working paper 11 03

The Cost of Inflation: A Mechanism Design Approach

Guillaume Rocheteau

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 11-03 January 2011

The Cost of Inflation: A Mechanism Design Approach

Guillaume Rocheteau

I apply mechanism design to quantify the cost of inflation that can be attributed to monetary frictions alone. In an environment with pairwise meetings, the money demand that is consistent with a constrained-efficient allocation takes the form of a continuous correspondence that can fit the data over the period 1900–2006. For such parameterizations, the cost of moderate inflation is zero. This result is robust to different assumptions regarding the observability of money holdings, the introduction of match-specific heterogeneity, and endogeneous participation decisions.

JEL Classification: D82, D83, E40, E50.

Keywords: Cost of inflation, pairwise trades, optimal mechanism.

Guillaume Rocheteau is a professor of economics at the University of California, Irvine, and a research associate at the Federal Reserve Bank of Cleveland. He can be reached at grochete@uci.edu. He thanks Nicolas Jacquet, Daniel Sanches, Serene Tan, and seminar participants at the Bank of Canada and the University of California at Irvine for useful comments and discussions. He is grateful to Peter Ireland, who shared his data on U.S. money demand. He thanks Monica Crabtree-Reusser for editorial assistance.

1 Introduction

A classical topic in monetary economics is measuring the burden that inflation imposes on society. The standard methodology, pioneered by Bailey (1956) and Friedman (1969) and reviewed in Lucas (2000), consists of estimating a reduced-form money demand function and measuring the welfare cost of inflation as the area underneath money demand.[1] The rationale for this approach is based on competitive general equilibrium models where money enters the utility function, or a cash-in-advance constraint.[2] Unfortunately, as pointed out by Wallace (2001), such models contain hidden inconsistencies and they are ill-suited for normative analysis as they fail to account for the social benefits that monetary exchange provides to the economy. To illustrate quantitatively the importance of this critique, Lagos and Wright (2005) – denoted LW hereafter – calibrate a model with microfoundations for monetary exchange and provide estimates for the annual cost of 10 percent inflation. Their estimates are multiple times larger than those of standard reduced-form monetary models, up to 5 percent of GDP. This result led Williamson and Wright (2010) in their review article on New Monetarist Economics to conclude that "the intertemporal distortion induced by inflation may be more costly than many economists used to think."

The quantitative insights of LW, however, are subject to the caveat of Hu, Kennan, and Wallace (2009), regarding the trading mechanisms that are typically assumed when measuring the welfare cost of inflation. The problem is that these trading mechanisms are socially inefficient, which raises the concern that the large welfare costs of inflation are not due to the frictions that make money essential but to the adoption of inefficient mechanisms (e.g., the Nash bargaining solution). It is for this reason and others – namely, to establish the essentiality of money and the robustness of policy prescriptions – that Wallace (2001, 2010) recommends that monetary theory be pursued by applying mechanism design. The objective of this paper is to do precisely that, i.e., to use mechanism design to determine the part of the welfare cost of inflation that can be attributed to monetary frictions alone – the *irreducible* cost of inflation. I will derive the money demand that is part of a constrained-efficient allocation, check whether the model can fit the data, and measure

[1] For an application of this method to the recent behavior of U.S. money demand, see Ireland (2009).

[2] For recent examples of general equilibrium models with cash-in-advance constraints or money-in-the-utility-function, see Dotsey and Ireland (1996), Burstein and Hellwig (2008) and da Costa and Werning (2008).

the cost of inflation.[3]

I show that the money demand generated by the LW model under socially optimal mechanisms takes the form of a correspondence. Below a threshold for the inflation rate, there is an interval of real balances that are consistent with the implementation of the first-best allocation, and the measure of this interval shrinks with inflation. Above a threshold for the inflation rate money demand is a singleton, and real balances and welfare are decreasing with the inflation rate.

When calibrated to fit the data, based on the methodology of Lucas (2000), I find parametrizations such that all the annual observations in the data for the U.S. over 1900-2006 are consistent with the model. For such parameterizations the welfare cost of 10 percent inflation is 0. Thus, for plausible calibrations of the LW model moderate inflation generates no burden for society when the only frictions in the environment are the ones that make money essential. This result turns on its head the prevailing wisdom that monetary environments generate large costs of inflation: they only do so to the extent that suboptimal mechanisms are employed. It also confirms that the standard approach to estimating the cost of inflation, the area underneath money demand, has dubious foundations.

I check the robustness of the results to different extensions. For instance, I consider different assumptions regarding the observability of agents' money holdings. I also introduce match-specific heterogeneity (idiosyncratic preference shocks) and private information. Finally, I refine money demand by introducing a participation decision that endogenizes the frequency at which agents trade. This extension provides a natural way to pin down the transfer of real balances in bilateral matches. The model generates a downward-sloping money demand consistent with the data and, for some parametrizations, the cost of inflation is zero. This is in contrast to Shi (1997) and Rocheteau and Wright (2009), who showed that under bargaining, social welfare can be nonmonotonic with inflation and small inflation can be beneficial, thereby justifying departures from the Friedman rule.

There is a growing literature, surveyed in Craig and Rocheteau (2008), measuring the welfare cost of inflation in the context of microfounded models of monetary exchange under different trading mechanisms.[4] In contrast to this literature I endogenize the trading mechanism in decentralized

[3] In the context of the labor market, the idea of imposing a socially optimal mechanism when calibrating a search-theoretic model can be found in Shimer (2005).

[4] Rocheteau and Wright (2009) compute the welfare cost of inflation under different mechanisms and with endogenous participation decisions. Reed and Waller (2006) introduce a risk-sharing motive. Aruoba, Waller, and Wright (2010) calibrate versions of the model with capital. Aruoba and Chugh (2010) show that the Friedman rule

markets so that it implements a constrained-efficient allocation.[5] Relative to Hu, Kennan, and Wallace (2009), I characterize money demand and calibrate the model under different assumptions regarding the observability of money holdings. I introduce match-specific heterogeneity and private information, and endogenous participation decisions to refine money demand.

The paper is organized as follows. The environment is described in Section 2. The optimal mechanism and the money demand correspondence are characterized in Section 3. The model is calibrated in Section 5. Match-specific heterogeneity and endogenous participation decisions are introduced in Sections 6 and 7, respectively.

2 The environment

The environment is similar to the one in Lagos and Wright (2005). Time is discrete and continues forever. There is a continuum of infinitely-lived agents with measure one. Each period is divided into two stages. In the first stage agents trade in a decentralized market with pairwise meetings, denoted DM, while in the second stage they trade in a centralized market, denoted CM.

In the DM, an agent is either a buyer, with probability $n \in (0,1)$, or a seller, with probability $1-n$. Up to Section 7 n is exogenous, while in Section 7 the composition of the market is endogenous. Buyers and sellers are matched bilaterally: a buyer meets a seller with probability α_b, while a seller meets a buyer with probability α_s, with $n\alpha_b = (1-n)\alpha_s$. In the CM, agents, who are price-takers, trade a perishable good, called the numéraire good, labor and money.

Agents' preferences are represented by the following utility function:

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t \mathcal{U}(q_t, e_t, c_t, h_t),$$

is not optimal, and the long-run capital income tax is not zero. Boel and Camera (2009) extend the model to obtain an equilibrium dispersion in wealth and show that the impact of inflation varies across segments of society. Boel and Camera (2010) compute the cost of inflation across OECD countries. Chiu and Molico (2010) introduce costly liquidity management and show that the cost of inflation is significantly lower than previous estimates, thanks to redistributional effects.

[5] Kocherlakota (1998) and Kocherlakota and Wallace (1998) were the first to use implementation theory to prove the essentiality of money. Applications of mechanism design to monetary theory include Cavalcanti and Wallace (1999) and Mattesini, Monnet, and Wright (2010) on banking and inside money, Cavalcanti and Erosa (2008) on the propagation of shocks in monetary economies, Cavalcanti and Nosal (2009) on cyclical monetary policy, Koeppl, Monnet, and Temzelides (2008) on settlement, and Deviatov and Wallace (2001) and Deviatov (2006) on the welfare gains of money creation. A related, but partial equilibrium, analysis was provided in Berentsen and Rocheteau (2002) in the context of a model with divisible money.

where $\beta \equiv (1+r)^{-1} \in (0,1)$ is the discount factor, q_t is DM consumption, e_t is the DM level of effort, c_t is CM consumption, and h_t is the supply of hours in the CM. For tractability, $\mathcal{U}$ is additively separable and linear in hours, $\mathcal{U}(q,e,c,h) = u(q) - \psi(e) + U(c) - h$. The technology in the DM is such that $q = e$. The utility function is well-behaved, and $q^* = \arg\max[u(q) - \psi(q)]$. I also assume without loss in generality that $u(0) = \psi(0) = 0$. Output in the CM is produced according to a linear production function in labor, which implies the (real) wage rate is equal to 1.

All goods are perishable across both stages and time. Agents cannot commit to future actions, and individual histories are private information. These assumptions rule out credit arrangements and generate an essential role for money. The quantity of fiat money per capita at the beginning of period t is $M_t > 0$, with $M_{t+1} = \gamma M_t$. The money growth rate, $\gamma \equiv 1 + \pi$, is constant and new money is injected by lump-sum transfers (or taxes if $\gamma < 1$) in the CM.[6] I will not impose that the money growth rate is chosen optimally since my focus is on socially optimal trading arrangements under different inflation rates. The price of goods in terms of money in the CM is denoted p_t.

Agents' money holdings in a match are not observable: an agent can hide his money holdings or overstate them.[7] This assumption limits the ability of the mechanism to punish the seller in a bilateral match who does not hold sufficient real balances. Also, it will be consistent with the definition of money that includes demand and checkable deposits, $M1$, when I calibrate the model.

3 Constrained-efficient allocations

I first consider a version of the model in which each agent receives an idiosyncratic shock at the beginning of the DM that determines whether he is a buyer (he wants to consume but cannot produce) with probability n, or a seller (he can produce but does not want to consume) with probability $1-n$. I set $n = 1/2$, so that each agent is equally likely to be a buyer or a seller in the DM, and $\alpha = \alpha_b = \alpha_s$, which is implied by bilateral matching, and denote $\sigma = \alpha/2$.

The terms of trade in the DM are determined according to the following game. In the first stage the buyer and the seller announce simultaneously their real balances, z^b and z^s, respectively.

[6] In the case where $\pi < 0$, the government has the power to impose infinite penalties on agents who do not pay taxes. The government, however, does not have the technology to monitor DM and CM trades and cannot observe agents' real balances. Hu, Kennan, and Wallace (2009) and Andolfatto (2010) study the case where agents can choose whether or not to participate in the CM and pay taxes.

[7] In Section 4 I determine the set of incentive-feasible allocations with unobservable money holdings.

A mechanism in the DM, $[q,d]:\mathbb{R}_{2+}\to\mathbb{R}_{+}\times\mathbb{R}_{+}$, maps the announced real money holdings of the buyer and the seller in a proposed allocation, $(q,d)\in\mathbb{R}_{+}\times\left[-z^{s},z^{b}\right]$, where q is the quantity produced by the seller and consumed by the buyer and d is a transfer of real balances from the buyer to the seller. The allocation is restricted to the pairwise core of the meetings in the DM.[8] In the second stage of the game the buyer and the seller simultaneously say "yes" or "no" to the proposed allocation. If they both say "yes," and if the transfer of money is feasible given the buyer's actual money holdings, then the trade takes place. Otherwise, there is no trade. This second stage guarantees that both agents can walk away from the proposed trade.

I consider stationary, symmetric allocations. Such an allocation is defined by a triple (q^{p},d^{p},z^{p}), where (q^{p},d^{p}) is the trade in all matches in the DM and z^{p} is agents' real balances or, equivalently, the production of the CM good by agents not holding money at the beginning of the CM. By the clearing of the money market in the CM, $M_{t}/p_{t}=z^{p}$, which implies $p_{t+1}=\gamma p_{t}$.

Given a mechanism, $[q,d]$, Bellman's equation for an agent in the DM holding $z=m/p$ units of real balances is

$$V(z)=\sigma\left\{u\left[q(z,z^{p})\right]+W\left[z-d\left(z,z^{p}\right)\right]\right\}+\\ \sigma\left\{-\psi\left[q\left(z^{p},z\right)\right]+W\left[z+d\left(z^{p},z\right)\right]\right\}+\left(1-2\sigma\right)W(z), \quad (1)$$

where $W(z)$ is the value function of the agent in the CM. Equation (1) has the following interpretation. An agent is a consumer who meets a producer with probability σ. He consumes q units of goods and delivers d units of real balances (expressed in terms of CM output) to his trading partner. The terms of trade (q,d) depend on the (truthfully) announced real balances of the buyer and the seller in the match. With probability σ, the agent is a producer who meets a consumer. He produces q for his trading partner and receives d real balances. With probability $1-2\sigma$, no trade takes place.

The CM problem of the agent is

$$W(z)=\max_{c,\hat{z}}\left\{U(c)-c+T+z-\gamma\hat{z}+\beta V(\hat{z})\right\}, \quad (2)$$

[8] Zhu (2008) proposes a coalition-proof game that guarantees that any trade in the DM is in the pairwise core. This game works as follows. First, the buyer and the seller in the match announce simultaneously their real money holdings. Second, an allocation that depends on the announced money holdings is proposed. The buyer and the seller simultaneously accept or reject the proposed allocation. If it is rejected by one of the two players, the game ends. Otherwise, the seller makes a counterproposal. Third, the buyer can choose which trade is carried out, the seller's counteroffer or the initial offer.

where T is the lump-sum transfer (expressed in numéraire goods), and $\hat{z}$ is the real balances taken into the next DM. I have used the budget constraint according to which the CM supply of hours is $h = c + \gamma\hat{z} - z - T$ and the relative price of real balances next period in terms of current-period CM output is $p_{t+1}/p_t = \gamma$. From (2), the maximizing choice of $\hat{z}$ is independent of z; and W is linear in z, with $W_z = 1$.

Substituting $V(\hat{z})$ by its expression given by (1), using the linearity of $W(z)$, and ignoring the constant terms, one can reformulate the agent's problem in the CM as

$$\max_{z\geq 0}\left\{-\gamma z + \beta\left\{\sigma\left\{u\left[q(z,z^p)\right] - d\left(z,z^p\right)\right\} + \sigma\left\{d\left(z^p,z\right) - \psi\left[q\left(z^p,z\right)\right]\right\} + z\right\}\right\}.$$

Divide the previous expression by β and denote $i \equiv (1+\pi)(1+r) - 1$, which can be interpreted as the nominal interest rate that would be paid on an illiquid bond, to get:

$$\max_{z\geq 0}\left\{-iz + \sigma\left\{u\left[q(z,z^p)\right] - d\left(z,z^p\right)\right\} + \sigma\left\{d(z^p,z) - \psi[q(z^p,z)]\right\}\right\}. \tag{3}$$

Given a mechanism, $[q(\cdot,\cdot), d(\cdot,\cdot)]$, a seller reports the level of real balances that maximizes his expected surplus, taking as given that the buyer will report his own real balances truthfully, $\hat{z}^s \in \arg\max_z \left\{d(z^p,z) - \psi[q(z^p,z)]\right\} = d^p - \psi\left(q^p\right)$. A necessary condition for the mechanism to be incentive-compatible is that the seller's expected surplus is independent of his announced real balances. Consequently, for any incentive-compatible mechanism, the choice of real balances, (3), can be reexpressed as

$$\max_{z\geq 0}\left\{-iz + \sigma\left\{u\left[q(z,z^p)\right] - d\left(z,z^p\right)\right\}\right\}. \tag{4}$$

The optimal choice of real balances maximizes the expected surplus of a buyer in the DM net of the cost of holding real balances.

While a seller can overstate his real balances without fear of negative consequences, the same is not true for a buyer. If a buyer holding z^b announces $\hat{z}^b$ and it turns out that $d\left(\hat{z}^b, z^s\right) > z^b$ then the trade is not feasible and cannot be carried out. The optimal announcement of a buyer who holds z^b is then $\hat{z}^b \in \arg\max_z \left\{u\left[q(z,z^p)\right] - d\left(z,z^p\right)\right\}\mathbb{I}_{\{d(z,z^p)\leq z^b\}}$, where $\mathbb{I}_{\{d\leq z^b\}}$ is an indicator function that is equal to one if $d \leq z^b$.

Given that money holdings are unobservable, agents will not hold more money than what they intend to spend, $z^p = d^p$. From (4), a necessary condition for the allocation to be incentive feasible is

$$-id^p + \sigma\left[u\left(q^p\right) - d^p\right] \geq 0. \tag{5}$$

The left side of (5) is the expected surplus of a buyer in the DM, net of the cost of holding real balances according to the proposed allocation. A deviation that is always feasible consists of not accumulating money in the CM and not trading as a buyer in the DM. The expected payoff associated with this defection is 0. The allocation must also satisfy the seller's participation constraint,

$$-\psi\left(q^p\right)+d^p\geq 0. \tag{6}$$

There is a similar condition for buyers, $u\left(q^p\right)-d^p\geq 0$, but it is implied by (5) and $d^p\geq 0$.

At this point it is useful to characterize the pairwise core of a meeting between a buyer holding z^b real balances and a seller holding z^s real balances. The pairwise core, denoted $\mathcal{C}(z^b,z^s)$, is the set of all feasible allocations, $(q,d)\in\mathbb{R}_+\times\left[-z^s,z^b\right]$, such that no alternative feasible allocations exist that would make the buyer and the seller in the match better off, with at least one of the two being strictly better off. Formally:

$$\begin{aligned}\mathcal{C}(z^b,z^s) &= \Big\{(q,d)\in\arg\max\left[u(q)-d\right]\ \text{s.t.}\ d\in\Big[-z^s,z^b\Big]\\ &\qquad\qquad\text{and}\ -\psi(q)+d\geq U^s\ \text{for some}\ U^s\geq 0\Big\}.\end{aligned}$$

This gives:

$$\begin{aligned}\mathcal{C}(z^b,z^s) &= \{q^*\}\times[\psi(q^*),u(q^*)]\ \ \text{if}\ \ z^b\geq u(q^*)\\ &= \{q^*\}\times\Big[\psi(q^*),z^b\Big]\cup\Big[u^{-1}(z^b),q^*\Big]\times\{z^b\}\ \ \text{if}\ \ z^b\in[\psi(q^*),u(q^*))\\ &= \Big[u^{-1}(z^b),\psi^{-1}(z^b)\Big]\times\{z^b\}\ \ \text{if}\ \ z^b<\psi(q^*).\end{aligned}$$

If the buyer's real balances are larger than the amount he is willing to pay for the first-best level of output, $u(q^*)$, then any allocation in the pairwise core implements the efficient level of output, and the transfer of real balances is between the seller's cost and the buyer's willingness to pay. If the buyer's real balances are less than his willingness to pay for the first-best level of output, $u(q^*)$, but greater than the seller's cost, $\psi(q^*)$, then the first-best allocation is achieved provided that the seller's surplus is not too large; otherwise, the buyer transfers all of his real balances, and output is less than the efficient level. Finally, if the buyer's real balances are not large enough to compensate the seller for the cost of producing the first-best level of output, then any allocation in the pairwise core is such that the buyer transfers all his real balances, and the output level is inefficiently low.

Lemma 1 *Any allocation* (q^p, d^p, z^p) *such that* $z^p = d^p$ *and* $(q^p, d^p) \in \mathcal{C}(z^p, z^p)$ *that satisfies (5) and (6) can be implemented by the following coalition-proof trading mechanism:*

$$\left[q(z^b, z^s), d(z^b, z^s)\right] = \arg\max_{q,d \leq z^b} [d - \psi(q)] \quad s.t. \quad u(q) - d \geq u(q^p) - d^p \quad if \quad z^b \geq d^p, \tag{7}$$

$$= \arg\max_{q,d \leq z^b} [d - \psi(q)] \quad s.t. \quad u(q) - d = 0 \quad otherwise. \tag{8}$$

According to (7) if the buyer holds more than d^p real balances, then the mechanism specifies a pairwise Pareto-efficient allocation that gives the buyer a surplus that is at least equal to what he would obtain under the trade (q^p, d^p). According to (8) if the buyer holds less than d^p real balances, then the mechanism chooses the allocation that maximizes the seller's surplus subject to the buyer being indifferent between trading or not trading. Figure 1 represents graphically the mechanism in (7)-(8). The buyer's surplus is $U^b = u(q) - d$, while the seller's surplus is $U^s = -\psi(q) + d$. The pairwise core (in the utilities space) is downward-sloping and concave. The utility levels associated with the proposed trade, (q^p, d^p), are denoted $\bar{U}^b$ and $\bar{U}^s$. If the buyer holds more than z^p real balances, then the Pareto frontier shifts outward. The mechanism selects the point on the Pareto frontier marked by a circle that assigns the same utility level, $\bar{U}^b$, to the buyer. If the buyer holds less than z^p real balances, the Pareto frontier shifts downward. The mechanism selects the point on the frontier that assigns no utility to the buyer, $U^b = 0$.

The proof of Lemma 1 is contained in Figure 2. The top panel of Figure 2 represents the buyer's surplus as a function of his real balances. The buyer's surplus is (weakly) monotonically increasing in his real balances, which implies that the buyer has no incentive to hide some of his money holdings. He has no incentive to overstate his real balances either since for all $z \geq d^p$ the buyer's surplus is constant and if the buyer holds less than d^p but reports $\hat{z} \geq d^p$ then the trade is not feasible. The bottom panel represents the buyer's surplus net of the cost of holding real balances. From (3) and the bottom panel of Figure 2 it is easy to check that the agent will choose $z = d^p$ if (5) holds.

I define an optimal mechanism as a trading mechanism described in Lemma 1 that maximizes society's welfare, denoted $\mathcal{W}$.

Definition 1 *An optimal mechanism is a* $[q(\cdot,\cdot), d(\cdot,\cdot)]$ *defined by (7)-(8), where* $(q^p, d^p) \in \mathcal{C}(d^p, d^p)$

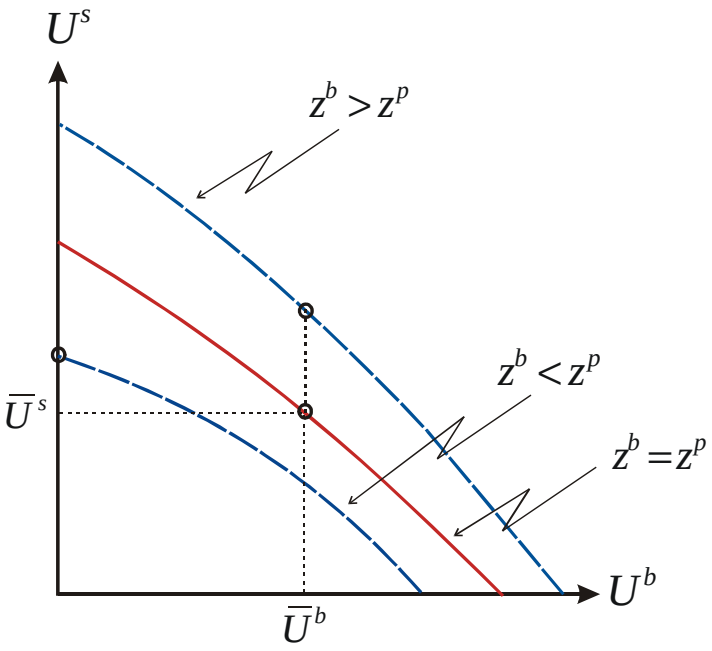


Figure 1: Implementation of coalition-proof trades

solves:

$$(q^p, d^p) \in \arg\max_{q,d} \mathcal{W} = \sigma\left[u(q) - \psi(q)\right] + U(c^*) - c^* \tag{9}$$

$$s.t. \quad -\psi(q) + d \geq 0 \tag{10}$$

$$-id + \sigma\left[u\left(q\right) - d\right] \geq 0. \tag{11}$$

The solution to (9)-(11) is in the pairwise core; otherwise, $q^p > q^*$ and one could reduce both q and d so as to increase the seller's surplus in (10), the buyer's expected surplus net of the cost of holding money in (11), and the whole match surplus in (9). The optimal mechanism proposes the DM allocation that maximizes the period expected utility of a representative household subject to the individual rationality constraints of the seller, (10), and the buyer, (11). It corresponds to the highest $q \leq q^*$ so that (10) and (11) hold. The solution is

$$(q^p, d^p) \in \{q^*\} \times \left[\psi(q^*), \frac{\sigma}{i+\sigma}u(q^*)\right] \quad \text{if} \quad \psi(q^*) \leq \frac{\sigma}{i+\sigma}u(q^*) \tag{12}$$

$$\in \{q(i)\} \times \{\psi\left[q(i)\right]\} \text{ otherwise,} \tag{13}$$

where $q(i)$ is the positive solution to $\psi(q) = \frac{\sigma}{i+\sigma}u(q)$. So while the output level is uniquely deter-

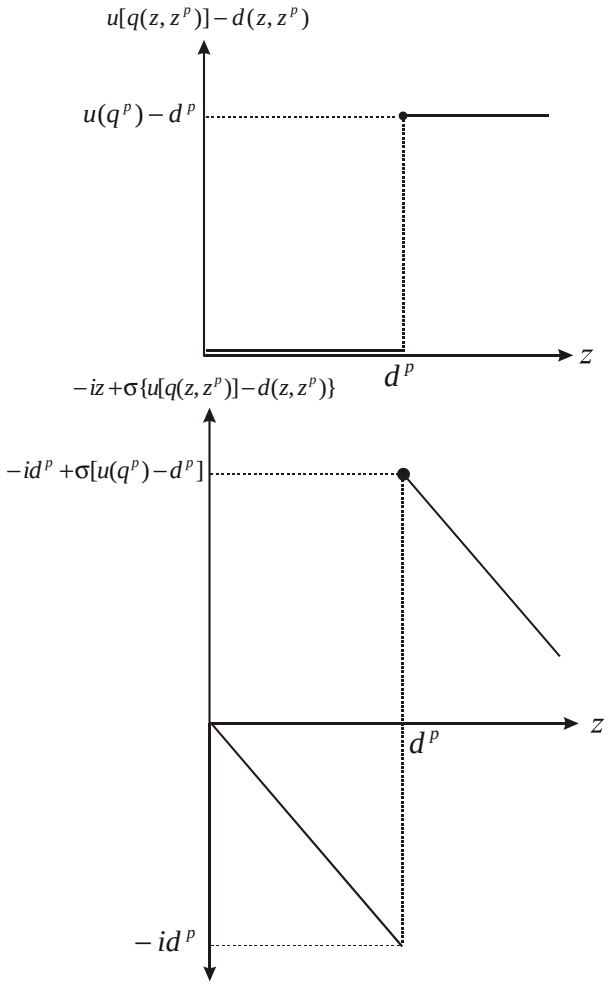


Figure 2: An incentive-feasible mechanism

mined, the transfer of real balances is not always unique. If the first-best level of output is feasible, $q^p = q^*$, then there is a range of real balances that are incentive feasible. This is simply saying that provided that an agent's participation constraint in the CM is not binding, there are many ways one can split the surplus of a bilateral match, $u(q^*) - \psi(q^*)$. In contrast, when the agent's participation constraint in the CM binds, then output and real balances are uniquely determined.

The solution to (9)-(11) is represented in Figure 3. If $\psi(q^*) \leq \frac{\sigma}{i+\sigma}u(q^*)$ then $q = q^*$ and there is an interval of real balances, between the curves $\psi(q)$ and $\frac{\sigma}{i+\sigma}u(q)$, that are consistent with the first-best allocation. If $\psi(q^*) > \frac{\sigma}{i'+\sigma}u(q^*)$, then the first-best allocation is not implementable and the quantity traded is $q(i') < q^*$ at the intersection of $\psi(q)$ and $\frac{\sigma}{i'+\sigma}u(q)$. The level of real balances

is $\psi[q(i')]$.

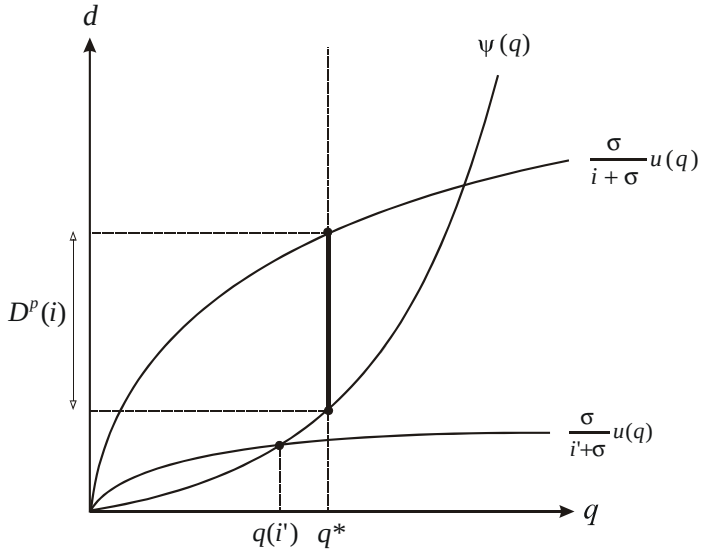


Figure 3: Constrained-efficient allocations

Denote $D^p(i)$ the money demand correspondence defined as:

$$\begin{aligned} D^p(i) &= \left[\psi(q^*), \frac{\sigma}{i+\sigma}u(q^*)\right] \text{ if } \psi(q^*) \leq \frac{\sigma}{i+\sigma}u(q^*) \\ &= \{\psi[q(i)]\} \text{ otherwise.} \end{aligned}$$

It specifies the set of real balances that are consistent with an optimal mechanism for a given inflation rate. The next proposition characterizes how money demand, output, and welfare vary with inflation.

Proposition 1 *There is*

$$\bar{\imath} = \frac{\sigma\left[u(q^*) - \psi(q^*)\right]}{\psi(q^*)} > 0 \tag{14}$$

such that

1. *For all* $i \in [0, \bar{\imath})$, $q^p(i) = q^*$, $\frac{\partial \mathcal{W}}{\partial i} = 0$, *and* $D^p(i) \subset D^p(i')$ *for all* $i' < i$;

2. *For all* $i > \bar{\imath}$, $\frac{\partial q^p}{\partial i} < 0$, $\frac{\partial d^p}{\partial i} < 0$, *and* $\frac{\partial \mathcal{W}}{\partial i} < 0$.

The quantity $\bar{\imath}$ is the highest nominal interest rate below which the first-best level of output is incentive-feasible. The right side of (14) can be interpreted as the expected nonpecuniary rate of

return of money. It is the probability that an agent is a buyer in the DM times the first-best surplus of a match expressed as a fraction of the cost to produce the first-best level of output. So the larger the nonpecuniary rate of return of money, the larger the range of inflation rates consistent with the first-best allocation. The first part of Proposition 1 also shows that money is super-neutral for low inflation rates, and money demand is decreasing in the sense that the set of implementable real balances at higher inflation rates is contained in the set of implementable real balances at lower inflation rates. See Figure 4 for an illustration of the money demand correspondence. When the buyer's participation constraint binds, $i > \bar{\imath}$, the nonpecuniary rate of return of currency evaluated at the first-best level of output is less than the cost of holding currency: so the first best is not implementable. In this case, money demand is a singleton, and the output produced and consumed in a match, social welfare, and the transfer of real balances are decreasing with the inflation rate.

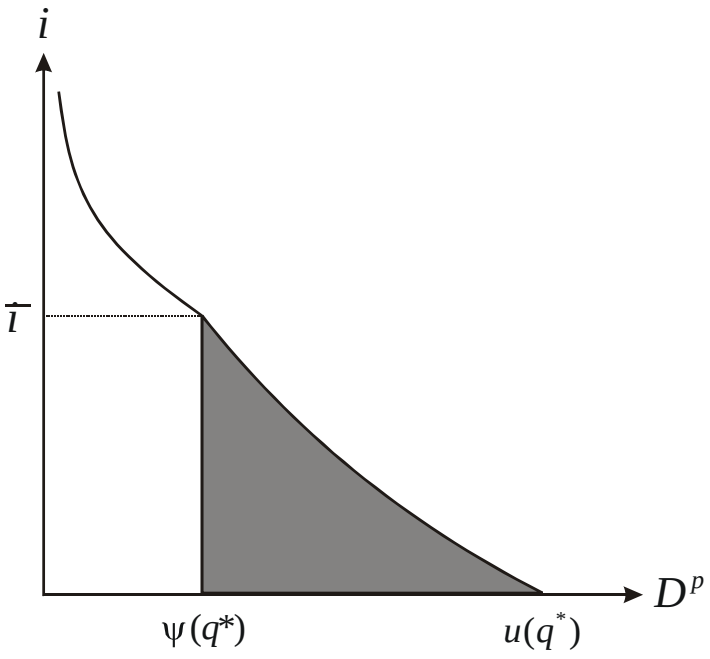


Figure 4: Money demand correspondence

4 The observability of money holdings

Throughout the paper I assume that money holdings are not observable, i.e., agents can both hide and overstate their money balances. In contrast, Hu, Kennan, and Wallace (2009) assume that

while agents can hide their money balances, they cannot overstate them. Before turning to the calibration of the model, it is useful to discuss the importance played by the nonobservability of money balances. To do this, I derive the set of stationary, symmetric, incentive-feasible allocations in the case where money holdings cannot be overstated.[9]

One crucial element to determine the set of implementable allocations is a necessary condition under which the deviation that consists of not accumulating money in the CM is not profitable. In the case where money holdings cannot be overstated, such a necessary condition takes the form $W(0) \geq \beta W(0)$, i.e.,

$$-iz^p + \sigma \left[u\left(q^p\right) - \psi(q^p) \right] \geq 0. \tag{15}$$

In contrast to (5), an agent who deviates in the CM and accumulates no money can no longer secure the surplus $-\psi(q^p) + d^p$ in the subsequent DM by overstating his money balances.[10] Indeed, the mechanism can potentially punish a seller who holds no money by assigning no surplus to this seller. A deviation that is always feasible, however, consists of not accumulating money in the CM and not trading in the subsequent DM. The expected surplus net of the cost of holding money from this deviation is zero. It should be emphasized from (15) that the money holdings of an agent, z^p, need not coincide with the transfer in the DM, d^p.

Lemma 2 *Any allocation (q^p, d^p, z^p) such that $(q^p, d^p) \in \mathcal{C}(d^p, d^p)$ that satisfies $d^p \leq z^p$, (15) and (6) can be implemented by the following coalition-proof trading mechanism:*

$$\begin{aligned}
(q, d) &= \arg \max_{q, d \leq z^b} \left[d - \psi(q) \right] \quad s.t. \quad u(q) - d \geq u(q^p) - d^p \quad if \quad \min(z^b, z^s) \geq z^p & (16) \\
&= \arg \max_{q, d \leq z^b} \left[d - \psi(q) \right] \quad s.t. \quad u(q) - d = 0 \quad if\ z^b < z^p & (17) \\
&= \arg \max_{q, d \leq z^b} \left[u(q) - d \right] \quad s.t. \quad -\psi(q) + d = 0 \quad if\ z^s < z^p \ and\ z^b \geq z^p. & (18)
\end{aligned}$$

According to (16), if both the seller and the buyer in a bilateral match hold (and announce) more than z^p units of real balances, then the trade is the allocation in the pairwise core that generates the same surplus for the buyer as the one he would have obtained under (q^p, d^p). According to (17), if the buyer holds less than z^p, then the mechanism proposes the preferred trade of the seller in the

[9] For instance, agents could choose to bring only a fraction of their money holdings to a bilateral match in the DM.

[10] For readers familiar with the literature, the analysis of the set of implementable allocations in Hu, Kennan, and Wallace (2009) is erroneous, as they impose (5) as a necessary condition for an implementable allocation instead of (15), which is the relevant condition when money holdings cannot be overstated.

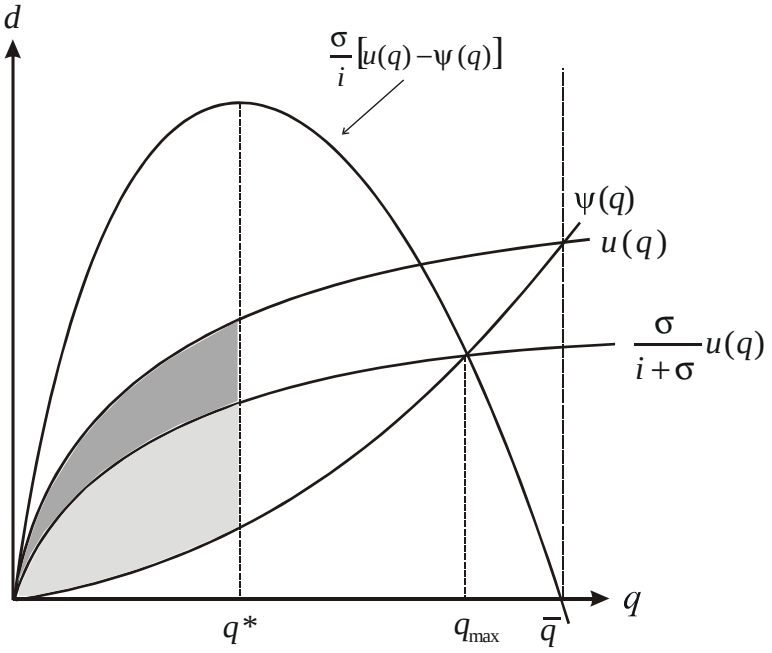


Figure 5: Implementable allocations under observable (dark and light grey) and nonobservable (light grey) money holdings.

pairwise core. Symmetrically, according to (18), if the seller holds less than z^p and the buyer holds at least z^p, then the mechanism proposes the preferred trade of the buyer in the pairwise core.

The proof of Lemma 2 goes as follows. By construction, the buyer and the seller have incentives to report truthfully their money holdings since their surpluses are nondecreasing with their money holdings. If an agent believes that all his potential partners adhere to the equilibrium play and hold z^p units of real balances, then he has no incentives to deviate and hold less than z^p since otherwise, from (17) and (18), he would receive no surplus in the DM irrespective of whether he turns into a buyer or a seller. He has no incentive to hold more than z^p either since from (16) his expected surplus in the DM is $u(q^p) - \psi(q^p)$, which is independent of his real balances (provided they are larger than z^p).

The set of stationary, symmetric, incentive-feasible allocations when agents cannot overstate their money holdings is

$$\mathcal{A}^o(i) \equiv \left\{(q,d,z) : (q,d) \in \mathcal{C}(z,z),\ \psi(q) \leq d \leq z \leq \frac{\sigma\left[u\left(q\right)-\psi(q)\right]}{i},\ d \leq u(q)\right\}.$$

Under unobservable money holdings,

$$\mathcal{A}^u(i) \equiv \left\{ (q,d,z) : (q,d) \in \mathcal{C}(z,z),\ \psi(q) \leq d = z \leq \frac{\sigma}{i+\sigma} u(q) \right\}.$$

The sets $\mathcal{A}^o$ and $\mathcal{A}^u$ are represented in Figure 5. Under unobservable money holdings, all the pairs (q,d) in the light grey area are implementable. If money holdings holding cannot be overstated, the pairs (q,d) in the light and dark grey areas can be implemented. Therefore, under the assumption of unobservable money holdings, the set of implementable real balances is smaller, i.e., for all $i \geq 0$, $\mathcal{A}^u(i) \subset \mathcal{A}^o(i)$. Intuitively, when money holdings are observable, there is more leverage to punish an agent who does not carry enough real balances. For my purpose, this implies that if the model can fit money demand when money holdings are unobservable, this would also be the case when money holdings are observable.

5 The irreducible cost of inflation

In the previous section I derived an individual money demand correspondence, $D^p(i)$. Following the methodology of Lucas (2000) and LW, the next step is to construct the aggregate money demand and to check whether there are parameter values for which it fits the data. The model can then be used to measure the cost of inflation.

The aggregate demand for money is defined as $L \equiv M/PY$, where M is the money supply, Y is real aggregate output, and P is the price level. In the data, Y is measured by GDP, P by the GDP deflator, M by M1, and i by the short-term commercial paper rate. Real aggregate output is composed of the CM output, A such that $U'(A) = 1$, and the DM output expressed in terms of the numéraire good, $\sigma M/p$. Hence, $Y = A + \sigma M/p$. Aggregate real balances are $M/p = d^p$. Therefore, the aggregate demand for money is a continuous correspondence defined as

$$L(i) = \left\{ \frac{d^p}{A + \sigma d^p} : d^p \in D^p(i) \right\}.$$

From (12)-(13),

$$L(i) = \left[\frac{\psi(q^*)}{A + \sigma\psi(q^*)}, \frac{\sigma u(q^*)}{(i+\sigma) A + \sigma^2 u(q^*)} \right] \text{ if } i \leq \bar{\imath} \equiv \frac{\sigma \left[u(q^*) - \psi(q^*) \right]}{\psi(q^*)} \tag{19}$$

$$= \frac{\psi\left[q(i)\right]}{A + \sigma\psi\left[q(i)\right]} \text{ if } i > \bar{\imath} \equiv \frac{\sigma \left[u(q^*) - \psi(q^*) \right]}{\psi(q^*)}. \tag{20}$$

I adopt the same functional forms as in LW: $U(c) = A \ln c$, $\psi(e) = e$, $u(q) = \frac{(q+b)^{1-a} - b^{1-a}}{1-a}$. I set $\beta^{-1} = 1.03$, as in Lucas (2000). This gives four parameters, (A, a, b, σ), to adjust to attempt to

fit the money demand in the model to the data. Following the literature, b is chosen to be close to 0 so that the utility function approximates a CRRA.

First, I represent the money demand correspondence under an optimal mechanism for the parameter values obtained in LW with symmetric Nash bargaining as the trading protocol. As noticed in Hu, Kennan, and Wallace (2009, p.136), for this parametrization the first-best allocation is implementable for all the interest rates observed in the data. However, as revealed by Figure 6, the money demand from the model is a poor fit for the data: all the observations except three lie outside of the money demand correspondence. Intuitively, the optimal mechanism gives agents incentives to accumulate enough real balances to trade the first-best level of output, whereas under Nash bargaining the first-best level of output is never achieved, even when the cost of holding money is zero. As a consequence, if the model is calibrated to fit the data under Nash bargaining, the money demand under the optimal mechanism will tend to overestimate the money demand in the data.

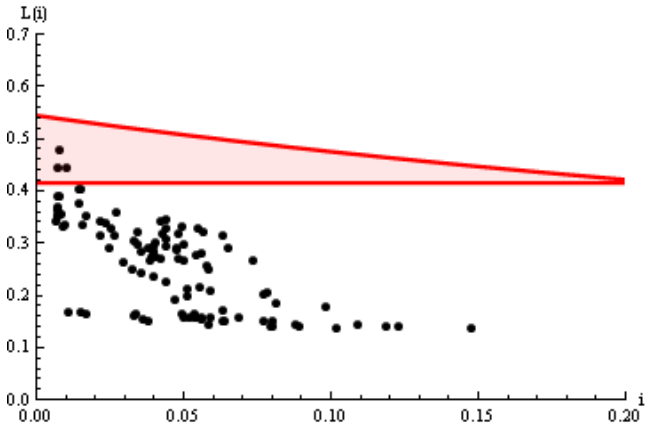


Figure 6: The LW specification: $(A, a, \sigma) = (1.91, 0.3, 0.5)$.

The next step is to recalibrate the model. Figures 7 and 8 show that the model is able to generate a money demand that is consistent with the observations in the data. In fact, all the observations over the period 1900-2006 are in the money demand correspondence generated by the model.[11] In Figure 7, I adopt a utility function in the DM similar to the one in the CM with a unit CRRA ($a = 1.001$), and I set the frequency of trade to its maximum value ($\sigma = 0.5$). I chose the

[11] A noteworthy implication of the fact that the model is consistent with all the observations in the data is that the downward shift in aggregate real balances observed at the end of the 70's does not need to be explained by a

value of CM output, A, to adjust the level of money demand. In Figure 8 I set $b = 0$ but choose (σ, A, a) to generate a shape for the money demand correspondence that is visually close to the data. Again, both parametrizations do equally well to account for the data.

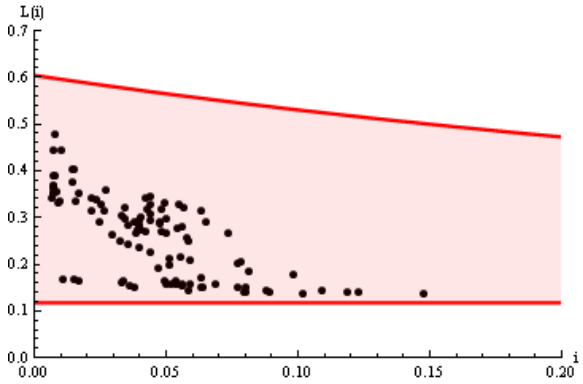


Figure 7: $\sigma = 0.5$, $A = 8$, $a = 1.001$, $b = 0.001$.

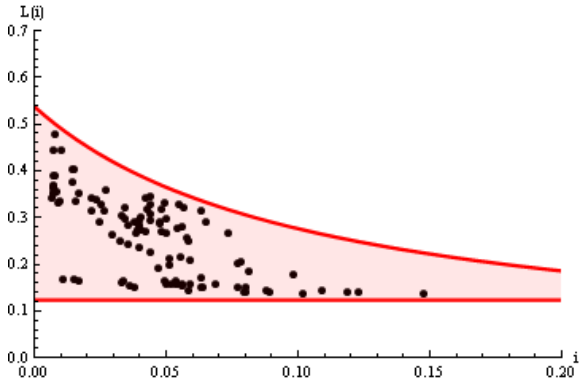


Figure 8: $\sigma = 0.1$, $A = 8$, $a = 0.78$, $b = 0$

To understand how the model can match the empirical money demand, consider the following two moments: $\underline{L} = \min(L_t)$ is the minimum real balance and $\bar{L} = \min(L_t)$ is the maximum real balance observed in the data. One can choose parameters to match these two moments. To do this, it is useful to rewrite (19) as

$$L(i) = \left[\frac{1}{\frac{A}{\psi(q^*)} + \sigma}, \frac{1}{\left(1 + \frac{i}{\sigma}\right) \frac{A}{u(q^*)} + \sigma} \right].$$

change in the fundamental structure of the economy. It could reflect the fact that under an optimal mechanism real balances need not be uniquely determined.

To match these two targets one needs $\frac{1}{\frac{A}{\psi(q^*)}+\sigma} = \underline{L}$ and $\frac{1}{\left(1+\frac{i_{\max}}{\sigma}\right)\frac{A}{u(q^*)}+\sigma} = \bar{L}$, where $i_{\max}$ is the highest interest rate in the data. For a given σ, one can choose $\frac{A}{\psi(q^*)}$ sufficiently high to obtain the lower bound for real balances. This is a condition on the relative size of the CM and DM production levels. For a given σ and $\frac{A}{\psi(q^*)}$, one can make $\frac{A}{u(q^*)} = \frac{A}{\psi(q^*)}\frac{\psi(q^*)}{u(q^*)}$ sufficiently low, or $\frac{\psi(q^*)}{u(q^*)}$ sufficiently high, in order to achieve the upper bound for real balances. This can be interpreted as a condition on the size of the gains from trade in the DM.

The welfare cost of a $\gamma - 1$ percent inflation is defined as the fraction of total consumption that agents would be willing to give up to reduce γ to 1. For the two parameterizations above, the first-best level of output is achieved for all inflation rates observed in the data. Therefore, the welfare cost of 10 percent inflation is 0. In LW, under Nash bargaining, the cost of inflation is 3.2 percent of GDP every year. In the case where buyers play an ultimatum game, the cost of inflation is lowered to 1.2 percent of GDP, in the same ballpark as the estimate of Lucas (2000). Using the same calibration methodology but applying a mechanism design approach, I just showed that the part of the welfare cost of inflation that can be attributed to monetary frictions alone, the *irreducible* cost of inflation, is zero.

The cost of inflation that is measured in LW and the follow-up literature (see the survey in Craig and Rocheteau, 2008) is essentially a welfare loss that can be attributed to suboptimal trading mechanisms. Monetary frictions create a large welfare loss to the extent that they make mechanisms that are optimal in pure credit economies—economies where bilateral credit can be enforced—suboptimal when credit is no longer available. This is not to say that these trading mechanisms are not empirically plausible—the data does not seem able to discriminate between different mechanisms. But the trading mechanisms that have been imposed in the literature are not part of the frictions that make money essential, and when measuring the cost of inflation, one should disentangle the cost associated with those pure monetary frictions from the costs that stem from socially inefficient trading protocols.

6 Match-specific heterogeneity

So far I have assumed that all buyers have the same preferences and all sellers have the same technology. If there is heterogeneity across matches, and if agents have some private information regarding their preferences or technologies, it is unclear whether the first-best allocation, which

involves different production levels in different matches, is incentive feasible for some inflation rates. For instance, in an environment in which sellers offer price schedules to buyers in bilateral matches, Ennis (2008) finds that the equilibrium is inefficient (even at the Friedman rule) and the cost of inflation is large.

In this section I extend the model by assuming that once a buyer and a seller are matched, a preference shock is realized that determines how much the buyer values the output produced by the seller. Preferences are represented by the utility function $\varepsilon u(q) - \psi(e) + U(c) - h$, where $\varepsilon \in \mathcal{E} \subset [0, \bar{\varepsilon}]$ is a match-specific component drawn from a distribution $G(\varepsilon)$.[12] The preference shock, ε, is private information to the buyer.

An allocation rule in the DM maps a triple, (ε, z^b, z^s), the announced match-specific component and the announced buyer's and seller's real balances, into a match allocation, $(q_\varepsilon, d_\varepsilon)$, which specifies the output in a match and the transfer of real balances. The mechanism must be incentive compatible: the buyer is willing to reveal truthfully his preference shock and real balances, while the seller is willing to reveal truthfully his real balances. The mechanism must also be individually rational, i.e., buyers and sellers are willing to go along with the proposed allocation. Because there is no clear notion of coalition-proof equilibrium in the presence of ex-post heterogeneity and private information, I adopt the weaker notion of individually rational (IR) implementability that requires the trades in pairwise meetings to be immune to individual defection.

Suppose the planner seeks to implement symmetric, stationary allocations $\{(q^p_\varepsilon, d^p_\varepsilon) : \varepsilon \in \mathcal{E}, z^p\}$. Buyers will be required to hold $z^p = \max_\varepsilon(d^p_\varepsilon)$ real balances (since with unobservable money holdings an agent will have no incentive to hold more money than what he spends in any match). A necessary condition for a buyer to have incentives to reveal truthfully his preference shock is

$$\varepsilon u(q^p_\varepsilon) - d^p_\varepsilon \geq \varepsilon u(q^p_{\varepsilon'}) - d^p_{\varepsilon'} \quad \text{for all } \varepsilon' \neq \varepsilon. \tag{21}$$

According to (21), the buyer will achieve a higher surplus by reporting his true preference shock, ε, instead of some other value, ε'. From (21)

$$\varepsilon \left[u(q^p_{\varepsilon'}) - u(q^p_\varepsilon)\right] \leq d^p_{\varepsilon'} - d^p_\varepsilon \leq \varepsilon' \left[u(q^p_{\varepsilon'}) - u(q^p_\varepsilon)\right].$$

If $\varepsilon' > \varepsilon$, then $q^p_{\varepsilon'} \geq q^p_\varepsilon$ and $d^p_{\varepsilon'} \geq d^p_\varepsilon$. Buyers with a high marginal utility of consumption receive (weakly) more output and spend (weakly) more real balances than buyers with a low marginal

[12] This formalization is borrowed from Lagos and Rocheteau (2005).

utility of consumption. This implies that an agent's real balances must be $z = d^p_{\bar{\varepsilon}}$. Individual rationality in a match requires

$$\psi(q^p_\varepsilon) \leq d^p_\varepsilon \leq \varepsilon u(q^p_\varepsilon) \quad \text{for all } \varepsilon \in [0, \bar{\varepsilon}]\,. \tag{22}$$

Both the buyer and the seller enjoy a positive surplus. Finally, necessary conditions for an agent in the CM to accumulate z units of real balances are:

$$-iz^p + \sigma \int_0^{\bar{\varepsilon}} \left[\varepsilon u(q^p_\varepsilon) - d^p_\varepsilon\right] dG(\varepsilon) \geq -iz' + \sigma \int_0^{\bar{\varepsilon}} \max_{\varepsilon'} \left[\varepsilon u(q^p_{\varepsilon'}) - d^p_{\varepsilon'}\right] \mathbb{I}_{\{d^p_{\varepsilon'} \leq z'\}} dG(\varepsilon), \tag{23}$$

for all $z' \in \{d^p_\varepsilon : \varepsilon \in \mathcal{E} \cup \{0\}\}$. The deviation that consists of reducing one's real balances from z to $z' < z$ and choosing the best offer such that $d^p_{\varepsilon'} < z'$ must not be profitable. Note that when evaluating this deviation I took into account only the buyer's expected surplus in the DM, net of the cost of holding real balances. Indeed, from the previous section, any incentive-compatible mechanism is such that the seller's expected surplus in the DM is independent of his real balances.

Lemma 3 *Any allocation* $\{(q^p_\varepsilon, d^p_\varepsilon) : \varepsilon \in \mathcal{E}, z^p\}$ *that satisfies* $z^p = d^p_{\bar{\varepsilon}}$*, (21), (22), and (23) is implemented by the following mechanism:*

$$\left[q(z^b, z^s, \varepsilon), d(z^b, z^s, \varepsilon)\right] = (q^p_\varepsilon, d^p_\varepsilon) \quad \textit{if} \quad z^b \geq d^p_{\bar{\varepsilon}}, \tag{24}$$

$$= (0, 0) \quad \textit{otherwise.} \tag{25}$$

The proof of Lemma 3 consists of showing that accumulating $z = d^p_{\bar{\varepsilon}}$ real balances is an optimal strategy. From (24), by choosing $z > d^p_{\bar{\varepsilon}}$ an agent does not increase his surplus in the DM but he incurs a higher cost of holding real balances, $iz > id^p_{\bar{\varepsilon}}$. If the agent chooses $z < d^p_{\bar{\varepsilon}}$, he can lie and announce $d^p_{\bar{\varepsilon}}$, but he will be restricted to offers such that $d_{\varepsilon'} \leq z < d^p_{\bar{\varepsilon}}$, which from (23) lowers his expected utility. Alternatively, the buyer can choose to reveal truthfully his real balances but then, from (25), the mechanism proposes no trade in the DM.

6.1 Discrete preference shocks

Following Curtis and Wright (2004) and Ennis (2008), I consider a specification in which the preference shock can take two values, $\varepsilon \in \{\varepsilon_\ell, \varepsilon_h\}$ with $\varepsilon_h > \varepsilon_\ell$. The probability of the high preference shock is π_h, while the probability of the low preference shock is $\pi_\ell = 1 - \pi_h$. In this

case, the individual-rationality constraints in the CM, (23), can be expressed as

$$-iz^p + \sigma \sum_{\chi\in\{\ell,h\}} \pi_\chi \left[\varepsilon_\chi u(q^p_{\varepsilon_\chi}) - d^p_{\varepsilon_\chi}\right] \geq 0 \tag{26}$$

$$-iz^p + \sigma \sum_{\chi\in\{\ell,h\}} \pi_\chi \left[\varepsilon_\chi u(q^p_{\varepsilon_\chi}) - d^p_{\varepsilon_\chi}\right] \geq -id^p_{\varepsilon_\ell} + \sigma \sum_{\chi\in\{\ell,h\}} \pi_\chi \left[\varepsilon_\chi u(q^p_{\varepsilon_\ell}) - d^p_{\varepsilon_\ell}\right]. \tag{27}$$

The condition (26) specifies that an agent prefers to accumulate z^p real balances instead of 0, i.e., the expected surplus of a buyer net of the cost of holding real balances is non-negative. According to (27), an agent does not decrease his surplus by holding z^p real balances instead of $d^p_{\varepsilon_\ell}$. If an agent chooses to hold $d^p_{\varepsilon_\ell}$ real balances, he incurs a lower cost of holding real balances, but in the event that he has a high marginal utility of consumption, he has to report a low marginal utility in order to be able to trade. The conditions (26) and (27) can be rearranged as

$$z^p \leq \frac{\sigma\pi_\ell \left[\varepsilon_\ell u(q^p_{\varepsilon_\ell}) - d^p_{\varepsilon_\ell}\right] + \sigma\pi_h\varepsilon_h u(q^p_{\varepsilon_h})}{i + \sigma\pi_h} \tag{28}$$

$$z^p \leq \frac{\sigma\pi_h\varepsilon_h \left[u(q^p_{\varepsilon_h}) - u(q^p_{\varepsilon_\ell})\right]}{i + \sigma\pi_h} + d^p_{\varepsilon_\ell}. \tag{29}$$

In the following I establish the conditions under which the first-best allocation is implementable. The first best requires $q_{\varepsilon_h} = q^*_h$ and $q_{\varepsilon_\ell} = q^*_\ell$, where $\varepsilon_h u'(q^*_h) = \psi'(q^*_h)$ and $\varepsilon_\ell u'(q^*_\ell) = \psi'(q^*_\ell)$.

Proposition 2 *Consider an economy with preference shocks in $\{\varepsilon_\ell, \varepsilon_h\}$ that are private information to buyers. There exists $\bar{\imath} \in (0, \infty)$ such that the first-best allocation is implementable for all $i \in [0, \bar{\imath}]$.*

There is a range of inflation rates, including the Friedman rule, that implement the first-best allocation. So, as in the case with homogenous matches, the first-best allocation can be obtained even with inflation rates above the Friedman rule.

To conclude this section, I provide a calibrated example with match-specific heterogeneity where the money demand correspondence fits the data. Real output in the DM expressed in terms of the numéraire good is $\sigma(\pi_h z + \pi_\ell d_{\varepsilon_\ell})$ since in h-type matches the buyer spends all his real balances and in ℓ-type matches the buyer spends $d_{\varepsilon_\ell} < z$. Hence, aggregate output is $Y = A + \sigma(\pi_h z + \pi_\ell d_\ell)$. The money demand correspondence is

$$L(i) = \left\{ \frac{z}{A + \sigma\pi_h z + \sigma\pi_\ell d_{\varepsilon_\ell}} \ : (z, d_{\varepsilon_\ell}) \in \mathcal{A}^*(i) \right\},$$

where $\mathcal{A}^*(i)$ is the set of pairs $(z, d_\ell) \in \mathbb{R}_{2+}$ such that (21), (22), and (23) hold with $q_{\varepsilon_h} = q_h^*$ and $q_{\varepsilon_\ell} = q_\ell^*$. I normalize $\varepsilon_h = 1$. To make the match-specific heterogeneity relevant I take $\varepsilon_\ell = 0.5$ and $\pi_h = 0.5$. The frequency of meetings is assumed to be maximum, $\sigma = 0.5$, and I set $a = 0.9$. I choose the parameter A so that all points in the data are in the money demand correspondence. See Figure 9. For this example, the welfare cost of (moderate) inflation is 0.

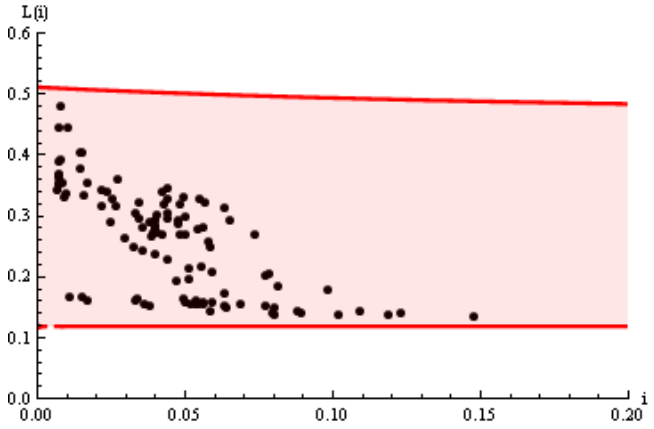


Figure 9: $\varepsilon_\ell = 0.5$, $\varepsilon_h = 1$, $\pi_h = 0.5$, $\sigma = 0.5$, $a = 0.9$, $A = 8$

6.2 Continuous preference shocks

Suppose that the preference shock is distributed uniformly on the interval $[0, 1]$. From (21)

$$\varepsilon \left[\frac{u(q_{\varepsilon'}^p) - u(q_\varepsilon^p)}{\varepsilon' - \varepsilon} \right] \leq \frac{d_{\varepsilon'}^p - d_\varepsilon^p}{\varepsilon' - \varepsilon} \leq \varepsilon' \left[\frac{u(q_{\varepsilon'}^p) - u(q_\varepsilon^p)}{\varepsilon' - \varepsilon} \right].$$

Taking the limit as ε' approaches ε,

$$\frac{\partial d_\varepsilon^p}{\partial \varepsilon} = \varepsilon u'(q_\varepsilon^p) \frac{\partial q_\varepsilon^p}{\partial \varepsilon}.$$

If the first-best level of output is implemented, $\varepsilon u'(q_\varepsilon^p) = \psi'(q_\varepsilon^p)$, which implies $\frac{\partial d_\varepsilon^p}{\partial \varepsilon} = \psi'(q_\varepsilon^*) \frac{\partial q_\varepsilon^*}{\partial \varepsilon}$. From (22), $d_0^p = 0$. Hence,

$$d_\varepsilon^p = \psi\left(q_\varepsilon^*\right).$$

In order to implement the first-best, the transfer of real balances must be equal to the disutility cost of the seller. So the participation constraints (22) are satisfied.

I now turn to the buyer's incentives to accumulate real balances in the CM. Consider a buyer with $d_{\tilde{\varepsilon}}^p$ real balances in a match of type ε. This buyer can buy at most $q_{\tilde{\varepsilon}}^*$ units of output. He will

announce a preference type, $\hat{\varepsilon}$, that maximizes $[\varepsilon u(q^*_{\hat{\varepsilon}}) - \psi(q^*_{\hat{\varepsilon}})]\,\mathbb{I}_{\{q^*_{\hat{\varepsilon}} \leq q^*_{\tilde{\varepsilon}}\}}$. The solution is $\hat{\varepsilon} = \varepsilon$ if $\tilde{\varepsilon} \geq \varepsilon$ and otherwise $\hat{\varepsilon} = \tilde{\varepsilon}$. Provided that the buyer has enough real balances to purchase the first-best level of output given his preference type, he will announce his type truthfully. If he doesn't hold enough real balances, he will announce the highest type consistent with his real balances, $\tilde{\varepsilon}$. Therefore, the buyer's choice of real balances in the CM is equivalent to the choice of a threshold, $\tilde{\varepsilon}$, below which the buyer consumes the first-best level of output and above which the buyer is constrained by his real balances. It solves

$$\tilde{\varepsilon} = \arg\max_{\tilde{\varepsilon}} \left\{ -i\psi(q^*_{\tilde{\varepsilon}}) + \sigma \int_0^{\tilde{\varepsilon}} [\varepsilon u(q^*_\varepsilon) - q^*_\varepsilon]\, d\varepsilon + \sigma \int_{\tilde{\varepsilon}}^1 [\varepsilon u(q^*_{\tilde{\varepsilon}}) - \psi(q^*_{\tilde{\varepsilon}})]\, d\varepsilon \right\}.$$

The first-order condition is

$$i = \sigma \int_{\tilde{\varepsilon}}^1 \left[\frac{\varepsilon u'(q^p_{\tilde{\varepsilon}})}{\psi(q^*_{\tilde{\varepsilon}})} - 1 \right] d\varepsilon.$$

Interestingly, this first-order condition is the same as the one obtained in Lagos and Rocheteau (2005) under buyers-take-all bargaining. Unless $i = 0$, the first-best is not implementable.[13] This also tells us that the cost of inflation will be bounded above by the cost obtained under buyers-take-all bargaining, which is about 1.2 percent of GDP.

So far I assumed that money holdings were not observable and could be overstated. If money holdings cannot be overstated, then agents will be willing to accumulate the socially efficient quantity of real balances, $d^p_1 = \psi(q^*_1)$, if

$$-i\psi(q^*_1) + \sigma \int_0^1 [\varepsilon u(q^*_\varepsilon) - q^*_\varepsilon]\, d\varepsilon \geq 0.$$

There is an interval of inflation rates above the Friedman rule that are consistent with the implementation of the first-best.

7 Endogenous participation

As shown in Section 5 under an optimal mechanism, aggregate money demand is a correspondence: for low inflation rates the division of the surplus in a bilateral match is not uniquely pinned down. A natural way to endogenize the division of the match surplus – thereby refining money demand – is to let agents choose which side of the DM market they participate in. Indeed, the division of the gains from trade in the DM affects the composition of the market in terms of buyers and sellers,

[13] This result is consistent with Faig and Jerez (2006), who assume that sellers compete to attract buyers.

the measure of trades, and therefore society's welfare. In order to endogenize participation, I follow the approach of Shi (1997) in the context of a large household model and Rocheteau and Wright (2009) in the LW model.

I assume that there is a unit measure of ex ante identical agents who choose to be either buyers or sellers in the DM. The decision to become a buyer or seller in period t is taken at the beginning of the previous CM, in period $t-1$. Suppose, for example, that at the beginning of the CM, individuals invest in a (costless) technology that allows them to either produce DM goods or consume them, and it is only possible to invest in one technology.[14]

Let n denote the measure of buyers in the DM, $\theta = \frac{1-n}{n}$ the ratio of sellers per buyer (market tightness), and $\alpha(\theta)$ the matching probability of a buyer. The matching function has standard properties: $\alpha(0) = 0$, $\alpha' > 0$, $\alpha'(0) \leq 1$, $\alpha'(\infty) = 0$, $\alpha'' < 0$. The matching probability of a seller is $\alpha(\theta)/\theta$. Society's welfare is measured by

$$\mathcal{W} = n\alpha\left(\frac{1-n}{n}\right)[u(q) - \psi(q)] + U(c) - c. \tag{30}$$

Let denote n^* the composition of the market that maximizes the number of trades,

$$n^* = \arg\max n\alpha\left(\frac{1-n}{n}\right). \tag{31}$$

The first-best allocation is such that $q = q^*$, $n = n^*$, and $c = c^*$.

A symmetric, stationary allocation is represented by the 5-tuple $(q^p, d^p, n^p, z_b^p, z_s^p)$, where z_b^p denotes the buyer's real balances and z_s^p the seller's real balances. Let W^b (W^s) denote the value function of an agent in the CM who chooses to be a buyer (seller) in the next DM, and V^b (V^s) denotes the value function of a buyer (seller) in the DM. The value function at the beginning of the CM is similar to (2) and satisfies

$$W^j(z) = T + z + \max_{c \geq 0}\{U(c) - c\} + \max_{\hat{z} \geq 0}\left\{-\gamma\hat{z} + \beta V^j(\hat{z})\right\}, \tag{32}$$

where $j \in \{b, s\}$. The value of being a buyer in the DM satisfies

$$V^b(z) = \alpha(\theta)\left\{u\left[q(z, z_s^p)\right] - d\left(z, z_s^p\right)\right\} + \max\left[W^b(z), W^s(z)\right]. \tag{33}$$

[14] One can think of the DM good as being an intermediate good, where sellers produce the intermediate good and buyers produce a final good that requires the intermediate good as an input. The final good is produced after the buyer and seller split apart. Therefore, the final good cannot be consumed by both the buyer and seller.

Substituting (33) into (32), and using the linearity of $W^b(z)$ and $W^s(z)$, the value of a buyer with z units of real balances at the beginning of the CM satisfies

$$W^b(z) = T + z + \max_{c\geq 0}\{U(c) - c\} + \beta \max\left[W^b(0), W^s(0)\right] \tag{34}$$
$$+ \max_{z\geq 0} \beta\left\{-iz + \alpha(\theta)\left\{u\left[q(z, z_s^p)\right] - d\left(z, z_s^p\right)\right\}\right\}.$$

By similar reasoning, the value of being a seller with z units of real balances satisfies

$$W^s(z) = T + z + \max_{c}\{U(c) - c\} + \beta \max\left[W^b(0), W^s(0)\right] \tag{35}$$
$$+ \max_{z\geq 0} \beta\left\{-iz + \frac{\alpha(\theta)}{\theta}\left\{-\psi\left[q(z_b^p, z)\right] + d\left(z_b^p, z\right)\right\}\right\}.$$

For the trading mechanism to be incentive-compatible, the seller's surplus must be independent of his real balances. Consequently, from (35), sellers do not carry real balances in the DM, $z_s^p = 0$.

A buyer will never find it optimal to carry more money than he spends (otherwise he would have an incentive to misreport his money balances), so an incentive-compatible mechanism must be such that $d^p = z_b^p$. Moreover, for all $n^p \in (0,1)$, $W^b(z) = W^s(z)$, which from (34) and (35) implies

$$-iz_b^p + \alpha(\theta)\left[u\left(q^p\right) - d^p\right] = \frac{\alpha(\theta)}{\theta}\left[d^p - \psi\left(q^p\right)\right] \geq 0, \tag{36}$$

where $\theta = \frac{1-n^p}{n^p}$. The left side of (36) is the expected surplus of a buyer, net of the cost of holding real balances. The right side of (36) is the expected surplus of a seller.

Lemma 4 *The allocation, $(q^p, d^p, n^p, z_b^p, z_s^p)$, that satisfies $z_s^p = 0$, (6) and (36) and $(q^p, d^p) \in \mathcal{C}(z_b^p, 0)$ can be implemented by the trading mechanism (7)-(8).*

The trading mechanism (7)-(8) guarantees that a buyer accumulates z_b^p real balances in the CM, reveals truthfully his money holdings, and agrees to the trade (q^p, d^p). Given the proposed terms of trade, agents are indifferent between being buyers or sellers, i.e., condition (36) holds, when the measure of buyers is equal to n^p.

Definition 2 *An optimal mechanism with endogenous participation is a $[q(\cdot,\cdot), d(\cdot,\cdot)]$ that satisfies (7)-(8) with*

$$(q^p, d^p, n^p) \in \max_{q,d,n} \mathcal{W} = n\alpha\left(\frac{1-n}{n}\right)[u(q) - \psi(q)] + U(c^*) - c^* \tag{37}$$

$$s.t. \quad -\psi(q) + d \geq 0 \tag{38}$$

$$-id + \alpha\left(\frac{1-n}{n}\right)\left[u\left(q\right) - d\right] = \frac{n}{1-n}\alpha\left(\frac{1-n}{n}\right)\left[d - \psi\left(q\right)\right]. \tag{39}$$

Proposition 3 *There is* $\bar{\imath} \equiv \frac{\alpha\left(\frac{1-n^*}{n^*}\right)[u(q^*)-\psi(q^*)]}{\psi(q^*)}$ *such that*

1. *For all* $i < \bar{\imath}$, $q^p = q^*$, $n^p = n^*$, *and*

$$d^p = \frac{(1-n^*)\,\alpha\left(\frac{1-n^*}{n^*}\right)u(q^*) + n^*\alpha\left(\frac{1-n^*}{n^*}\right)\psi(q^*)}{\alpha\left(\frac{1-n^*}{n^*}\right) + i\,(1-n^*)}. \tag{40}$$

   Moreover, $\frac{\partial \mathcal{W}}{\partial i} = 0$ *and* $\frac{\partial (M_t/p_t)}{\partial i} < 0$*;*

2. *For all* $i > \bar{\imath}$, $q^p < q^*$, $n^p < n^*$. *Moreover,* $\frac{\partial \mathcal{W}}{\partial i} < 0$ *and* $\frac{\partial (M_t/p_t)}{\partial i} < 0$.

In contrast to Shi (1997) and Rocheteau and Wright (2009), where terms of trade are determined by bargaining, under an optimal mechanism the first-best allocation can be implemented for low inflation rates. If the cost of holding money is lower than the threshold $\bar{\imath}$, then the first-best allocation is incentive feasible. According to (40), the transfer of real balances is a decreasing function of the nominal interest rate. As the nominal interest rate increases, it is more costly to hold money and in order to keep the buyers' and sellers' incentives to participate in the market unchanged, buyers must be compensated by a larger share of the match surplus (which implies that they hold fewer real balances). If the cost of holding money is larger than the threshold $\bar{\imath}$, then the first-best allocation is no longer incentive feasible. The allocation is chosen so that agents are just indifferent between participating and not participating in the market. In this case, output and the measure of buyers in the DM decrease with inflation.

I now turn to the calibration of the model. As above, $Y = A + n^p\alpha\left(\frac{1-n^p}{n^p}\right)d^p$ and $M/P = n^p d^p$. The aggregate demand for money is

$$L(i) = \frac{n^p d^p}{A + n^p\alpha\left(\frac{1-n^p}{n^p}\right)d^p}. \tag{41}$$

I adopt the same functional forms as before, $\psi(e) = e$, $u(q) = \frac{(q+b)^{1-a}-b^{1-a}}{1-a}$, but here I set $b = 0$ and restrict a in $(0,1)$. I adopt the matching function that is commonly used in the literature, $\alpha(\theta) = \frac{\sigma\theta}{1+\theta}$. The matching probability of a buyer, $\sigma(1-n)$, is proportional to the measure of sellers. Symmetrically, the matching probability of a seller, σn, is proportional to the measure of buyers. The measure of DM trades is maximum when $n = 1/2$.

In Figure 10 below I set $a = 0.95$ and I look for the pair (σ, A) that provides the best fit with the data by minimizing squared residuals. I consider the whole sample, 1900-2006, as well as the subsample 1981-2006. For the whole sample a good fit requires a low frequency of trade, $\sigma = 0.06$.

This suggests that shocks where the liquidity constraints bind in the DM have to be infrequent for the model to match the data. For the subsample 1981-2006 money demand is flatter and a good fit is obtained for a much larger frequency of trades, $\sigma = 0.47$. For both examples, a 10 percent inflation imposes no cost on society.

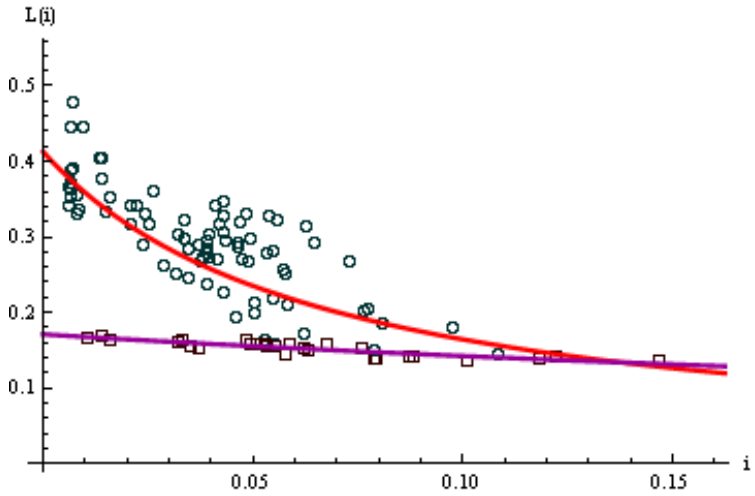


Figure 10: 1900-2006: $\sigma = 0.06$, $A = 12.55$, $a = 0.95$; 1981-2006: $\sigma = 0.47$, $A = 29.43$, $a = 0.95$.

8 Conclusion

I have studied the welfare cost of inflation in an environment in which money plays an essential role. When the trading mechanism in pairwise meetings is chosen optimally, aggregate money demand takes the form of a correspondence that can fit the data over the period 1900-2006. The welfare cost of moderate inflation that can be attributed to monetary frictions alone is zero. Hence, in contrast to some common wisdom, inflation does not need to impose a large burden on society when the only frictions in the environment are the ones that make money useful. This insight is robust to different assumptions regarding the observability of money holdings, the introduction of match-specific heterogeneity, and endogenous participation decisions. It is important to recall that this prediction relies on the trading mechanism being socially efficient. If agents were to trade according to some other mechanism – some mechanisms despite being socially inefficient have strong strategic foundations – the cost of inflation would be large, as argued in the literature. This reinforces Williamson and Wright's (2010) assessment that "getting into the details of monetary

theory can make a big difference for quantitative and policy analysis."

References

[1] Andolfatto, David (2010). "Essential interest-bearing money," Journal of Economic Theory 145, 1319-602.

[2] Aruoba, Boragan, Christopher Waller and Randall Wright (2010). "Money and capital," Journal of Monetary Economics (Forthcoming).

[3] Aruoba, Boragan, and Sanjay Chugh (2010). "Optimal fiscal and monetary policy when money is essential," Journal of Economic Theory 145, 1618-1647.

[4] Bailey, Martin (1956). "The welfare cost of inflationary finance," Journal of Political Economy 64, 93-110.

[5] Berensten, Aleksander and Guillaume Rocheteau (2002). "Money in bilateral trade," Swiss Journal of Economics and Statistics 138, 489-506.

[6] Boel, Paola and Gabriele Camera (2009). "Financial sophistication and the distribution of the welfare cost of inflation,"Journal of Monetary Economics 56, 968-978.

[7] Boel, Paula, and Gabriele Camera (2010). "The welfare cost of inflation in OECD countries," Macroeconomic Dynamics (Forthcoming).

[8] Burstein, Ariel, and Christian Hellwig (2008). "Welfare costs of inflation in a menu cost model," American Economic Review 98, 438–43.

[9] Cavalcanti, Ricardo, and Andrés Erosa (2008). "Efficient propagation of shocks and the optimal return on money," Journal of Economic Theory 142, 128-148

[10] Cavalcanti, Ricardo, and Ed Nosal (2009). "Some benefits of cyclical monetary policy," Economic Theory 39, 195-216.

[11] Cavalcanti, Ricardo, and Neil Wallace (1999). "Inside and outside money as alternative media of exchange," Journal of Money, Credit, and Banking 31, 443–457.

[12] Chiu, Jonathan, and Miguel Molico (2010). "Liquidity, redistribution, and the welfare cost of inflation," Journal of Monetary Economics 57, 428–438.

[13] Craig, Ben and Guillaume Rocheteau (2008). "Inflation and welfare: A search approach," Journal of Money, Credit and Banking 40, 89-120.

[14] Curtis, Elisabeth, and Randall Wright (2004). "Price setting, price dispersion, and the value of money: or, the law of two prices," Journal of Monetary Economics 51, 1599-1621.

[15] da Costa, Carlos and Ivan Werning (2008). "On the optimality of the Friedman Rule with heterogeneous agents and nonlinear income taxation," Journal of Political Economy 116, 82-112.

[16] Deviatov, Alexei (2006). "Money creation in a random matching model," Topics in Macroeconomics 6.

[17] Deviatov, Alexei, and Neil Wallace (2001). "Another example in which money creation is beneficial", Advances in Macroeconomics 1.

[18] Dotsey Michael and Peter Ireland (1996). "The welfare cost of inflation in general equilibrium," Journal of Monetary Economics 37, 29-47.

[19] Ennis, Huberto (2008). "Search, money, and inflation under private information," Journal of Economic Theory 138, 101-131.

[20] Faig, Miquel and Belen Jerez (2006). "Inflation, prices, and information in competitive search," Advances in Macroeconomics 6, Article 3.

[21] Friedman, Milton (1969). "The optimum quantity of money." In The Optimum Quantity of Money and Other Essays. Chicago: the Aldine publishing company.

[22] Hu, Tai-wei, Kennan, John, and Neil Wallace (2009). "Coalition-proof trade and the Friedman rule in the Lagos-Wright model," Journal of Political Economy 117, 116-137.

[23] Ireland, Peter (2009). "On the welfare cost of inflation and the recent behavior of money demand," American Economic Review 99, 1040–1052.

[24] Kocherlakota, Narayana (1998). "Money is memory," Journal of Economic Theory 81, 232–251.

[25] Kocherlakota, Narayana and Neil Wallace (1998). "Incomplete record-keeping and optimal payment arrangements," Journal of Economic Theory 81, 272–289.

[26] Koeppl, Thorsten, Cyril Monnet, and Ted Temzelides (2008). "A dynamic model of settlement," Journal of Economic Theory 142, 233–246.

[27] Lagos, Ricardo, and Guillaume Rocheteau (2005). "Inflation, output and welfare," International Economic Review 46, 495-522.

[28] Lagos, Ricardo and Wright, Randall (2005). "A unified framework for monetary theory and policy analysis," Journal of Political Economy 113, 463-484.

[29] Lucas, Robert (2000). "Inflation and welfare," Econometrica 68, 247-274.

[30] Mattesini, Fabrizio, Cyril Monnet and Randall Wright (2010). "Banking: A mechanism design approach," Working Paper.

[31] Reed, Robert, and Christopher Waller (2006). "Money and risk sharing." Journal of Money, Credit, and Banking 38, 1599-1618.

[32] Rocheteau, Guillaume, and Wright, Randall (2009). "Inflation and welfare in models with trading frictions," in Monetary Policy in Low Inflation Economies, David Altig and Ed Nosal, eds., Cambridge University Press.

[33] Shi, Shouyong (1997). "A divisible search model of fiat money," Econometrica 65, 75-102.

[34] Shimer, Robert (2005). "The cyclical behavior of unemployment and vacancies," American Economic Review 95, 25-49.

[35] Wallace, Neil (2001). "Whither monetary economics?" International Economic review 42, 847-869.

[36] Wallace, Neil (2010). "The mechanism-design approach to monetary theory," Handbook of Monetary Economics (Forthcoming).

[37] Williamson, Stephen, and Randall Wright (2010). "New monetarist economics: Methods," Federal Reserve Bank of St. Louis Review 92, 265-302.

[38] Zhu, Tao (2008). "Equilibrium concepts in the large-household model," Theoretical Economics 3, 257-281.

APPENDIX

Proof of Proposition 1.

Part 1. From (12), the threshold $\bar{\imath}$ for i below which the first-best level of output is implementable, $q^p(i) = q^*$, is the largest value of i such that $\psi(q^*) \leq \frac{\sigma}{i+\sigma} u(q^*)$. Hence, $\psi(q^*) = \frac{\sigma}{\bar{\imath}+\sigma} u(q^*)$, which gives (14). From (9) $\mathcal{W} = \sigma\left[u(q^*) - \psi(q^*)\right] + U(c^*) - c^*$ is independent of i. Hence, $\frac{\partial \mathcal{W}}{\partial i} = 0$ for all $i < \bar{\imath}$. Finally, it is straightforward that $\left[\psi(q^*), \frac{\sigma}{i+\sigma} u(q^*)\right] \subset \left[\psi(q^*), \frac{\sigma}{i'+\sigma} u(q^*)\right]$ for all $i' < i$.

Part 2. From (13), for all $i > \bar{\imath}$, $q(i)$ is the positive solution to $\psi(q) = \frac{\sigma}{i+\sigma} u(q)$. Hence, $u'(q^p) < \frac{i+\sigma}{\sigma} \psi'(q^p)$ (since $\frac{\sigma}{i+\sigma} u(q)$ intersects $\psi(q)$ by above) and

$$\frac{\partial q^p}{\partial i} = \frac{\psi(q^p)}{\sigma u'(q^p) - (i+\sigma)\,\psi'(q^p)} < 0.$$

Since $d^p = \psi(q^p)$ with $\psi' > 0$, $\frac{\partial d^p}{\partial i} < 0$. From (9)

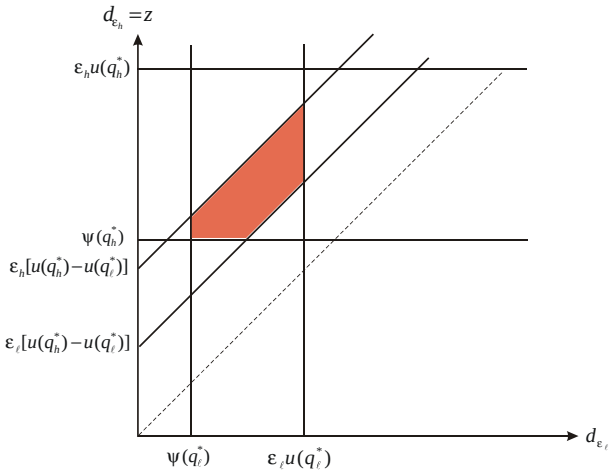
$$\frac{\partial \mathcal{W}}{\partial i} = \sigma\left[u'(q^p) - \psi'(q^p)\right] \frac{\partial q^p}{\partial i} < 0,$$

where I used that $q^p < q^*$ and hence $u'(q^p) - \psi'(q^p) > 0$. ■

Proof of Proposition 2. Denote $\mathcal{A}^*(i)$ the set of pairs $(z, d_\ell) \in \mathbb{R}_{2+}$ such that (21), (22), and (23) hold with $q_{\varepsilon_h} = q_h^*$ and $q_{\varepsilon_\ell} = q_\ell^*$. If $i = 0$, the constraints (21) and (22), which imply (23), can be reexpressed as:

$$\begin{aligned} \varepsilon_\ell\left[u(q_h) - u(q_\ell)\right] &\leq d_h - d_\ell \leq \varepsilon_h\left[u(q_h) - u(q_\ell)\right] \\ \psi(q_\varepsilon) &\leq d_\varepsilon \leq \varepsilon u(q_\varepsilon) \quad \text{for all } \varepsilon \in \{\varepsilon_\ell, \varepsilon_h\}. \end{aligned}$$

The set $\mathcal{A}^*(0)$ is illustrated in the figure below.



The measure of the set $\mathcal{A}^*(0)$ is

$$\mu\left[\mathcal{A}^*(0)\right] = \int_{\psi(q_\ell^*)}^{\hat{d}} x + \varepsilon_h\left[u(q_h^*) - u(q_\ell^*)\right] - \psi(q_h^*)dx + \int_{\hat{d}}^{\varepsilon_\ell u(q_\ell^*)} \left(\varepsilon_h - \varepsilon_\ell\right)\left[u(q_h^*) - u(q_\ell^*)\right]dx,$$

where $\hat{d} = \min\left\{\psi(q_h^*) - \varepsilon_\ell\left[u(q_h^*) - u(q_\ell^*)\right], \varepsilon_\ell u(q_\ell^*)\right\} > \psi(q_\ell^*)$. Moreover, $x + \varepsilon_h\left[u(q_h^*) - u(q_\ell^*)\right] - \psi(q_h^*) > 0$ for all $x > \psi(q_\ell^*)$. So $\mu\left[\mathcal{A}^*(0)\right] > 0$.

Define $\bar{z}(i,x)$ as the upper bound on real balances consistent with (21). (28) and (29), i.e.,

$$\min\left\{\frac{\sigma\pi_\ell\left[\varepsilon_\ell u(q_\ell^*) - x\right] + \sigma\pi_h\varepsilon_h u(q_h^*)}{i + \sigma\pi_h}, \frac{\sigma\pi_h\varepsilon_h\left[u(q_h^*) - u(q_\ell^*)\right]}{i + \sigma\pi_h} + x, x + \varepsilon_h\left[u(q_h^*) - u(q_\ell^*)\right]\right\}.$$

Then,

$$\mu\left[\mathcal{A}^*(i)\right] = \int_{\psi(q_\ell^*)}^{\hat{d}} \bar{z}(i,x) - \psi(q_h^*)dx + \int_{\hat{d}}^{\varepsilon_\ell u(q_\ell^*)} \bar{z}(i,x) - \varepsilon_\ell\left[u(q_h^*) - u(q_\ell^*)\right]dx.$$

It follows that $\mu\left[\mathcal{A}^*(i)\right]$ is nonincreasing and continuous with i. Finally, from (21) and (29),

$$\varepsilon_\ell\left[u(q_h^*) - u(q_\ell^*)\right] \leq d^p_{\varepsilon_h} - d^p_{\varepsilon_\ell} \leq \frac{\sigma\pi_h\varepsilon_h}{i + \sigma\pi_h}\left[u(q_h^*) - u(q_\ell^*)\right].$$

A necessary condition for $\mathcal{A}^*(i) \neq \varnothing$ is

$$\frac{\sigma\pi_h}{i + \sigma\pi_h} \geq \frac{\varepsilon_\ell}{\varepsilon_h}.$$

Consequently, there is an $i < \infty$ such that $\mathcal{A}^*(i)$ is empty and $\mu\left[\mathcal{A}^*(i)\right] = 0$. By the continuity of $\mu\left[\mathcal{A}^*(i)\right]$, there is a threshold, $\bar{\imath} > 0$, such that for all $i < \bar{\imath}$, $\mu\left[\mathcal{A}^*(i)\right] > 0$ and $\mathcal{A}^*(i) \neq \varnothing$. ■

Proof of Proposition 3. The solution to (37)-(39) is

$$\begin{aligned} q^p &= q^* \\ n^p &= n^* \\ d^p &= \frac{\left(1 - n^*\right)\alpha\left(\frac{1-n^*}{n^*}\right)u\left(q^*\right) + n^*\alpha\left(\frac{1-n^*}{n^*}\right)\psi\left(q^*\right)}{\alpha\left(\frac{1-n^*}{n^*}\right) + i\left(1 - n^*\right)}, \end{aligned} \tag{42}$$

if $i \leq \bar{\imath} \equiv \frac{\alpha\left(\frac{1-n^*}{n^*}\right)\left[u(q^*) - \psi(q^*)\right]}{\psi(q^*)}$, where $\bar{\imath}$ is obtained from (38) at equality with q^p and d^p as defined above. Otherwise, (38) holds at equality and

$$\left(q^p, n^p\right) \in \arg\max_{q,n} n\alpha\left(\frac{1-n}{n}\right)\left[u(q) - \psi(q)\right] \tag{43}$$

$$\text{s.t.} \quad -i\psi\left(q\right) + \alpha\left(\frac{1-n}{n}\right)\left[u\left(q\right) - \psi\left(q\right)\right] = 0. \tag{44}$$

The case where $i < \bar{\imath}$ follows from (40). Consider next the case $i > \bar{\imath}$. From (44)

$$n = \frac{1}{1 + \alpha^{-1}\left[\frac{i\psi(q)}{u(q)-\psi(q)}\right]}. \tag{45}$$

Substituting into (43), the mechanism design problem becomes

$$\mathcal{W}(i) = \max_{q} \frac{i\psi\left(q\right)}{1 + \alpha^{-1}\left[\frac{i\psi(q)}{u(q)-\psi(q)}\right]} + U(c^*) - c^*.$$

The first-order condition for the optimal choice of q can be rearranged to read as

$$-\left[1 - \frac{\alpha'\left(\frac{1-n^p}{n^p}\right)}{n^p\alpha\left(\frac{1-n^p}{n^p}\right)}\right] i + \alpha\left(\frac{1-n^p}{n^p}\right)\left[\frac{u'(q^p)}{\psi'(q^p)} - 1\right] = 0, \tag{46}$$

where I used $\frac{\psi(q^p)}{u(q^p)-\psi(q^p)} = \frac{\alpha\left(\frac{1-n^p}{n^p}\right)}{i}$. From the Envelope Theorem,

$$\frac{\partial \mathcal{W}}{\partial i} = \psi\left(q^p\right) n^p \left[1 - \frac{n^p\alpha\left(\frac{1-n^p}{n^p}\right)}{\alpha'\left(\frac{1-n^p}{n^p}\right)}\right].$$

Using (46),

$$\frac{\partial \mathcal{W}}{\partial i} = -\frac{\psi\left(q^p\right)\left[n^p\alpha\left(\frac{1-n^p}{n^p}\right)\right]^2}{i\alpha'\left(\frac{1-n^p}{n^p}\right)}\left[\frac{u'(q^p)}{\psi'(q^p)} - 1\right].$$

Suppose $q^p > q^*$. One can reduce q^p to q^* and increase $u(q^p) - \psi(q^p)$, and d^p can be chosen such that the buyers' and sellers' incentives to participate are unaffected. To see this denote d_0 and d_1 as the values for the transfers of real balances such that

$$-id_1 + \alpha\left(\frac{1-n^p}{n^p}\right)\left[u\left(q^*\right) - d_1\right] = 0 \tag{47}$$

$$\frac{n^p}{1-n^p}\alpha\left(\frac{1-n^p}{n^p}\right)\left[d_0 - \psi\left(q^*\right)\right] = 0. \tag{48}$$

From $q^* < q^p$, $\psi\left(q^*\right) < \psi\left(q^p\right)$, $u\left(q^*\right) - \psi\left(q^*\right) > u\left(q^p\right) - \psi\left(q^p\right)$, and

$$-i\psi\left(q^*\right) + \alpha\left(\frac{1-n^p}{n^p}\right)\left[u\left(q^*\right) - \psi\left(q^*\right)\right] > -i\psi\left(q^p\right) + \alpha\left(\frac{1-n^p}{n^p}\right)\left[u\left(q^p\right) - \psi\left(q^p\right)\right] = 0.$$

Consequently, $d_1 > d_0 = \psi\left(q^*\right)$. Therefore,

$$-id_1 + \alpha\left(\frac{1-n^p}{n^p}\right)\left[u\left(q^*\right) - d_1\right] = 0 < -id_0 + \alpha\left(\frac{1-n^p}{n^p}\right)\left[u\left(q^*\right) - d_0\right],$$

and

$$d_0 - \psi\left(q^*\right) = 0 < d_1 - \psi\left(q^*\right).$$

By continuity, there is a $d \in [d_0, d_1]$ such that (39) holds with $q = q^*$ and $n = n^p$. From (37), such a deviation raises $\mathcal{W}$. Moreover, from (46), $q^p \neq q^*$ since otherwise $n^p = n^*$ and $i \leq \bar{\imath}$. Consequently, $q^p < q^*$ and $\frac{\partial \mathcal{W}}{\partial i} < 0$. To show that $n^p < n^*$, notice first from (46) that $n^p \neq n^*$. From (45) n is a decreasing function of q. So if $n^p > n^*$, one can reduce n and increase q, which would raise welfare. From (38) at equality and (44), $\mathcal{W}(i) = in^p d^p + U(c^*) - c^*$, with $n^p d^p = \frac{M_t}{p_t}$. The result $\frac{\partial \mathcal{W}}{\partial i} < 0$ implies $\frac{\partial (M_t/p_t)}{\partial i} < 0$. ■