Inflation and Welfare: A Search Approach*

Ben Craig
Federal Reserve Bank of Cleveland
Ben.Craig@clev.frb.org

Guillaume Rocheteau
Federal Reserve Bank of Cleveland
Guillaume.Rocheteau@clev.frb.org

This version: 5/24/2006

Abstract

This paper uses a search model of monetary exchange to provide new insights for evaluating the welfare costs of inflation. We first show that the search model of money can rationalize the estimates of the welfare cost of inflation based on the "welfare triangle" methodology of Bailey (1956) and Lucas (2000). The welfare cost of inflation predicted by the model approximates the area underneath money demand when pricing mechanisms are such that buyers appropriate the social marginal benefit of their real balances. For other mechanisms, the measure given by the welfare triangle has to be scaled up by a factor that increases with sellers' market power. We also show that the choice of the bargaining solution matters for the ability of the Friedman rule to generate the first-best allocation and therefore for the welfare cost of low inflation. We introduce capital in the model and establish that under bargaining, both capital and the quantities traded are inefficiently low. These inefficiencies tend to increase the welfare cost of inflation significantly. Finally, we illustrate how endogenous participation decisions can mitigate or exacerbate the cost of inflation, and we provide calibrated examples in which a deviation from the Friedman rule is optimal.

Keywords: Inflation, search, money.

J.E.L. Classification: E40, E50.

* We have benefited from the comments of Ray Batina, Charles Carlstrom, Richard Dutu, Sebastien Lotz, Ed Nosal, Randall Wright and Christopher Waller as well as seminar participants at the universities of Akron and Paris 2 (Pantheon-Assas), at the Banque de France and at the Federal Reserve Banks of Cleveland and Richmond. We thank Patrick Higgins for his research assistance and Monica Crabtree-Reusser and Michele Lachman for editorial assistance. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

1. Introduction

Assessing the welfare costs of inflation requires a sound understanding of the benefits of monetary exchange. The search theory of money, developed in the last 15 years from the pioneering works of Kiyotaki and Wright (1989, 1993), offers such a framework. However, the first generation of search models of money were based on assumptions that were too restrictive to deliver useful insights for monetary policy (goods and money were indivisible; individuals' portfolios were limited to one unit of one object, and so forth). These severe restrictions have been relaxed in several recent extensions of the theory, by Shi (1997, 1999), Molico (1999), Faig (2004) and Lagos and Wright (2005), allowing the search model of money to be integrated with standard neoclassical growth theory (Shi, 1998; Aruoba, Waller, and Wright, 2006). This integration with mainstream macroeconomics has opened up the perspectives for a better understanding of the costs —and also maybe benefits— associated with inflationary finance.

As an example, Lagos and Wright provide estimates for the annual cost of 10 percent inflation ranging from 1.4 percent of GDP to 4.6 percent of GDP. These numbers are significantly larger than estimates based on the traditional method developed by Bailey (1956), which consists of computing the area underneath a money demand function. For instance, Lucas (2000), using Bailey's approach, estimates the cost of 10 percent inflation at slightly less than 1 percent of GDP.[1]

In this paper, we clarify and extend recent findings regarding the cost of inflation in search environments. Our approach consists of relating the measures of the welfare cost of inflation obtained from a search-theoretic model of monetary exchange —we use the formulation by Lagos and Wright (2005)— to the traditional measures based on the Bailey-Lucas methodology. We show the conditions under which the two measures are consistent, and those under which they differ. We also disentangle the various effects of

---

[1] Based on this methodology, Fischer (1981) and Lucas (1981) obtained estimates for the cost of 10 percent inflation ranging from 0.3 percent of GDP to 0.45 percent of GDP.

inflation by considering different extensions of the search model: We allow for different pricing mechanisms, participation decisions, and a choice to accumulate capital.

Our first result establishes that the traditional estimates for the welfare cost of inflation provided by the Bailey-Lucas methodology —the area underneath money demand— can be rationalized by a particular version of the search model. If money holders can appropriate the marginal social return of their real balances, the welfare cost of inflation as predicted by the search model is essentially the same as the area underneath the money demand function. This condition is satisfied when buyers have all the bargaining power to set prices in bilateral trades or when pricing is competitive. Any discrepancy between the estimates from the search model and the previous estimates in the literature arises from the use of different strategies to fit money demand.

Our second contribution is to clarify why alternative pricing mechanisms can exacerbate the welfare cost of inflation. We consider a simple rent-sharing rule (the proportional bargaining solution) and we establish a relationship between the cost of inflation, the area underneath the money demand function, and the buyer's market power. If the surplus of a trade is shared evenly between the buyer and the seller in a match, the welfare cost of inflation is twice as large as the estimate provided by the area underneath money demand. When sellers' bargaining power is chosen to be consistent with a realistic markup, the cost of 10 percent inflation is about 2.5 percent of GDP. The Bailey-Lucas measure of the cost of inflation is inaccurate because of a rent-sharing externality when buyers do not have all the bargaining power in bilateral matches. We also discuss the ability of the Friedman rule to generate the first best allocation under various pricing mechanisms. Under the generalized Nash solution or under a constant-markup pricing policy, the quantities traded are always inefficiently low (provided that buyers do not have all the bargaining power), which matters when quantifying the benefits of optimal deflation.

Third, we introduce a decision to accumulate capital that affects sellers' productivity in bilateral matches. If the terms of trade are determined according to some bargaining solution, individuals are subject to a double-holdup problem: they do not get the full marginal return of their real balances (away from the Friedman rule), and they do not get the full marginal return of their capital stock. As a consequence of these two inefficiencies, the welfare cost of inflation is larger in the presence of capital accumulation. For all our calibrated examples, the welfare cost of 10 percent inflation is larger than 5 percent of GDP.

Finally, we extend the model by introducing participation decisions and search externalities. We show that the Friedman rule may no longer be optimal and we illustrate how the presence of search frictions can mitigate or exacerbate the welfare cost of inflation. In the case of a proportional bargaining solution, the Friedman rule is not optimal and the welfare cost of 10 percent inflation is less than 0.5 percent of GDP for plausible values of the markup. This result, however, is sensitive to the choice of the pricing mechanism.

The paper is organized as follows: In Section 2, we present the methodology that Bailey (1956) developed to compute the welfare cost of inflation. In Section 3, we use a search-theoretic model to provide microfoundations for the Bailey-Lucas methodology. In Section 4, we discuss the robustness of the results to alternative pricing mechanisms, and in Section 5, we introduce capital accumulation. Section 6 considers participation decisions.[2]

2. The "Welfare Triangle"

The traditional approach to measuring the cost of inflation was developed by Bailey (1956) who measures the welfare cost of inflation by calculating the area underneath a money demand curve over an appropriate interval. In Figure 1, we plot the (inverse) demand for real balances, where the cost of holding real balances (the nominal interest rate) is represented on the vertical axis. The demand for real balances is downward-sloping because individuals reduce their money balances and resort to alternative payment arrangements, such as credit or barter, as the interest rate increases. The area underneath the money demand relationship over the interval $[m_1, m^*]$, the "triangle" *ABC* in Figure 1, measures the welfare cost of having a positive interest rate r_1 instead of zero. (In this analysis, the interest rate is assumed to vary one-to-one with the inflation rate.) Obviously, the welfare cost of inflation is minimized when the nominal

[2] A more detailed presentation of the models and data is available in Craig and Rocheteau (2006).

interest rate is zero.[3] This corresponds to the Friedman (1969) rule for optimal monetary policy. In what follows, we will measure the cost of inflation as the cost of raising the interest rate from r_0=3%, interpreted as the interest rate consistent with zero inflation, to r_1, say, the interest rate associated with 10 percent inflation. Graphically, this cost is measured by the area *ABDE*.

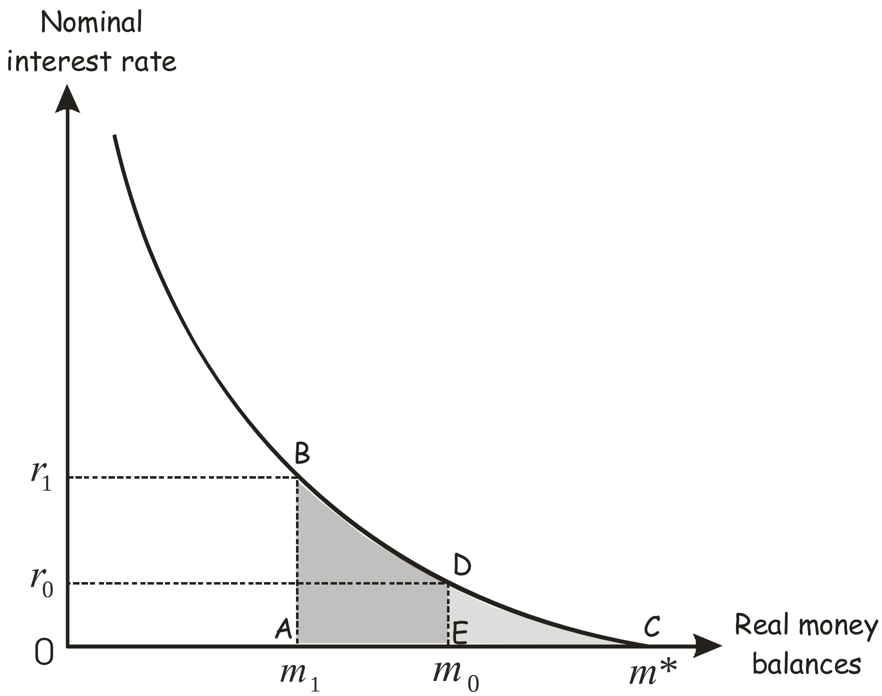

Figure 1. The welfare triangle

Money demand is then defined as the aggregate money balances M1 (currency and demand deposits), divided by nominal gross domestic product.[4] The nominal interest rate, *r*, is measured by the short-term commercial paper rate. In Figure 2, we represent each observation (*m,r*) with a circle for the period 1900-2000.

---

[3] Since the interest rate is approximately the sum of a constant real interest rate and the inflation rate, the Friedman rule would imply that the inflation rate is negative and approximately equal to the opposite of the real interest rate.

[4] Alternatively, several authors, including Fischer (1981), define money as high-powered money. Faig and Jerez (2005) exclude the currency in circulation abroad. For a discussion of the appropriate definition of money in this context, see Lucas (1981) and Marty (1997). By measuring real balances as a fraction of domestic output, the area of the money triangle can be interpreted as the fraction of income that is needed to compensate individuals for an interest rate of r_1 instead of zero (Lucas, 2000).

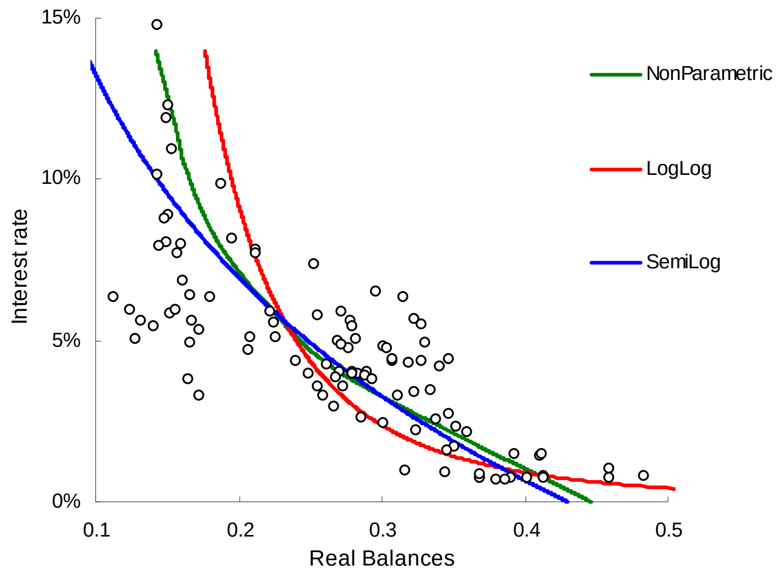

Figure 2. Fitting money demand

To measure the welfare triangle, one estimates a curve that fits the observations in Figure 2 and then computes the appropriate area underneath the implied money-demand curve. Lucas (2000) considers two specifications for money demand: the log-log specification, $m(r) = Ar^{-\eta}$, where m is aggregate real balances divided by output, r is the interest rate, and A and η are two estimated parameters; and the semi-log specification, where $m(r) = Ae^{-\eta r}$. In order to estimate parameters A and η, we use nonlinear least squares.[5] We also estimate the money demand curve by using a kernel regression.[6] In Table 1, we report the estimated parameters for the different specifications as well as the cost of 10 percent inflation (Δ) for each specification.

[5] We regress real balances on the nominal interest rate. This method is different from the one used by Lucas (2000), who constrains the curves to pass through the geometric means of the data and who uses a visual test to identify the best fit.
[6] The kernel was estimated with a local bandwidth computed using plug-in techniques, modified at each boundary. See Brockman *et al.* (1993).

	A	η	R^2	Δ (%)
$m(r)=A^{-\eta r}$	0.10	0.30	0.62	0.66
$m(r)=A\text{Exp}[-\eta r]$	0.43	11.03	0.67	1.51
Non-parametric			0.68	1.04

Table 1. Estimates of the parameters.

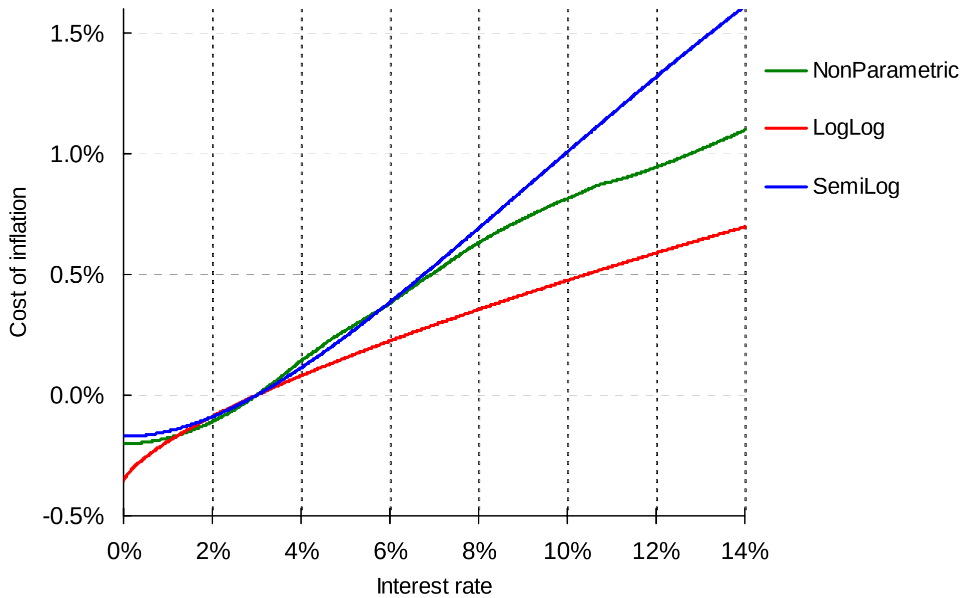

Figure 3. Welfare cost of inflation (Bailey method)

It can be seen in Table 1 and Figure 3 that the welfare cost associated with 10 percent inflation is quite different across specifications for the money demand function, ranging from slightly more than 0.5 percent to 1.5 percent. These differences simply reflect different ways to fit the data. The kernel regression evaluates the cost of 10 percent inflation at about 1 percent of GDP, which is similar to Lucas's measure. The estimate from the semi-log specification, about 1.5 percent of GDP, is comparable to Lagos and Wright's (2005) smallest estimate of the welfare cost of inflation.[7]

[7] These estimates are almost invariant to the time period, provided that the initial year is before 1940. If, however, we start the period of time after 1950, the welfare cost of inflation gets smaller. For instance, if we use data later than 1970 the cost of 10 percent inflation can be less than 0.1 percent of GDP.

3. Search-theoretic foundations for the "welfare triangle"

The objective of this section is to show that the search of monetary exchange pioneered by Kiyotaki and Wright (1989, 1993) can be used to provide microfoundations for the Bailey methodology to measure the welfare cost of inflation. In contrast to the Bailey approach, the search model spells out explicitly the benefits of monetary exchange for society and it provides both microfoundations for the demand for real balances and a measure of welfare.

We adopt the search framework of Lagos and Wright (2005), in which trades take place under different market structures.[8] The economy is composed of a unit-measure of agents. Time is discrete, and each period of time is divided into two subperiods, day and night. During the day, trades take place in a decentralized market where agents are matched bilaterally. The probability of a single-coincidence-of-wants meeting in which an agent meets someone who produces a good he likes is $\sigma \leq ½$. There are no double-coincidence-of-wants meetings. So, with probability σ, an agent is a buyer in a bilateral match, with probability σ he is a seller and probability 1-2σ he is unmatched. Night-trades take place in a competitive market.

An agent's utility function is

$$u(q^b) - c(q^s) + x,$$

where q^b is the consumption and q^s the production in a bilateral match, and x is the net consumption in the centralized market (x is negative if an agent produces more at night than he consumes).[9] The discount factor is $\beta = (1+\rho)^{-1} \in (0,1)$. The stock of fiat money at the beginning of period t is denoted M_t. Money is introduced into the economy

[8] Alternatively, one could use the model proposed by Shi (1997), in which households are composed of a large number of members who pool their money holdings at the end of each period. This model has been used for calibration purposes by Shi (1998) and Wang and Shi (2005).

[9] The linearity of the preferences is what guarantees that the distribution of real balances at the beginning of each period is degenerate. Notice that Lagos and Wright (2005) only require quasi-linear preferences.

through lump-sum transfers in the centralized market, and the supply of money is growing at rate π.

A social planner who can dictate the quantities to be produced and consumed in the decentralized market would choose $q^b = q^s = q^*$, where q^* solves $u'(q^*) = c'(q^*)$.

Turning to the equilibrium, let p_t denote the price in the centralized market. The expected discounted utility of an agent holding m_t units of money during the night of period *t*, denoted by $W(m_t / p_t)$, satisfies

$$W\left(\frac{m_t}{p_t}\right) = \max_{x_t, \hat{m}_{t+1}}\left\{x_t + \beta V\left(\frac{\hat{m}_{t+1}}{p_{t+1}}\right)\right\},$$

$$\text{s.t. } p_t x_t + \hat{m}_{t+1} = m_t + T_t,$$

where $V(\cdot)$ is the expected utility of the individual in the decentralized market. The individual receives a lump-sum transfer $T_t = \pi M_t$ and chooses his net consumption of night-goods, x_t, and his money balances, $\hat{m}_{t+1}$, in the next period. Substituting x_t by its expression from the budget constraint into the Bellman equation, it is straightforward to show that $W(m_t / p_t)$ is linear and that the choice of $\hat{m}_{t+1}$ is independent of m_t. Because $V(\cdot)$ will be strictly concave over the relevant range, the distribution of money balances at the beginning of each period is degenerate.

The utility of an agent in the decentralized market satisfies

$$V\left(\frac{m_t}{p_t}\right) = \sigma\left[u(q^b) + W\left(\frac{m_t}{p_t} - d^b\right)\right] + \sigma\left[-c(q^s) + W\left(\frac{m_t}{p_t} + d^s\right)\right]$$
$$+ (1-2\sigma)W\left(\frac{m_t}{p_t}\right),$$

where (q^b,d^b) are the terms of trade when the agent is the buyer, (q^s,d^s) are the terms of trade when the agent is the seller, q is the output, and d is the real transfer of money from the buyer to the seller (expressed in terms of the night-good). For the pricing mechanisms we will consider in this paper, terms of trade only depend on the money balances of the buyer in the match. Therefore, the terms of trade (q^s,d^s) are essentially taken as given by the agent. Furthermore, in equilibrium, individuals do not bring more money than what they need to trade in the decentralized market, $d^b = m_t / p_t$. Denote as $z(q) = m_t / p_t$ the real balances that an agent must hold in order to buy the quantity q in a bilateral match. The specific form for *z(q)* will depend on the assumed pricing mechanism in the decentralized market. The Bellman equation above can be rewritten as

$$V\left(\frac{m_t}{p_t}\right) = \sigma\left[u(q^b) - z(q^b)\right] + \sigma\left[-c(q^s) + d^s\right] + W\left(\frac{m_t}{p_t}\right),$$

with $z(q^b) = m_t / p_t$. Substitute $V(m_t / p_t)$ into the Bellman equation for $W(m_t / p_t)$ and rearrange in order to obtain

$$q^b = \arg\max_q \{-rz(q) + \sigma[u(q) - z(q)]\},$$

where $r = (1+\rho)(1+\pi) - 1$ at the steady state. This equation says that an agent chooses the quantity q to consume in the decentralized market so as to maximize the expected surplus he gets as a buyer, $\sigma[u(q) - z(q)]$, minus the cost of holding real balances, $rz(q)$. Since all agents hold the same money balances, $q = q^b = q^s$, a monetary equilibrium is a $q>0$ solution to the problem above.

In order to calibrate the model, we adopt the functional forms $u(q) = q^{1-\eta}/(1-\eta)$, where $\eta \geq 0$, and $c(q) = q$. Furthermore, the matching probability σ is set to ½ , so that each agent trades with probability one. The money demand function, *L(r)*, is defined as aggregate money balances divided by aggregate nominal output. It is equal to $L(r) = z(r)/[\sigma z(r) + A]$, where A is the real output in the centralized market (this quantity is indeterminate in the model, but it will be determined by the data).[10]

One needs to take a stand on how prices (or terms of trade) are determined in the decentralized day-market. In order to provide microfoundations for the Bailey methodology, we will assume that the monetary transfer from the buyer to the seller is such that the seller is exactly compensated for his production cost, $z(q) = c(q)$. This bargaining solution, called the *dictatorial* solution, is the outcome of a game in which the buyer makes an offer which the seller can accept or reject. If the offer is rejected, no trade

[10] Since agents readjust their real balances in the centralized market, the output A must be at least equal to σz. Under a quasi-linear utility function $U(x) - y$, where x is the consumption and y the production in the centralized market, the output in the centralized market is determined by $U'(x) = 1$. For instance, if $U(x) = A\ln x$, then $x = A$.

takes place. (We discuss alternative pricing mechanisms in the following section). The first-order condition for the choice of q is then

$$\frac{r}{\sigma} = \frac{u'(q)}{c'(q)} - 1.$$

Thus the equilibrium allocation is socially efficient when $r=0$ (i.e., the Friedman rule is implemented). Notice that for a given $r>0$, search frictions, as represented by σ, exacerbate the wedge introduced by a positive nominal interest rate. Furthermore, the quantity q traded in bilateral matches is a decreasing function of the average (opportunity) cost of holding money, r/σ.

After some calculation, aggregate real balances satisfy

$$L = \left[\sigma + A\left(1 + \frac{r}{\sigma}\right)^{\frac{1}{\eta}} \right]^{-1}.$$

The parameters A and η are estimated from the data for the U.S. economy from 1900 to 2000. We find $A=1.82$ and $\eta=0.14$.

In order to measure the welfare cost associated with a given interest rate, r, relative to 3 percent (the interest rate consistent with zero inflation), we ask the following question: What is the fraction Δ of total consumption that individuals would be willing to sacrifice in order to be in the steady state associated with an interest rate of 3 percent instead of the one associated with r? Let q_r denote the quantities traded in a steady state when the interest rate is r. The cost of inflation Δ solves

$$\sigma\left[u\left[q_{0.03}(1-\Delta)\right] - c(q_{0.03})\right] - A\Delta = \sigma\left[u(q_r) - c(q_r)\right].$$

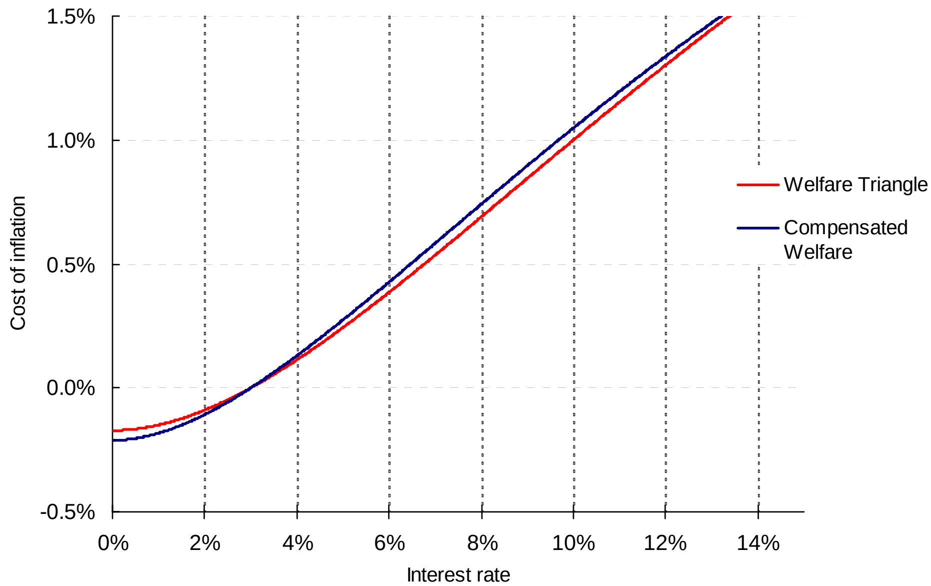

Figure 4. Compensated welfare and the welfare triangle

In Figure 4, we compare the two measures of the welfare cost of inflation, namely, the compensated measure and the welfare triangle measure. (The welfare triangle measure is the area underneath the money demand function as estimated from the search model.) Figure 4 shows that the two measures are nearly identical. In order to understand this result, consider the individual demand for real balances given by

$$r = \sigma\left[u'(q)\frac{dq}{dz} - 1\right],$$

where $dq/dz = 1/z'(q)$. Compute the area underneath this (individual) money demand function over the interval $[z_0, z_1]$,

$$\int_{z_0}^{z_1} r(z)dz = \sigma\{u[q(z_1)] - z_1\} - \sigma\{u[q(z_0)] - z_0\}.$$

Using the fact that $z = c(q)$, it is easy to see that the right-hand side of the previous expression is just the change in steady-state welfare. So the area underneath the individual demand for real balances coincides with the change in steady-state welfare.[11]

The welfare cost of a nominal interest rate of 13 percent relative to a 3 percent interest rate is about 1.5 percent. This estimate is bigger than those of Lucas (2000) but it agrees with the nonlinear least square estimate based on the semi-log specification. The difference between the numbers simply stems from different strategies for fitting the points in the data.

We can relate this estimate of the cost of inflation to the estimate that would be obtained under competitive pricing in the decentralized market. To this end, reinterpret the idiosyncratic trading shocks as preference and productivity shocks. With probability σ, an agent wants to consume during the day; with probability σ, he has the ability to produce during the day; and with $1 - \sigma$, he neither consumes nor produces. Denote ω the dollar price of goods in the decentralized market. The choice of q^b obeys

$$q^b = \arg\max_q \left\{ -r\frac{\omega q}{p} + \sigma\left[u(q) - \frac{\omega q}{p} \right] \right\}.$$

The first-order condition gives $u'(q^b) = \left(1 + r/\sigma\right)\omega/p$. Sellers in the decentralized market maximize $-c(q^s) + q^s\,\omega/p$, which gives $c'(q^s) = \omega/p$. Given that there is the same measure σ of buyers and sellers, market clearing requires $q^b = q^s = q$ and therefore

$$\frac{u'(q)}{c'(q)} = 1 + \frac{r}{\sigma}.$$

This is the same equation as the one obtained under the assumption that buyers make a take-it-or-leave-it offer. If prices are determined competitively, buyers can obtain the marginal return of their real balances so that the welfare cost of inflation coincides with the Bailey-Lucas measure.[12]

---

[11] These two measures do not exactly coincide in figure 4 because we express real balances as a fraction of aggregate output and we look not at the change in steady-state welfare but at a compensated measure of welfare.

[12] Assuming a competitive pricing mechanism, Rocheteau and Wright (2004, 2005) and Reed and Waller (2004) find similar estimates for the welfare cost of inflation, between 1 percent and 1.5 percent of GDP.

In summary, the search model of money can, for some pricing mechanisms, provide theoretical foundations for the Bailey-Lucas methodology to estimate the cost of inflation. In this case, our estimate is in the same ballpark as previous studies, that is, around 1 percent of GDP per year.

4. Inflation and pricing

The estimate for the welfare cost of inflation reported in the previous section was obtained under the assumption that buyers can extract the whole (marginal) surplus from trade. In this section, we describe alternatives in which sellers get a fraction of the surplus of the match. We consider first a simple bargaining solution, called the *proportional solution*, which will illustrate the role played by the pricing mechanism in assessing the welfare cost of inflation. The proportional bargaining solution requires that the buyer obtains a constant fraction, θ, of the surplus of a match, $u(q)-c(q)$. Therefore, $u(q)-z(q)=\theta\left[u(q)-c(q)\right]$ and $z(q)=\theta\, c(q)+(1-\theta)u(q)$. As before, the quantity traded in individual matches solves

$$q=\arg\max\left\{-rz(q)+\sigma\left[u(q)-z(q)\right]\right\}.$$

When $z(q)$ is replaced by its new expression, the first-order condition for the choice of q is

$$\frac{r}{\sigma\theta}=\frac{u'(q)-c'(q)}{\theta c'(q)+(1-\theta)u'(q)}.$$

The first-best allocation (where $u'(q)=c'(q)$) is achieved when the Friedman rule is implemented, $r=0$.[13] The monetary wedge generated by a positive nominal interest rate, the left-hand side of the equation above, is amplified when search frictions become more severe (a lower σ) or when sellers have more market power (a lower θ).

We estimate the parameters A and η of money demand as before. The parameter θ of the bargaining solution can be chosen so as to generate a markup $1+\mu$ (price over marginal cost) consistent with the data. In the model, the real marginal cost is $c'(q)$, and

[13] For a monetary equilibrium to exist, r must be smaller than $\sigma\theta/(1-\theta)$.

the real price in the decentralized market is $z(q)/q$. Therefore, the markup in the decentralized market is $z(q)/[c'(q)q]$, which for our specification for the utility function, yields

$$1+\mu=\theta+\frac{(1-\theta)}{(1-\eta)q^{\eta}}.$$

We target a markup of 20 percent for a value of q associated with 2 percent inflation. We find $\theta=0.61$. For the sake of comparison with the results in Lagos and Wright (2005), we also compute θ that generates an average markup for the entire economy equal to 10 percent. We find $\theta=0.33$.[14] In Figure 5, we also report the welfare cost of inflation for $\theta=1$ (the benchmark studied in the previous section) and $\theta=0.5$ (the egalitarian solution).

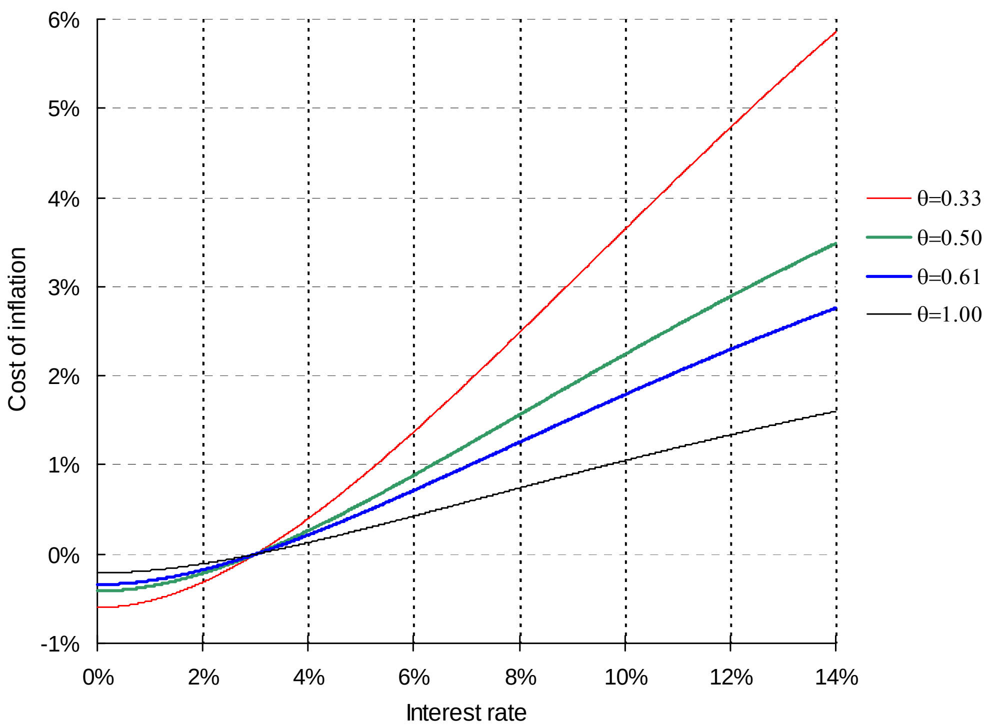

Figure 5. Cost of inflation under proportional bargaining

[14] Since the markup in the centralized sector is one, the markup in the decentralized market must exceed the one we find when we target only the markup in the decentralized market.

Pricing mechanisms	*Market power*	*Coefficients of L(r)*		*Allocation: q*			*Cost of inflation*
	θ or μ	***A***	η	***r=0***	***r=3%***	***r=13%***	*Δ(%)*
Proportional bargaining	1.00	1.82	0.14	1.00	0.67	0.20	1.51
	0.61	2.05	0.22	1.00	0.65	0.15	2.60
	0.50	2.19	0.26	1.00	0.63	0.13	3.28
	0.33	2.64	0.37	1.00	0.60	0.07	5.49
Nash bargaining	1.00	1.82	0.14	1.00	0.67	0.20	1.51
	0.60	1.80	0.23	0.88	0.56	0.16	2.70
	0.50	1.77	0.26	0.81	0.50	0.14	3.35
	0.34	1.67	0.35	0.62	0.36	0.09	5.09
Constant Markup	1.00	1.82	0.14	1.00	0.67	0.20	1.51
	1.10	1.04	0.14	0.52	0.34	0.10	2.50
	1.20	0.62	0.14	0.28	0.19	0.06	3.32

Table 2: Inflation and pricing.

When the buyer's share is less than 1, the cost of inflation typically is higher than the measure given by the welfare triangle. When we match the markup (*θ=0.61*), the welfare cost of 10 percent inflation is about 2.5 percent of GDP. As one varies the buyer's share from 1/3 to 1, the cost of 10 percent inflation varies from 1.5 percent to 5.5 percent of GDP. More generally,

$$\text{Cost of inflation} \approx \frac{\text{Area of welfare triangle}}{\theta}.$$

For instance, if the buyer's share is 50 percent, the welfare cost of inflation is about twice the area of the money triangle (see Figure 6). To understand this result, consider the area underneath the individual demand for real balances, which satisfies

$$\int_{z_0}^{z_1} r(z)dz = \sigma\{u[q(z_1)] - z_1\} - \sigma\{u[q(z_0)] - z_0\}$$
$$= \theta\sigma\{u[q(z_1)] - c[q(z_1)]\} - \theta\sigma\{u[q(z_0)] - c[q(z_0)]\}.$$

It is equal to the change in steady-state welfare multiplied by θ.

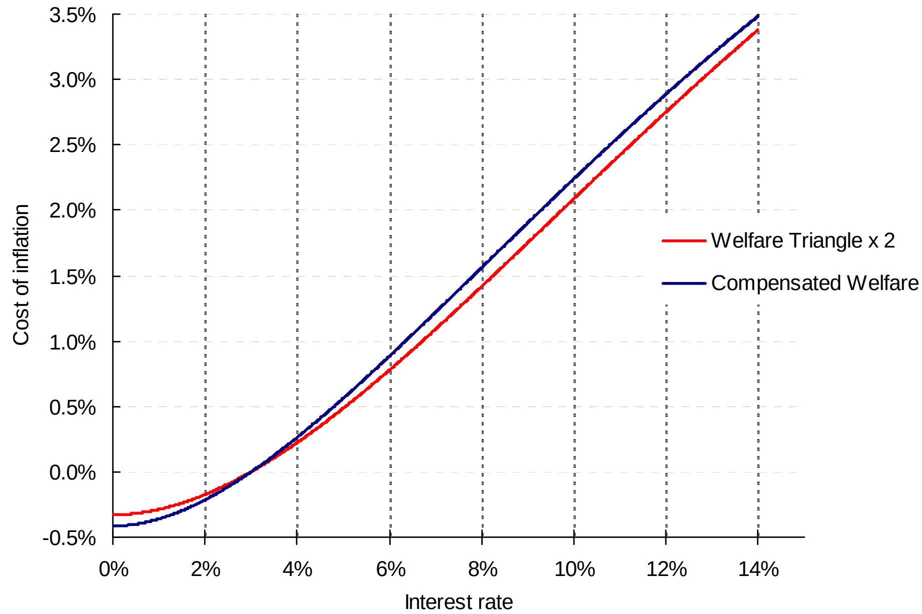

Figure 6. Welfare triangle vs compensated welfare (θ=0.5)

To understand why the welfare triangle can underestimate the true welfare cost of inflation, consider the following example: If each unit of good produced in a bilateral match is worth \$1 to the buyer and costs \$0.9 to produce, the marginal surplus of a trade is \$0.1. Suppose that the price is \$0.95, so that both the buyer and the seller get a surplus of \$0.05. The private return of money is equal to the buyer's surplus divided by the amount of money that the buyer must carry to buy the good, 0.05/0.95=5.2 percent. The social return of money is the total surplus divided by the price of the good, 0.1/0.95=10.5 percent. If the interest rate is 10 percent, the buyer's cost of holding \$0.95 exceeds the marginal gain of \$0.05. So the buyer has no incentive to bring an additional dollar, even though the social return of this dollar is greater than the opportunity cost incurred by the buyer.

The discrepancy between the private and social benefits of real balances arises from a *rent-sharing externality* generated by any pricing rule that stipulates that the buyer does not get the full marginal return of his real balances.[15] The marginal benefit of the

[15] This rent-sharing externality is closely related to holdup problems noted in the investment literature. We adopt the terminology "rent-sharing externality" to distinguish it with the holdup problem emphasized in

real balances is greater from society's viewpoint than from the buyer's. Since money demand captures only the marginal benefit of money from the buyer's side, the welfare triangle misses a fraction of the welfare cost of inflation.

Formally, a marginal unit of real balances allows a buyer in a bilateral match to buy $\partial q / \partial z$ units of goods. The expected private marginal utility of real balances is then $\sigma[u'(q) - z'(q)]dq / dz = \sigma[u'(q) / z'(q) - 1]$, which is precisely r from the first-order condition for the choice of real balances. The expected social marginal utility of real balances is

$$\sigma[u'(q) - c'(q)]\frac{dq}{dz} = \sigma\left[\frac{u'(q) - c'(q)}{z'(q)}\right] = \frac{\sigma}{\theta}\left[\frac{u'(q) - z'(q)}{z'(q)}\right] = \frac{r}{\theta}.$$

Note that the private and social returns of real balances are equal if and only if θ=1, a point illustrated in Figure 7. For a given stock of real balances m_0, the marginal benefit that money provides to the buyer, represented by the length of segment *AB*, is less than its marginal social benefit, the length of segment *AD*. If prices are determined according to a proportional solution, the ratio *BA*/*DA* is the buyer's share. As a consequence, when measuring the *ABC* area, one underestimates money's social benefit, the *ADC* area, by a factor equal to the inverse of the buyer's share.

Lagos and Wright (2005). In particular, this rent-sharing externality does not generate an inefficiently low output at the Friedman rule.

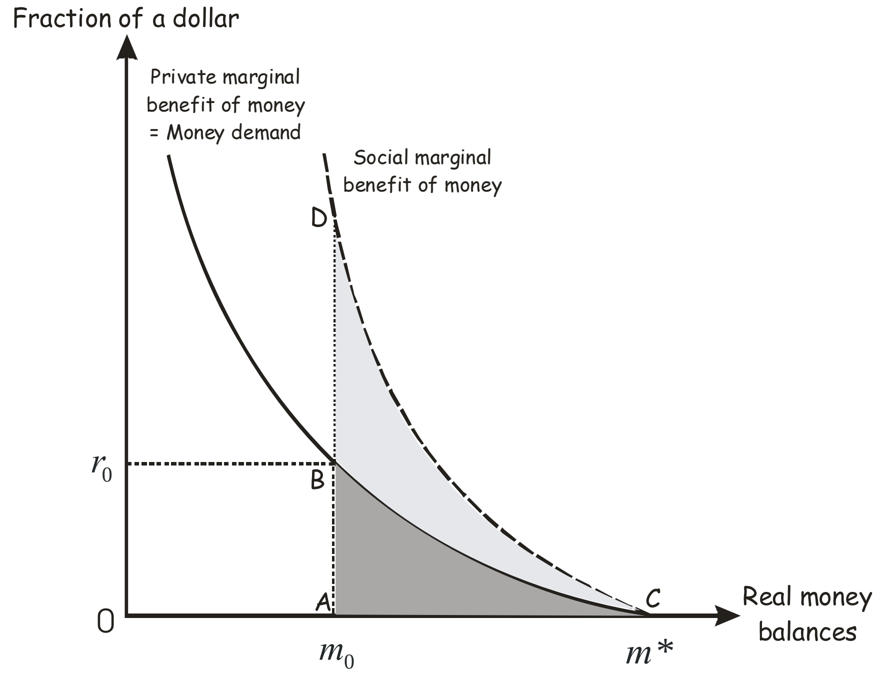

Figure 7. The rent-sharing externality

In Figure 7, the curves representing the private and social benefits of real balances intersect the horizontal axis (*r=0*) for the same value $m*$ of real balances. At this point, the benefits of holding real balances are maximized for both buyers and society. This observation implies that the Friedman rule yields the best allocation q^* for all values of θ. In Table 2, *q=1* at *r=0* for all θ.

However, the previous result, where the Friedman rule generates a first-best allocation, does not hold for all bargaining solutions. For instance, as emphasized in the search-money literature, it does not hold for the Nash solution, which implies that $z(q) = \Theta(q)c(q) + [1 - \Theta(q)]u(q)$ with

$$\Theta(q) = \frac{\theta u'(q)}{\theta u'(q) + (1-\theta)c'(q)}.$$

The parameter θ now denotes the bargaining power of the buyer, whose share, $\Theta(q)$, depends on both θ and *q*, and it is equal to θ when $q = q^*$.[16] In particular, Θ decreases as

[16] The bargaining power θ, which varies from 0 to 1, is a measure of the buyer's strength in the bargaining process. In an explicit bargaining game with offers and counteroffers, an individual's bargaining power depends, among other things, on his ability to terminate the negotiation if his offer is rejected.

q increases. Under the Nash solution, the Friedman rule is optimal but the quantities traded in the decentralized market are too low (in Table 2, $q<1$ at $r=0$ for all $\theta<1$). This result is illustrated in Figure 8. Under the Friedman rule ($r=0$), the economy's real balances are $\tilde{m}$, while the real balances that would maximize society's welfare are m^*. In other words, if the interest rate is zero, an individual's demand for real balances is satiated, even though the marginal benefits of money to society are still positive.

This inefficiency arises from the fact that the buyer's surplus is not necessarily monotonic with the size of the match surplus (the bargaining is said to be non-monotonic).[17] Put differently, the buyer's surplus $u(q)-z(q)$ reaches a maximum for a value of q smaller than q^*. As a consequence of this inefficiency, a small increase of the interest rate above $r=0$ will have a larger effect on welfare than it would under proportional bargaining. Indeed, the welfare cost of a small interest rate can be approximated by the change in real balances multiplied by the social benefits of real balances at $r=0$, the length of segment EC in Figure 8.

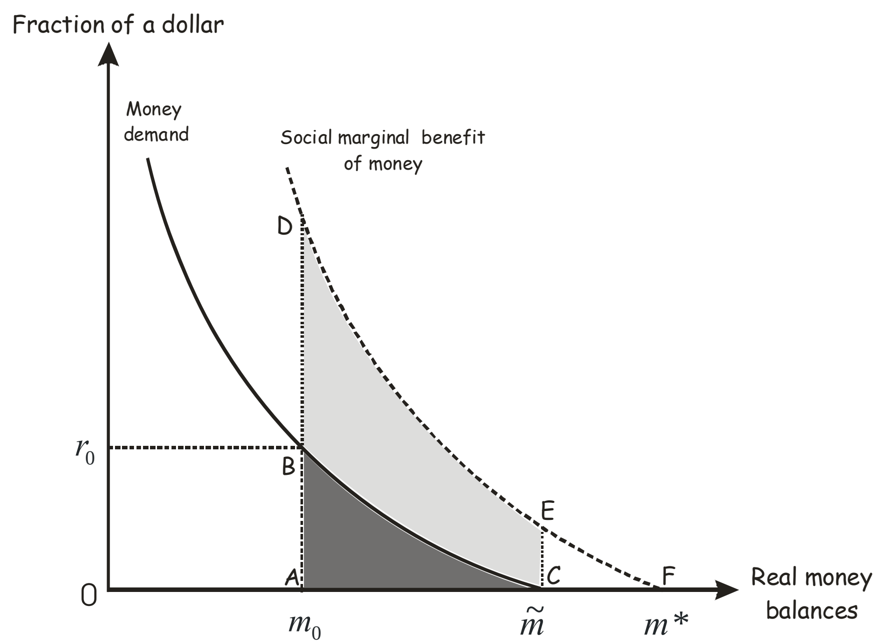

Figure 8. Bargaining inefficiencies

[17] For a detailed treatment of alternative bargaining solutions and their properties in monetary economies, see Rocheteau and Waller (2005).

Figure 9 plots the welfare cost of inflation when prices are determined according to the Nash solution for different values of the buyer's bargaining power, θ. The value of θ that corresponds to a markup of 20 percent in the decentralized market is θ=*0.60*, where the expression for the markup is

$$1+\mu=\frac{1-\eta\theta}{(1-\eta)\left[\theta+(1-\theta)q^{\eta}\right]}.$$

(To generate a 10 percent average markup for the entire economy, as in Lagos and Wright (2005), θ must be set equal to 0.34.) A comparison of Figures 5 and 9 reveals that the welfare cost of 10 percent inflation under the Nash solution is of the same magnitude as the cost under the proportional bargaining solution (see also Table 2). In particular, when we target a 20 percent markup, the welfare cost of 10 percent inflation is slightly more than 2.5 percent of GDP. Under both bargaining solutions, a rent-sharing externality is at work, which amplifies the cost of inflation. However, under the Nash solution, the buyer's share, Θ, gets larger at higher inflation rates, so the rent-sharing externality gets smaller.

The most noticeable difference between Figures 5 and 9 is the gain associated with a reduction of the interest rate from 3 percent to zero (which corresponds to the optimal deflation rate). This gain can be as high as 2 percent of GDP when the buyer's bargaining power is 0.34. Under the proportional solution, the gain is about 0.5 percent of GDP. So the non-monotonicity inefficiency generated by the Nash solution is important in that it predicts large welfare gains if inflation is reduced from zero to the optimal deflation rate.

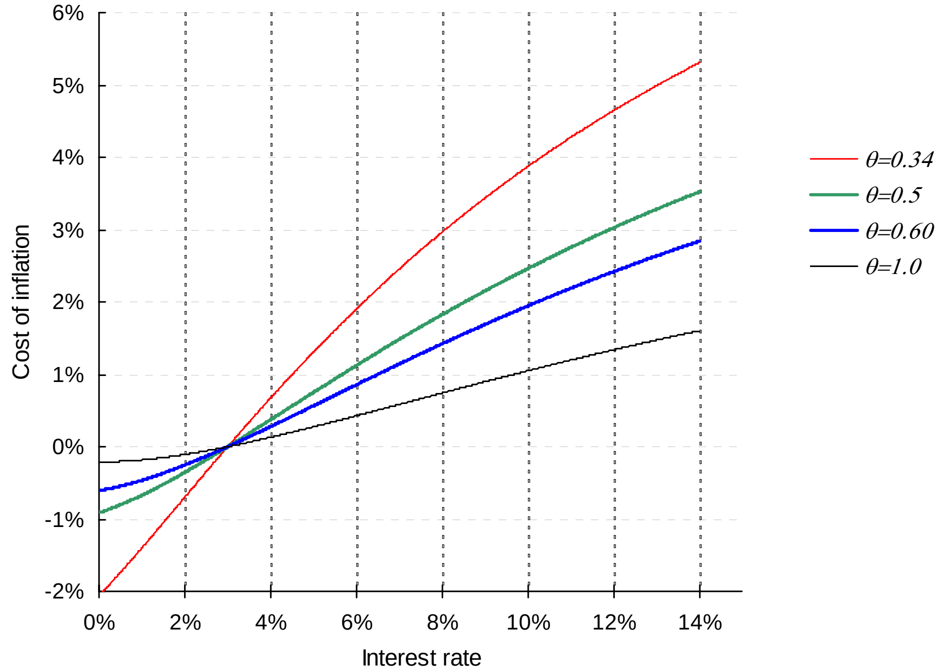

Figure 9. Cost of inflation under Nash bargaining

To conclude this section, we present a pricing mechanism which exhibits the same type of inefficiencies as the Nash bargaining solution but is based on the more familiar idea that prices are set as a markup over the cost incurred by sellers.[18] More precisely, the transfer of money from the buyer to the seller corresponds to the cost incurred by the seller in producing the amount asked for by the buyer multiplied by a constant factor, $1+\mu$ (where $\mu \geq 0$), which we interpret as the "markup," $z(q) = (1+\mu)c(q)$. As shown in Figure 10 (and Table 2), the cost of inflation increases with the markup. When the markup is 20 percent, the cost of 10 percent inflation is slightly less than 3 percent of GDP, which is similar to the prediction of the model under the symmetric Nash solution.

[18] For a search model with price posting by sellers, see Ennis (2004), who describes an economy in which buyers have private information about their tastes, and sellers make take-it-or-leave-it offers. The annual welfare cost of 10 percent inflation in this model is between 4 percent and 7 percent of GDP.

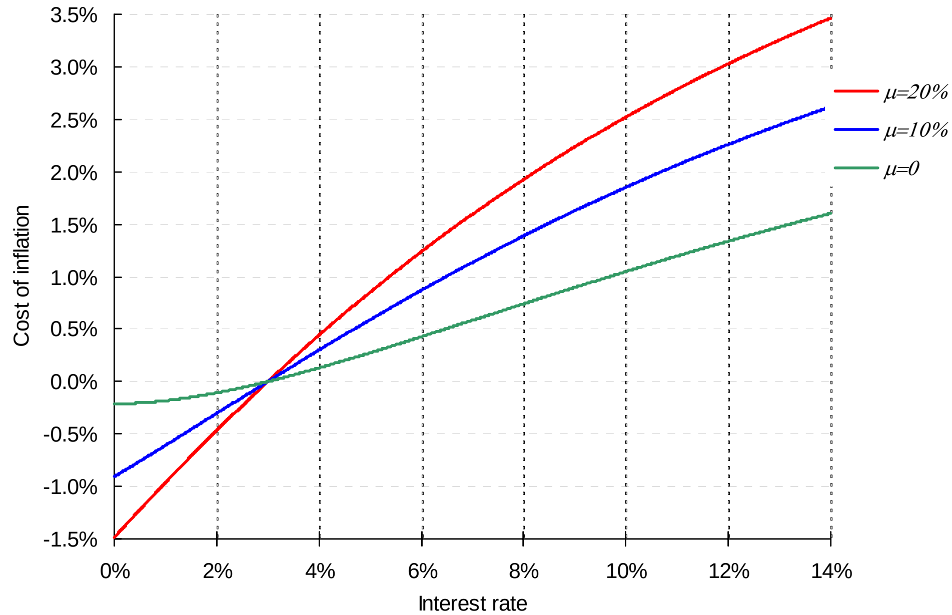

Figure 10. Cost of inflation under constant markup

Figure 11 illustrates how a constant markup affects assessment of the welfare cost of inflation. The constant markup, μ, shifts the curve that specifies the social return of real balances up by a constant amount (*BD=CE*). The larger the markup, the larger the difference between the private and social benefits of real balances. Also, it is clear from Figures 8 and 11 that both the Nash solution and pricing with constant markup induce qualitatively similar effects of inflation. In both cases, the quantities traded under the Friedman rule ($r=0$) are too low.

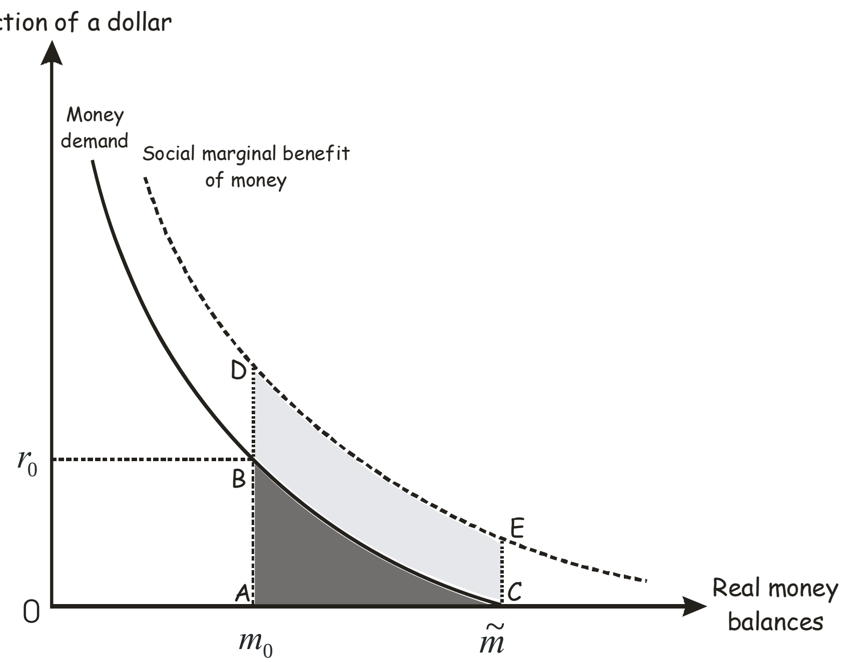

Figure 11. Constant markup

5. Inflation and capital[19]

Casual observations indicate that sellers must incur costs of purchasing capital or setting up stores before they can sell goods. If the terms of trade are determined according to some bargaining protocol, as assumed in the previous section, then buyers can hold up sellers on their initial investments. Focusing on the holdup problem on real balances only may create a biased view of the frictions in search environments; as such, it may affect our estimates of the welfare cost of inflation. Also, by omitting sellers' investment decisions, one neglects an important channel through which inflation can affect equilibrium allocation and therefore welfare.

In this section, we let agents choose a capital stock k that enters as an input of production technology in the decentralized market. We assume that the capital stock is accumulated in the centralized market: Each unit of good can be turned into a unit of

[19] We thank an anonymous referee who suggested this section to us. For a related model of money and capital, see Shi (1999) and Aruoba, Waller, and Wright (2006) who consider various pricing mechanisms and various ways in which capital can affect the decentralized market.

capital. Capital depreciates at rate δ between the decentralized market and the centralized market. The output produced by a seller in a bilateral match is now $q = q(k,h)$, where $q(\cdot,\cdot)$ is a constant-return-to-scale production function with two inputs, *k*, the capital stock of the seller, and *h*, his supply of hours. We impose the restriction that capital, or claims on capital, cannot be used as a means of payment in bilateral matches in the decentralized market.[20]

The utility of an agent holding m_t units of money and k_t units of capital when entering the centralized market of period *t*, denoted by $W(m_t/p_t, k_t)$, satisfies

$$W\left(\frac{m_t}{p_t}, k_t\right) = \max_{x_t, \hat{m}_{t+1}} \left\{ x_t + \beta V\left(\frac{\hat{m}_{t+1}}{p_{t+1}}, \hat{k}_{t+1}\right)\right\},$$
$$\text{s.t. } p_t x_t + \hat{m}_{t+1} + p_t \hat{k}_{t+1} = m_t + p_t k_t + T_t,$$

where $V(\cdot,\cdot)$ is the expected utility of an agent in the decentralized market. The wealth of an individual at the beginning of the centralized market is composed of the lump-sum transfer, $T_t = \pi M_t$, and his initial portfolio (m_t, k_t). The individual chooses his net consumption x_t, his money balances $\hat{m}_{t+1}$, and capital stock $\hat{k}_{t+1}$ in the next period. It is straightforward to check that $W\left(\frac{m_t}{p_t}, k_t\right) = \frac{m_t}{p_t} + k_t + W(0,0)$ and that the choice of $(\hat{m}_{t+1}, \hat{k}_{t+1})$ is independent of the initial portfolio (m_t, k_t). Thus, the distribution of money and capital is degenerate at the beginning of the decentralized market.

The utility of an agent in the decentralized market satisfies

$$V\left(\frac{m_t}{p_t}, k_t\right) = \sigma\left[u(q^b) + W\left(\frac{m_t}{p_t} - d^b, (1-\delta)k_t\right)\right] + \sigma\left[-c(h^s) + W\left(\frac{m_t}{p_t} + d^s, (1-\delta)k_t\right)\right]$$
$$+ (1-2\sigma)W\left(\frac{m_t}{p_t}, (1-\delta)k_t\right).$$

[20] Arguably, this restriction amounts to a cash-in-advance constraint. Aruoba, Waller, and Wright (2006) justify this restriction by assuming that physical capital is fixed in place in the centralized market, and therefore cannot be traded in the decentralized market. They rule out the circulation of claims to capital by assuming that such claims can be counterfeited at no cost. For a model of money and capital that places no restriction on which asset can be used as a means of payment, see Lagos and Rocheteau (2004).

The interpretation of this Bellman equation is similar to that in the previous section. The novelty is the fact that the terms of trade (h^s, d^s), in matches where the individual is a seller now depend on the agent's choice of capital. Also, notice that capital depreciates between the decentralized market and the centralized market. Using the linearity of $W(\cdot,\cdot)$, the previous Bellman equation can be simplified to

$$V\left(\frac{m_t}{p_t}, k_t\right) = \sigma\left[u(q^b) - d^b\right] + \sigma\left[-c(h^s) + d^s\right] + W\left(\frac{m_t}{p_t}, (1-\delta)k_t\right).$$

Substituting V by its expression into the Bellman equation for W, and after some simplification, the choice of the capital stock solves

$$\max_k\left[-(\rho+\delta)k + \sigma S^s(k,z)\right],$$

where $S^s(k,z) \equiv -c(h^s) + d^s$ is the seller's surplus from a trade as a function of his investment k and the buyer's real balances z (taken as given by the seller). So individuals choose their capital stock in order to maximize their expected surplus as sellers minus the user cost of capital. It is clear from the previous problem that if buyers have all the bargaining power to set terms of trade in bilateral matches, then $S^s = 0$ and $k=0$: The decentralized sector shuts down. This result is a manifestation of the holdup problem on capital.

Assuming that the terms of trade are determined in accordance with the proportional bargaining solution, the seller's surplus satisfies

$$S^s(k,z) = \max_{h,d}\left[d - c(h)\right]$$

$$\text{s.t.} \quad (1-\theta)\left[u\left[q(h,k)\right] - d\right] = \theta\left[d - c(h)\right]$$

$$d \le z.$$

So the seller chooses the terms of trade (h,d) to maximize his surplus, subject to the constraint that the buyer's surplus is a fraction $\frac{\theta}{1-\theta}$ of his own. We show in the appendix that the choice of capital satisfies

$$\rho + \delta = \sigma(1-\theta)u'(q)q_k(h,k)\left(1 - \frac{1-\theta}{\theta}\frac{r}{\sigma}\right),$$

where $q_k(h,k)$ is the partial derivative of q with respect to k. First, the extent of search frictions affects the return of capital, since an agent uses his capital stock only when he is the seller in a match, an event occurring with probability σ. Second, the seller's market power, as represented by $1-\theta$, also affects the private return of capital through a standard holdup problem. Third, an increase in inflation above the Friedman rule reduces buyers' real balances and therefore sellers' willingness to invest in capital. As shown in the Appendix, the term $\frac{1-\theta}{\theta}\frac{r}{\sigma}$ is the Lagrange multiplier associated with the constraint $d \leq z$. As r increases, or σ decreases, this constraint becomes more binding and the return of capital falls. Also, if θ falls, the holdup problem on real balances is more severe and the constraint $d \leq z$ becomes tighter, which reduces the return of capital.

Agents' choice of real balances is essentially the same as that described in the previous section. It gives

$$\frac{r}{\sigma} = \frac{\theta[u'(q)q_h(h,k) - c'(h)]}{(1-\theta)u'(q)q_h(h,k) + \theta c'(h)},$$

where $q_h(h,k)$ is the partial derivative of q with respect to h.

We next compare the equilibrium allocation with the allocation that a social planner would choose. The first-best allocation satisfies

$$\rho + \delta = \sigma u'(q)q_k(h,k)$$

$$u'(q)q_h(h,k) = c'(h)\,.$$

Comparing the equilibrium conditions and the first-order conditions of the planner's problem reveals various inefficiencies. First, there is the monetary wedge r/σ that distorts both the choice of hours in the decentralized market and the choice of capital. Second, there are two holdup problems on capital and real balances (captured by the terms θ and $1-\theta$) that tend to make both capital and the output in the decentralized market inefficiently low. The equilibrium allocation coincides with the first-best if the Friedman rule is implemented ($r=0$) and sellers have all the bargaining power ($\theta=0$).[21] However, this equilibrium is tenuous, since for all $r>0$ there is no monetary equilibrium when $\theta=0$.

We calibrate the model as in the previous section. Aggregate money demand is defined as

---

[21] Alternatively, one can consider the sequence (r_n, θ_n) such that $r_n \leq \sigma\theta_n/(1-\theta_n)$ for all n and $(r_n, \theta_n) \to (0,0)$ as n goes to infinity. For all n, there exists a monetary equilibrium and the equilibrium allocation approaches the first-best allocation as n tends to infinity.

$$L = \frac{z(q,h)}{\sigma z(q,h) + \delta k + A},$$

where q and h are functions of the nominal interest rate and

$$z(q,h) = \theta c(h) + (1-\theta)u(q).$$

Notice that aggregate output is defined as the output in the decentralized sector, σz, investment, δk, and consumption in the centralized sector, A.[22] We choose (A,η) to fit aggregate money demand to the data. We need to choose a production function and the depreciation rate δ. There is no hard evidence on what the technology should be in the decentralized market. Therefore, we report the results for the following CES production function:

$$q(h,k) = \left[\frac{2}{3}h^{0.5} + \frac{1}{3}k^{0.5}\right]^2.$$

We have chosen a relatively high elasticity of substitution because it facilitates the calibration of the model and the choice of θ that matches the markup. Finally, we put more weight on hours to capture the idea that the technology is intensive in hours in the decentralized market.[23] The depreciation rate is set equal to 4 percent. The welfare cost of inflation is computed as the change in consumption that individuals would be willing to sacrifice to avoid a given inflation rate. To make it comparable with our estimates in the previous sections we express it as a function of GDP.

Table 3 reports the first-best allocation for the different sets of parameter values that we consider in our calibration exercise. Table 4 reports the equilibrium allocation as well as the welfare cost of inflation. The comparison of Tables 3 and 4 illustrates the double holdup problems described earlier. When θ is high (e.g., $\theta=0.70$), the holdup problem on capital is severe, so the equilibrium capital stock is much smaller than the first-best capital stock. For lower values of θ (e.g., $\theta=0.27$), the equilibrium capital stock

[22] As before, the consumption in the centralized sector would be pinned downed, if we were to adopt a quasi-linear specification for the utility function in the decentralized market.

[23] We have tried several specifications for the production function, including a Cobb-Douglas one. Our findings are fairly robust across specifications.

gets closer to its first-best value but the effect on hours of an inflation increase is more pronounced.[24]

Let us turn to the welfare cost of inflation. Consider first the case of symmetric bargaining ($\theta=1/2$). As indicated in Figure 12 and Table 3, the welfare cost of 10 percent inflation is about 7.5 percent of GDP, a much larger estimate than the one given in the previous section. The intuition for this result is the following: when $\theta=1/2$, the buyer in the match gets only half of the marginal contribution of his real balances to the match surplus. This is the rent-sharing externality emphasized in the previous section. On the other hand, the seller gets only half of the contribution of his capital stock to the match surplus. As a result, sellers underinvest in capital, even when $r=0$. Inflation makes this inefficiency even worse, thereby exacerbating the welfare cost of inflation.

Similarly, if we choose θ so as to generate a markup of 20 percent in the decentralized sector, the welfare cost of 10 percent inflation is slightly less than 6 percent of GDP, which is also much higher than our previous estimates. To generate a markup of 20 percent, the buyer's share must be equal to $\theta=0.7$. For this value, the holdup problem on money is not too severe, but the holdup problem on capital is huge. As a consequence, capital tends to be inefficiently low, which exacerbates the cost of inflation.

Reciprocally, if we generate a 10 percent markup for the entire economy (including the centralized sector), then $\theta=0.27$. The holdup problem on capital is not too severe, but the one on money is large. Again, this makes the welfare cost of inflation very large, more than 10 percent of GDP, for the reasons emphasized in the previous section.

To summarize, in the presence of a seller's investment decision, the welfare cost of inflation is larger than the one found in the model without capital. This is so because the seller's capital choice generates an additional holdup problem, an inefficiency that can only be eliminated by giving all the bargaining power to sellers, which gets rid of the monetary equilibrium altogether.

---

[24] Under the Nash solution the model happens to be much harder to calibrate. The capital stock is very close to 0 for most parameter values and it is very insensitive to inflation.

η	q	k	h
0.16	3.92	14.50	1.14
0.21	2.78	10.27	0.80
0.39	1.72	6.37	0.50

Table 3. First-best allocation

Market power	***Coefficients of L***		***Allocation***						***Cost of inflation***
			r=0		***r=3%***		***r=13%***		
θ	**A**	η	***k***	***h***	***k***	***h***	***k***	***h***	***Δ(%)***
0.70	0.19	0.16	0.10	0.08	0.07	0.05	0.02	0.01	5.89
0.50	0.83	0.21	0.88	0.27	0.60	0.17	0.14	0.03	7.56
0.27	2.56	0.39	3.08	0.45	1.93	0.25	0.11	0.01	10.97

Table 4: Inflation and capital.

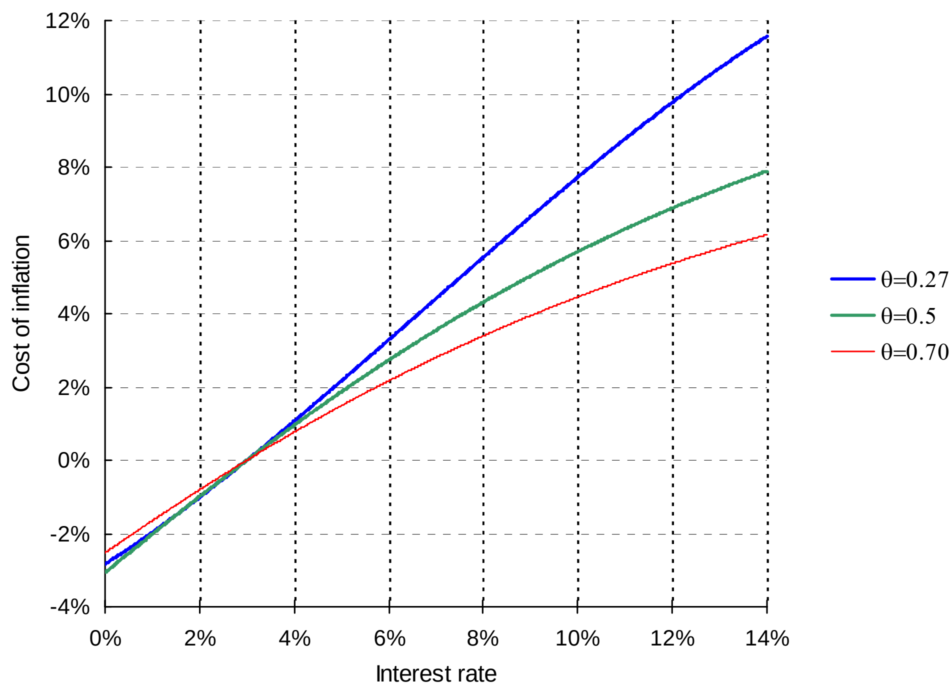

Figure 12: Cost of inflation and capital

6. Inflation and participation

Up to now, the frequency of trade σ has been assumed to be constant and independent of monetary policy: Inflation affects the quantities traded in bilateral meetings but not the number of meetings. If agents can choose whether or not to participate in the market, or on which side of the market to participate in, then the number of trades is endogenous and it can be affected by policy.[25] Also, in all variations of the model presented in the previous sections, the Friedman rule is optimal. In contrast, we will show that the Friedman rule is not necessarily optimal when participation decisions are endogenous.

We consider an extension of the model based on an assumption in Shi (1997). At the beginning of each period, before matches are formed, individuals can choose to be buyers or sellers in the decentralized market. An agent who chooses to be a buyer can only consume (he cannot produce) in the decentralized market, while an agent who chooses to be a seller can only produce (he cannot consume). The composition of buyers and sellers is then endogenous. Let n denote the fraction of sellers in the economy. Assume further that the matching process is such that a buyer meets a seller with probability $\sigma^b = n$, whereas a seller meets a buyer with probability $\sigma^s = 1 - n$.[26]

The value of an agent in the decentralized market satisfies

$$W\left(\frac{m_t}{p_t}\right) = \frac{m_t + T_t}{p_t} + \max_{\hat{m}_{t+1}}\left\{-\frac{\hat{m}_{t+1}}{p_t} + \beta \max\left[V^b\left(\frac{\hat{m}_{t+1}}{p_{t+1}}\right), V^s\left(\frac{\hat{m}_{t+1}}{p_{t+1}}\right)\right]\right\},$$

where V^b (V^s) is the value of a buyer (seller) in the decentralized market, and T_t is the lump-sum money transfer received by each agent. As before, the value function in the

[25] Monetary search models with participations decisions include Li (1995, 1997), Faig (2004), Rocheteau and Wright (2004, 2005), and Shi (1997, 1998). The model in this section is similar to the one in Rocheteau and Wright (2004), except that we consider different pricing mechanisms.

[26] The specification for the matching function is the same as the one used in most monetary models, including Kiyotaki and Wright (1993).

centralized market is linear in the agent's wealth. Each agent chooses his real balances in the centralized market and enters the subsequent decentralized market as a buyer or a seller. The value of being a buyer in the decentralized market satisfies

$$V^b\left(\frac{m_t}{p_t}\right) = n\left[u(q^b) - d^b\right] + W\left(\frac{m_t}{p_t}\right),$$

where we have used the linearity of $W(m_t/p_t)$. According to the previous Bellman equation, a buyer meets a seller in the decentralized market with probability *n*, the measure of sellers in the market. In this event, the buyer enjoys the surplus $u(q^b) - d^b$, and his continuation value in the centralized market is $W(m_t/p_t)$. Similarly, the value of a seller in the decentralized market satisfies

$$V^s\left(\frac{m_t}{p_t}\right) = (1-n)\left[u(q^s) - d^s\right] + W\left(\frac{m_t}{p_t}\right).$$

In equilibrium, agents are indifferent to being buyers or sellers, which requires

$$\max_{\hat{m}_{t+1}}\left\{-\frac{\hat{m}_{t+1}}{p_t} + \beta V^b\left(\frac{\hat{m}_{t+1}}{p_{t+1}}\right)\right\} = \max_{\hat{m}_{t+1}}\left\{-\frac{\hat{m}_{t+1}}{p_t} + \beta V^s\left(\frac{\hat{m}_{t+1}}{p_{t+1}}\right)\right\}.$$

It is easy to check that sellers choose $\hat{m}_{t+1} = 0$ since they do not need money in the decentralized market. Substituting V^b and V^s by their expressions and rearranging, we obtain

$$-rz(q) + n\left[u(q) - z(q)\right] = (1-n)\left[z(q) - c(q)\right].$$

The left-hand side of the previous equation is the expected trade surplus of a buyer net of the cost of carrying real balances; the right-hand side is the expected trade surplus of a seller, and *q* is the equilibrium quantity traded in bilateral matches. We assume that prices are determined according to the proportional bargaining solution, $z(q) = \theta c(q) + (1-\theta)u(q)$. The measure of sellers then satisfies

$$n = 1 - \theta + r\left[\frac{\theta c(q) + (1-\theta)u(q)}{u(q) - c(q)}\right].$$

For a given *q*, an increase in inflation tends to reduce the measure of buyers and increase the measure of sellers.

Society's welfare is measured by $n(1-n)[u(q)-c(q)]$. It is maximized when $n = 1/2$ and $q = q^*$. The equilibrium allocation maximizes welfare if and only if $\theta = 0.5$ and $r = 0$. The second requirement, which corresponds to the Friedman rule, guarantees that $q = q^*$. The first requirement, $\theta = 0.5$, is the condition under which $n = ½$ when $r=0$.[27]

The aggregate demand for money is defined as the money held by the $1-n$ buyers divided by the sum of the output in the decentralized market, $n(1-n)z$, and the output in the centralized market, A,

$$L = \frac{(1-n)z}{n(1-n)z + A}.$$

We use the same strategy as before to estimate the parameters A and η of money demand.

In Figure 13 and Table 5, we report the welfare cost of inflation. If buyers and sellers are symmetric in terms of their bargaining powers (θ=0.5), then the welfare cost of 10 percent inflation is slightly below 3 percent of GDP. In this case, participation decisions do not much affect the cost of inflation because the composition of the market is close to the efficient benchmark.

If $\theta < 0.5$, then the number of buyers is too low, and inflation lowers the fraction of buyers even further. In this case, the participation decision amplifies the welfare cost of inflation. For instance, if we choose θ to target an average markup of 10 percent, for the entire economy, we find $\theta=0.36$. In this case, agents are willing to give up about 5.5 percent of GDP to avoid a 10 percent inflation rate. Reciprocally, the welfare gains from reducing the interest rate to zero are also large.

Finally, if $\theta > 0.5$, then a deviation from the Friedman rule is optimal. If $r = 0$, then the number of buyers is too high and the number of trades is too low. Since inflation has a direct negative effect on buyers' expected utility, the number of buyers falls with r, while the number of sellers rises.[28] The composition of the market becomes more even and the number of trades expands. When we calibrate θ to match a 20 percent markup in the decentralized market, we find θ=0.90. The optimal inflation is about 2 percent,

---

[27] For an elaboration of this idea, see Berentsen, Rocheteau, and Shi (2004).

[28] This effect is sensitive to the choice of the pricing mechanism. Under the Nash solution, the number of buyers can increase with inflation because the buyer's share in the surplus of a match increases.

whereas the cost of 10 percent inflation is slightly less than a fifth of a percent of GDP. So the welfare cost of inflation is much lower than what was found in the absence of participation decisions.

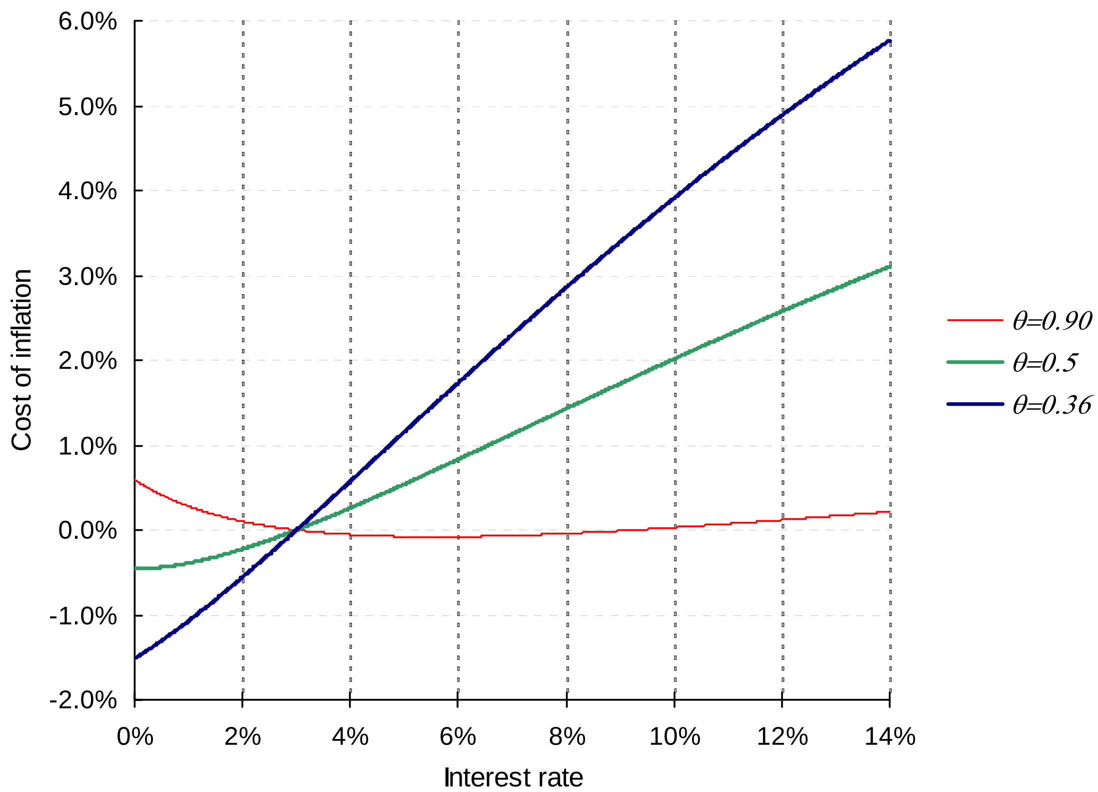

Figure 13. Endogenous participation (proportional bargaining)

Pricing mechanisms	*Market power*	*Coefficients of L(r)*		*Allocation*						*Cost of inflation*
				r=0		*r=3%*		*r=13%*		
	θ	A	η	*q*	*n*	*q*	*n*	*q*	*n*	*Δ(%)*
Proportional bargaining	0.90	2.03	0.52	1.00	0.01	0.61	0.12	0.24	0.15	0.18
	0.50	1.05	0.29	1.00	0.50	0.70	0.57	0.26	0.69	2.92
	0.36	0.80	0.35	1.00	0.64	0.71	0.70	0.24	0.81	5.47
Nash bargaining	0.89	1.63	0.55	0.87	0.11	0.48	0.09	0.04	0.03	4.72
	0.50	0.89	0.27	0.80	0.48	0.51	0.51	0.15	0.53	3.25
	0.32	0.52	0.29	0.70	0.66	0.41	0.68	0.10	0.67	5.45

Table 5. Inflation and participation

Not surprisingly, the choice of the pricing mechanism matters considerably for the welfare cost of inflation and the result according to which the Friedman rule may be suboptimal when participation decisions are endogenous. For instance, under the Nash solution, the Friedman rule is optimal for all our calibrated examples. The reason for this result is intuitive. Since the Friedman rule fails to generate the first best q under Nash bargaining, the envelope argument does not apply, and a deviation from the Friedman rule has a first-order effect on society's welfare. For values of θ that match the markup ($\theta=0.89$), the welfare cost of 10 percent inflation is about 4.7 percent of GDP, much bigger than was found previously.

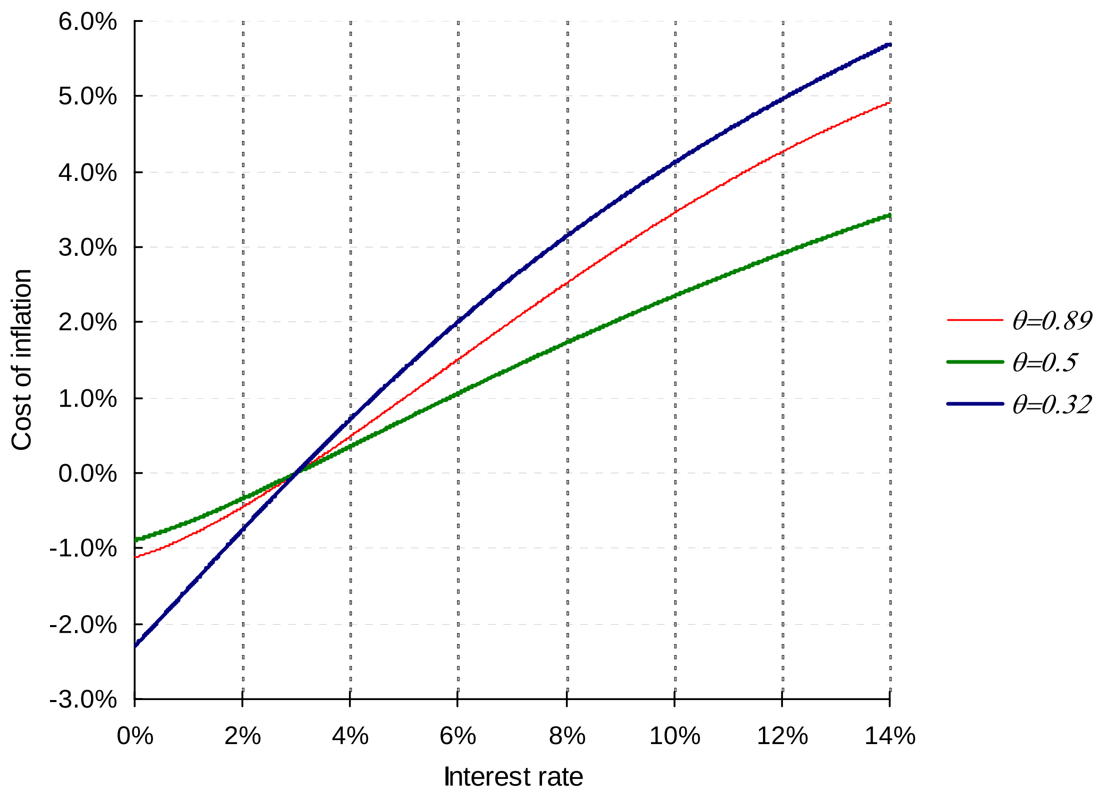

Figure 14: Endogenous participation (Nash bargaining)

To summarize, the introduction of endogenous participation decisions has several implications. First, the welfare triangle is a misleading measure because it does not capture the distortionary effects of inflation on individuals' participation decisions. Second, the presence of search externalities can mitigate or exacerbate the welfare cost of inflation. For values of the bargaining power that generate a reasonable markup, the welfare cost of inflation is much smaller than what was found in the previous sections. Third, the Friedman rule may no longer be optimal, because inflation's positive effect on the composition of the market and the frequency of trades can counteract the negative effect of inflation on real balances.

7. Conclusion

Using different extensions of a simple search-theoretic model of monetary exchange, we have identified and quantified various effects of inflation on welfare. First, inflation introduces a wedge in the decision to invest in real balances. The extent of this distortion depends on the assumed pricing mechanism. If buyers receive the full marginal benefit of their money balances, the cost of inflation is essentially the one given by the Bailey-Lucas measure, that is, the area underneath the money demand function. If buyers do not receive the full margin return of their real balances, then the Bailey-Lucas measure has to be scaled up by a factor that is an increasing function of sellers' market power. We have also provided examples of bargaining solutions under which the Friedman rule fails to generate the first-best allocation. For such mechanisms, the social benefit of implementing the optimal deflation can be large, since a deviation from the Friedman rule has a first-order effect on welfare.

Second, if capital is an input of the production technology in the decentralized market, then inflation affects agents' incentives to accumulate capital. The private marginal return of capital depends on the quantity of real balances held by buyers, which is a decreasing function of the inflation rate. Since there are two investment decisions, the choice of real balances and the choice of capital stock, bargaining introduces a double holdup problem. On the one hand, real balances increase with buyers' bargaining power. On the other, sellers' capital decreases with buyers' bargaining power. Because of this double holdup problem, the equilibrium allocation is generically inefficient. Furthermore, the welfare cost of inflation for plausible values of the markup is larger than what was found in the model without capital.

Third, inflation also affects agents' decisions to participate in the market and, consequently, the number of trades. Since participation decisions generate search externalities, the Friedman rule may no longer be optimal. For reasonable values of the markup, we provide examples in which the optimal inflation rate is positive and the welfare cost of 10 percent inflation is small. This result, however, is sensitive to the choice of the pricing mechanism.

The simple search model we have considered abstracts from the distributional effects of money creation. In Craig and Rocheteau (2006), we capture the distributional effects of inflation through a simple extension of the model in which individuals live two periods and generations overlap. We assume that individuals are heterogeneous in their abilities to produce when young, and we establish that a low inflation rate can raise welfare.

Additional extensions are worth considering. For instance, one should take into account distortionary taxes and other assets beside money, such as government bonds and credit. The calibration strategy should also be refined to provide a better sense of the extent of search externalities. These extensions are left for future investigation.

Literature.

1. Aruoba, Boragan, Christopher Waller, and Randall Wright (2006). "Money and capital," Working Paper.
2. Bailey, Martin (1956). "The welfare cost of inflationary finance," *Journal of Political Economy* 64, 93-110.
3. Berentsen, Aleksander, Guillaume Rocheteau, and Shouyong Shi (*2004). "Friedman meets Hosios: Efficiency in search models of money," Economic* Journal (forthcoming).
4. Brockman, Michael, Theo Glasser, and Eva Herrmann (1993). "Locally adaptive bandwidth choice for kernel estimators," *Journal of the American Statistical Association* 88, 1302-09.
5. Craig and Rocheteau (2006). "Inflation and welfare: A search approach," Policy Discussion Paper of the Federal Reserve Bank of Cleveland 12 (January).
6. Ennis, Huberto (2004). "Search, money, and inflation under private information," *Federal Reserve Bank of Richmond Discussion Paper* 142.
7. Faig, Miquel (2004). "Divisible money in an economy with villages," mimeo.
8. Faig, Miquel and Belen Jerez (2005). "Precautionary balances and the velocity of circulation of money," *Journal of Money, Credit, and Banking* (forthcoming).
9. Fischer, Stanley (1981). "Towards an understanding of the costs of inflation: II," *Carnegie-Rochester Conference Series on Public Policy* 15, 5-42.
10. Friedman, Milton (1969). "The optimum quantity of money," in *The Optimum Quantity of Money and Other Essays,* Aldine publishing company, Chicago.
11. Kiyotaki, Nobuhiro, and Randall Wright (1989). "On money as a medium of exchange," *Journal of Political Economy* 97, 927-54.
12. Kiyotaki, Nobuhiro, and Randall Wright (1993). "A search-theoretic approach to monetary economics," *American Economic Review* 83, 63-77.
13. Lagos, Ricardo, and Randall Wright (2005). "A unified framework for monetary theory and policy analysis," *Journal of Political Economy* 113, 463-84.

14. Lagos, Ricardo, and Guillaume Rocheteau (2004). “Money and capital as competing media of exchange,” mimeo.

15. Li, Victor (1995). “The optimal taxation of fiat money in search equilibrium,” *International Economic Review* 36, 927-42.

16. Li, Victor (1997). “The efficiency of monetary exchange in search equilibrium,” *Journal of Money, Credit, and Banking* 29, 61-72.

17. Lucas, Robert (1981). “Discussion of: Stanley Fischer, Towards an understanding of the costs of inflation: II,” *Carnegie-Rochester Conference Series on Public Policy* 15, 43-52.

18. Lucas, Robert (2000). “Inflation and welfare,” *Econometrica* 68, 247-74.

19. Marty, Alvin (1997). “The welfare cost of inflation: A critique of Lucas and Bailey”, *Federal Reserve Bank of St. Louis Review* 81, 41-46.

20. Molico, Miguel (1999). “The distribution of money and prices in search equilibrium,” *International Economic Review* (forthcoming).

21. Reed, Robert, and Christopher Waller (2004). “Money and risk sharing,” *Journal of Money, Credit, and Banking* (forthcoming).

22. Rocheteau, Guillaume, and Christopher Waller (2005). “Bargaining and the value of money,” Federal Reserve Bank of Cleveland Working Paper 05-01.

23. Rocheteau, Guillaume, and Randall Wright (2004). "Inflation and welfare in models with trading frictions,” in *Monetary Policy in Low Inflation Economies*, David Altig and Ed Nosal, eds., Cambridge University Press (forthcoming).

24. Rocheteau, Guillaume, and Randall Wright (2005). “Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium,” *Econometrica* 73, 175-202.

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

26. Shi, Shouyong (1998). “Search for a propagation mechanism,” *Journal of Economic Theory* 81, 314-52.

27. Shi, Shouyong (1999). “Search, inflation and capital accumulation,” *Journal of Monetary Economics* 44, 81-103.

28. Wang, Weimin and Shouyong Shi (2005). "The variability of velocity of money in a search model," *Journal of Monetary Economics* (forthcoming).

Appendix: Inflation and capital

The capital choice of an agent solves

$$\max_k \left\{-k + \beta\left[\sigma S^s(k,z) + (1-\delta)k\right]\right\},$$

where $S^s(k,z)$ is the seller's surplus from a trade as a function of his investment k and the buyer's real balances z. The agent incurs the cost k in the centralized market to buy k units of capital. In the subsequent period, the agent can use his capital and his hours if he has an opportunity to produce with probability σ. Each unit of capital depreciates at rate δ between the decentralized market and the centralized market. The problem can be rewritten as

$$\max_k \left[-(\rho+\delta)k + \sigma S^s(k,z)\right].$$

If the terms of trade are determined according to the proportional bargaining solution, the seller's surplus satisfies

$$S^s(k,z) = \max_h (1-\theta)\left[u\left[q(h,k)\right] - c(h)\right] + \lambda^s\left[z - (1-\theta)u\left[q(h,k)\right] - \theta c(h)\right],$$

where λ^s is the Lagrange multiplier associated with the cash constraint $d \le z$.

The first-order condition for the choice of capital is $\rho + \delta = \sigma S_k^s(k,z)$, where

$$S_k^s(k,z) = (1-\theta)u'(q)q_k(h,k)\left(1-\lambda^s\right),$$

with $\lambda^s = 0$ if the constraint $d \le z$ is not binding and where q_k is the partial derivative of the production function. The first-order condition with respect to h gives

$$\lambda^s = \frac{(1-\theta)\left[u'(q)q_h(h,k) - c'(h)\right]}{(1-\theta)u'(q)q_h(h,k) + \theta c'(h)}.$$

The choice of real balances solves

$$\max_z \left[-rz + \sigma S^b(k,z)\right],$$

where the buyer's surplus from a trade satisfies

$$\begin{aligned} S^b(k,z) &= \max_{h,d} \theta\left[u\left[q(h,k)\right] - d\right] \\ \text{s.t.} \quad & (1-\theta)\left[u[q(h,k)] - d\right] = \theta\left[d - c(h)\right] \\ & d \le z. \end{aligned}$$

This problem can be rewritten as

$$S^b(k,z) = \max_h \theta\left[u\left[q(h,k)\right] - c(h)\right] + \lambda^{\mathrm{b}}\left[z - (1-\theta)u\left[q(h,k)\right] - \theta c(h)\right].$$

The first-order condition for the choice of real balances yields $\lambda^b = r/\sigma$. Furthermore, the proportional bargaining solution implies $\lambda^b = \frac{\theta}{1-\theta}\lambda^s$: The benefit of an additional unit of money for the buyer is $\frac{\theta}{1-\theta}$ times the benefit for the seller. From this last observation we deduce that the choice of real balances obeys

$$\frac{r}{\sigma} = \frac{\theta\left[u'(q)q_h(h,k) - c'(h)\right]}{(1-\theta)u'(q)q_h(h,k) + \theta c'(h)}.$$

Using the fact that $\lambda^s = \frac{1-\theta}{\theta}\frac{r}{\sigma}$, the choice of capital can be simplified to

$$\rho + \delta = \sigma(1-\theta)u'(q)q_k(h,k)\left(1 - \frac{1-\theta}{\theta}\frac{r}{\sigma}\right).$$