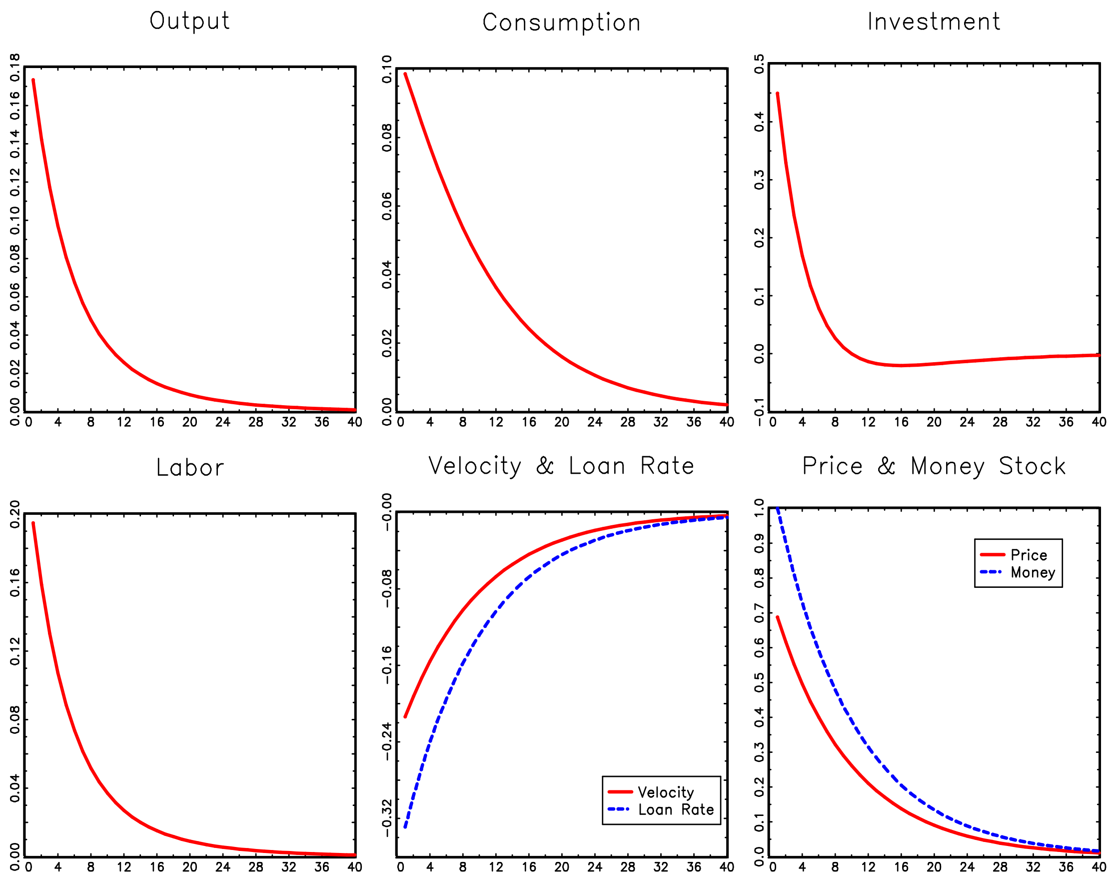

Figure 6. Impulse Responses to 1% Money Injection.

A selected set of predicted business-cycle statistics is reported in Table 1, along with their counterparts of the U.S. sample (top row) and the predictions from a standard RBC model augmented with capacity utilization (second row). For the postwar sample period, aggregate consumption is less volatile than output, which in turn is less volatile than investment. In addition, the consumption-velocity of money (defined as the ratio of aggregate consumption to $M1$) is about 1.75 times more volatile than GDP in terms of standard deviations, and the correlation between velocity and real money balances is -0.96. These stylized facts of the U.S. sample are well captured by the model driven by monetary shocks (see the third row). In particular, the model predicts that velocity should be about 1.6 times more volatile than output and its correlation with real money balances should be about -1. In addition to the dynamic behaviors of velocity, the model driven by monetary shocks can generate comparable predictions for real activities to those of a standard RBC model driven by technology shocks. For example, consumption is less volatile than output while investment is more volatile than output. The bottom row shows the predictions of the model when it is driven by technology shocks, which are similarly to the RBC model. That is, introducing money and heterogeneity into the model does not deteriorate the model's performance under technology shocks.

Table 1. Predicted Business-Cycle Moments*

	$\frac{\sigma_v}{\sigma_y}$	$\frac{\sigma_c}{\sigma_y}$	$\frac{\sigma_i}{\sigma_y}$	$\frac{\sigma_n}{\sigma_y}$	$\sigma_{v,m}$	$\sigma_{c,y}$	$\sigma_{i,y}$	$\sigma_{n,y}$	$\sigma_{y,y_{-1}}$	$\sigma_{c,c_{-1}}$	$\sigma_{i,i_{-1}}$	$\sigma_{n,n_{-1}}$	$\sigma_{v,v_{-1}}$
U.S.Data (59:1–08:4)	1.75	0.51	3.17	0.83	-0.96	0.82	0.79	0.80	0.86	0.88	0.91	0.94	0.94
RBC (cap. util.)	n.a.	0.26	4.25	0.89	n.a.	0.52	0.98	0.96	0.84	0.98	0.83	0.83	n,a,
Model (M shock)	1.58	0.78	2.10	1.10	-0.99	0.95	0.91	0.99	0.83	0.91	0.72	0.82	0.90
Model (A shock)	0	0.31	4.24	0.89	0	0.49	0.97	0.95	0.72	0.98	0.69	0.69	0

*U.S data (1959:1 - 2007:4) are real GDP, consumption of nondurable goods and services, business investment, aggregate non-farm employment, and M1. Velocity is measured as the ratio of nominal consumption to M1. All data are HP filtered. "RBC" refers to representative-agent model with capacity utilization and indivisible labor (see, e.g., King and Rebelo, 2000).

4 Conclusion

This paper extends Bewley's (1980) model to a dynamic stochastic general equilibrium setting. The generalized model is applied to study, among other things, the monetary business cycle, the welfare costs of inflation, the liquidity trap, the determination of the nominal interest rate, and the welfare gains of financial intermediation. The most important findings include (i) transitory monetary shocks have real expansionary effects despite their lump-sum nature and flexible prices; (ii) the welfare costs of inflation can be astonishingly many times larger with heterogeneous agents than with a representative agent; (iii) the liquidity trap is a natural general-equilibrium consequence of liquidity preference in a banking economy; and (iv) welfare costs of the collapse of the banking system can be potentially as large as $10 - 68\%$ of GDP, depending on the degree of heterogeneity and the inflation rate.

The neoclassical feature, especially the simplicity and analytical tractability of my model, makes it a useful framework for studying the relationship between money on the one hand and capital accumulation, asset pricing, banking and finance, international trade, exchange rate determination, wealth distribution, the business cycle and optimal government policies on the other hand. In this regard, the particular advantage of this framework is that it becomes much easier to incorporate and absorb many of the recent advances in the RBC literature and the New Keynesian literature.[27]

[27] For example, preliminary analysis shows that allowing for only a moderate degree of price stickiness in the model can generate highly persistent output movements under shocks to the money growth rate.

My analysis may shed new light on the liquidity preference theory of Keynes (1936), which is perhaps the single most controversial issue in *The General Theory.* There Keynes presents liquidity preference theory as a “liquidity theory of interest.” According to Keynes, the proper place of the theory of interest is at the level of portfolio decisions, and it is simple "arithmetic" to require that interest rates must be such that the general public’s desire to hold money be determined at the margin given the amount of liquidity the banking system decides to provide. Given the crucial importance of interest rates in affecting the real economy through asset prices and financial decisions, the key practical matter for Keynes is how deliberate monetary control can be applied to attain acceptable real performance. To sort out these problems and the possible confusions caused by Keynes himself, there is the need to formulate a liquidity preference-based model of money and banking in which currency, reserves, deposits, credit, and loans can be studied explicitly in a choice-theoretic dynamic general-equilibrium setting. For related reasons, Lucas (2000, p. 270) emphasizes that "[s]uch a model is essential if one wants to consider policies like reserve requirements, interest on deposits, and other measures that affect different components of the money stock differently." Hopefully, the model presented in this paper is a promising start in this direction.

References

[1] Aiyagari, S.R. and M. Gertler, 1991, Asset returns with transactions costs and uninsured individual risk, Journal of Monetary Economis 27(3), 311-332.

[2] Aiyagari, S.R. and S.D. Williamson, 2000, Money and dynamic credit arrangements under private information. Journal of Economic Theory 91(2), 248–279.

[3] Akyol, A., 2004, Optimal monetary policy in an economy with incomplete markets and idiosyncratic risk, Journal of Monetary Economics 51(6), 1245–1269.

[4] Alvarez, F., A. Atkeson and C. Edmond, 2008, On the sluggish response of prices to money in an inventory theoretic model of money demand, Quarterly Journal of Economics (forthcoming).

[5] Alvarez, Fernando, Andrew Atkeson and Patrick J. Kehoe.(1999). Money and Interest Rates with Endogeneously Segmented Markets, NBER Working Paper w7060.

[6] Andolfatto, D., Nosal, E., 2003. A theory of money and banking. Federal Reserve Bank of Cleveland, Working Paper 03-10.

[7] Baumol, William J. (1952). The Transactions Demand for Cash: An Inventory Theoretic Approach, Quarterly Journal of Economics. 66(4): 545-556.

[8] Berentsen, Aleksander, Camera, Gabriele and Christopher Waller, 2007, "Money, Credit and Banking," Journal of Economic Theory, 135(1), pp. 171-195.

[9] 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.

[10] Bewley, T., 1983, A difficulty with the Optimum Quantity of Money, Econometrica 51(5), 1485-1504.

[11] Blanchard, Olivier J. and Charles M. Kahn, 1980, "The Solution of Linear Difference Models under Rational Expectations," Econometrica, 48(5), pp. 1305-1311.

[12] Chatterjee, S. and D. Corbae, 1992, Endogenous Market Participation and the General Equilibrium Value of Money, The Journal of Political Economy, Vol. 100, No. 3 (Jun., 1992), pp. 615-646.

[13] Chiu, J., 2007, Endogenously Segmented Asset Market in an Inventory Theoretic Model of Money Demand, Bank of Canada Working Paper 2007-46.

[14] Chiu, J. and C. Meh, 2008, Financial internediation, liquidity and inflation. Working Paper, Bank of Canada.

[15] Christiano L., Eichenbaum M. and C. Evans (1995), Liquidity Effects, Monetary Policy, and the Business Cycle, Journal of Money, Credit and Banking 27(4), 1113-1136.

[16] Clementi, F. and M. Gallegati, 2005, Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States, in Econophysics of Wealth Distributions. Milan: Springer-Verlag Italia. ed. by Chatterjee, A., Yarlagadda, S., and Chakrabarti B. K.

[17] Clower, R.W., 1967, A reconsideration of the microfoundations of money, Western Economic Journal 6(4), 1-9.

[18] Cooley, T. and G. Hansen, 1989, The inflation tax in a real business cycle model, American Economic Review 79(4), 733-748.

[19] Fermi, D. (Ed.), QUID (Paris: RTL, 1998).

[20] Grandmont, J.M. and G. Laroque, 1976, The liquidity trap, Econometrica, Vol. 44 (1), 129-135.

[21] Greenwood, J., Z. Hercowitz, and G. Huffman, 1988, Investment, capacity utilization, and the real business cycle, American Economic Review 78(3), 402-417.

[22] Greenwood, J. and S. Williamson, 1989, International financial intermediation and aggregate fluctuations under alternative exchange rate regimes, Journal of Monetary Economics 23(3), 401-431.

[23] Imrohoroglu, A., 1992, The welfare cost of inflation under imperfect insurance, Journal of Economic Dynamics and Control 16(1), 79-91.

[24] Keynes, John M., 1930, Treatise on Money. (London: Macmillan).

[25] Keynes, John M., 1936, The General Theory of Employment, Interest and Money. London: Macmillan.

[26] Khan, Aubhik and Julia K. Thomas, 2007, Inflation and Interest Rates with Endogenous Market Segmentation, Federal Reserve Bank of Philadelphia working paper 07-1.

[27] King, Robert G., Plosser, Charles I., and Sergio T. Rebelo, 1988, "Production, Growth, and Business Cycles: I. The Basic Neoclassical Model," Journal of Monetary Economics, 21(2-3), pp. 195-232.

[28] King, R.G. and S. Rebelo, 2000, Resuscitating Real Business Cycles, NBER Working Paper No. W7534.

[29] King, R.G. and J. Thomas, 2007, Breaking the New Keynesian Dichotomy: Asset Market Segmentation and the Monetary Transmission Mechanism, Working Paper.

[30] Lagos, R. and R. Wright, 2005, A Unified Framework for Monetary Theory and Policy Analysis, Journal of Political Economy 113(3), 463-484.

[31] Levy, M. and H. Levy, 2003, Investment talent and the Pareto wealth distribution: Theoretical and empirical analysis, The Review of Economics and Statistics 85(3), 709-725.

[32] Liu, L.Q., L. Wang, and R. Wright, 2008, The "Hot Potato" Effect of Inflation, Working Paper, University of Pennsylvania.

[33] Lucas, R.E. Jr., 1980, Equilibrium in a Pure Currency Economy, Economic Inquiry 18(2), 203-220.

[34] Lucas, R.E. Jr., 2000, Inflation and welfare, Econometrica, Vol. 68(2), 247-274.

[35] Nirei, M. and W. Souma, 2007, A two factor model of income distribution dynamics, Review of Income and Wealth 53(3), 440-459.

[36] Sidrauski, M, 1967, "Rational Choice and Patterns of Growth in a Monetary Economy," American Economic Review, 57(2), 534 544.

[37] Svensson, Lars, E.O., 1985, "Money and Asset Prices in a Cash-in-Advance Economy," Journal of Political Economy, 93(5), pp. 919-944.

[38] Tobin, J., 1956, The Interest-Elasticity of the Transactions Demand for Cash, Review of Economics and Statistics. 38(3): 241-247.

[39] Wolff, Edward, 1996, Trends in Household Wealth during l989-1992, submitted to the Department of Labor (New York University).

[40] Wen, Y., 2008, Input and output inventory dynamics, Federal Reserve Bank of St. Louis Working Paper 2008-008B.

[41] Williamson, S., 1986. Costly monitoring, financial intermediation, and equilibrium credit rationing. Journal of Monetary Economics 18(2), 159–179.