The Cost of Inflation: A Mechanism Design Approach*

Guillaume Rocheteau
University of California, Irvine
Federal Reserve Bank of Cleveland

December 28, 2011

Abstract

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 an optimal, incentive feasible 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 the introduction of match-specific heterogeneity and endogenous participation decisions.

JEL Classification: D82, D83, E40, E50

Keywords: Cost of inflation, pairwise trades, optimal mechanism

---

*I thank three anonymous referees, an editor and associate editor, Nicolas Jacquet, Ryan Baranowski, Daniel Sanches, Serene Tan, and seminar participants at the Bank of Canada and the University of California at Irvine for useful comments and discussions. I am also grateful to Peter Ireland, who shared his data on U.S. money demand. I thank Monica Crabtree-Reusser for editorial assistance. E-mail: grochete@uci.edu.
1 Introduction

A classical topic in monetary economics is measuring the burden that inflation imposes on society. The standard methodology, reviewed in [20], consists of estimating a reduced-form money demand function and measuring the welfare cost of inflation as the area underneath money demand. The rationale for this approach is based on competitive general equilibrium models where money enters the utility function, or a cash-in-advance constraint. Unfortunately, as pointed out by [25], 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, [19] – 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, and can be up to 5 percent of GDP.

The quantitative insights of LW, however, are subject to the caveat of [15], 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 ([25] and [26]) 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 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 [20], 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.

There is a growing literature, surveyed in [9], measuring the welfare cost of inflation in the context of microfounded models of monetary exchange under different trading mechanisms. In contrast to this literature I endogenize the trading mechanism in decentralized markets so that it implements an optimal, incentive-feasible allocation for all inflation rates. Relative to [15], I characterize money demand (pointing out a mischaracterization of the true set of incentive-feasible allocations) and I calibrate the model to quantify the cost of inflation following the methodology in [20]. While they focus on output levels, I characterize incentive-feasible allocations in terms of output and real balances and I show that the model can be made consistent with the data. I also check the robustness of the results to different extensions. I introduce match-specific heterogeneity (idiosyncratic preference shocks) and private information, as in [13], and a participation decision that endogenizes the frequency at which agents trade, as in [22] and [24].

2 The environment

The environment is similar to the one in LW. 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 5.3 \( n \) is exogenous, while in Section 5.3 the composition of the market is endogenous. Buyers and sellers are matched bilaterally: a buyer meets a seller with probability \( \alpha_b \), while a seller meets a buyer with probability \( \alpha_s \), 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 utility function, \( E \sum_{t=0}^{\infty} \beta^t U(q_t, e_t, c_t, h_t) \), 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

---

1This literature includes [2], [3], [4], [8], and [22].

2[16] and [17] were the first to use implementation theory to prove the essentiality of money. Applications of mechanism design to monetary theory include [5], [6], [7], [11], [12], [18], [21].
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 $q = e$. Similarly, output in the CM is produced according to a linear production function in labor, which implies the (real) wage rate is equal to 1. The utility function is well-behaved, and $q^* = \arg \max [u(q) - \psi(q)] > 0$. I also assume without loss in generality that $u(0) = \psi(0) = 0$.

All goods are perishable across both stages and time. Agents cannot commit to future actions, and individual histories are private information. These assumptions 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, $\gamma = 1 + \pi$, is constant and new money is injected by lump-sum transfers (or taxes if $\gamma < 1$) in the CM.\(^3\) 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 private information.

### 3 Optimal, incentive-feasible 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. A mechanism in the DM, $[q, d] : \mathbb{R}_{2+} \rightarrow \mathbb{R}_+ \times \mathbb{R}_+$, maps the announced real money holdings of the buyer and the seller in a proposed allocation, $(q, d) \in \mathbb{R}_+ \times [z^s, z^b]$, 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.\(^4\)

\(^3\)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. [1] and [15] study the case where agents can choose whether or not to participate in the CM and pay taxes.

\(^4\) See [27] for a coalition-proof game that guarantees that any trade in the DM is in the pairwise core.
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.

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

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

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 \(\sigma\). 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 \(\sigma\), 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

\[
W(z) = \max_{c, \tilde{z}} \{U(c) - c + T + z - \gamma \tilde{z} + \beta V(\tilde{z})\},
\]

where \(T\) is the lump-sum transfer (expressed in numéraire goods), and \(\tilde{z}\) is the real balances taken into the next DM. I have used the budget constraint according to which the CM supply of hours is \(h = c + \gamma \tilde{z} - z - T\) and the relative price of real balances next period in terms of current-period CM output is \(p_{t+1}/p_t = \gamma\). Substituting \(V(\tilde{z})\) 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

\[
\max_{z \geq 0} \{-iz + \sigma \{u[q(z, z^p)] - d(z, z^p)\} + \sigma \{d(z^p, z) - \psi[q(z^p, z)]\}\},
\]

where \(i \equiv (1 + \pi)(1 + r) - 1\), which can be interpreted as the nominal interest rate that would be paid on an illiquid bond.
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, \(z^s \in \arg \max_z \{d(z^p, z) - \psi[q(z^p, z)]\} \mathbb{I}_{\{d(z, z^p) \leq z^p\}} = d^p - \psi(q^p)\), where \(\mathbb{I}_{\{d \leq z^p\}}\) is an indicator function that is equal to one if \(d \leq z^p\). A necessary condition for the mechanism to be incentive-compatible is that the seller’s expected surplus is independent of his announced real balances.\(^5\) Consequently, for any incentive-compatible mechanism, the choice of real balances, (3), can be reexpressed as

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

The optimal choice of real balances maximizes the expected surplus of a buyer in the DM net of the cost of holding real balances. If a buyer holding \(z^b\) announces \(\hat{z}^b\) and it turns out that \(d(\hat{z}^b, z^s) > z^b\) then the trade is not feasible and cannot be carried out. Therefore, the optimal announcement of a buyer who holds \(z^b\) is \(\hat{z}^b \in \arg \max_z \{u[q(z, z^p)] - d(z, z^p)\} \mathbb{I}_{\{d(z, z^p) \leq z^b\}}\).

Given that money holdings are private information, 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

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

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,

\[
-\psi(q^p) + d^p \geq 0. \tag{6}
\]

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

Let \(C(z^b, z^s)\) denote the pairwise core, i.e., the set of all feasible allocations, \((q, d) \in \mathbb{R}_+ \times [-z^s, z^b]\), 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.

\(^5\)The seller’s money holdings will not affect the feasibility of the transfer in a match because this transfer needs to be non-negative as a result of the seller’s participation constraint.
Lemma 1  Any allocation \((q^p, d^p, z^p)\) such that \(z^p = d^p\) and \((q^p, d^p) \in C(z^p, z^p)\) that satisfies (5) and (6) can be implemented by the following coalition-proof trading mechanism:

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

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

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.

The proof of Lemma 1 is contained in Figure 1. The top panel of Figure 1 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 1 it is easy to check that the agent will choose \(z = d^p\) if (5) holds.

An optimal, incentive-feasible allocation is an allocation implementable by the trading mechanism in Lemma 1 that maximizes society’s welfare, denoted \(W\).

**Definition 1** An optimal, incentive-feasible allocation is

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

\[
s.t. \quad -\psi(q) + d \geq 0 \quad (10)
\]

\[
-id + \sigma[u(q) - d] \geq 0. \quad (11)
\]

It is straightforward to establish that the solution to (9)-(11) is in the pairwise core, and it corresponds to the highest \(q \leq q^*\) so that (10) and (11) hold. The solution is

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

\[
\in \begin{cases} 
\{q^*\} \times \left[ \psi(q^*), \frac{\sigma}{i + \sigma}u(q^*) \right] & \text{if} \quad \psi(q^*) \leq \frac{\sigma}{i + \sigma}u(q^*) \\
\{q(i)\} \times \{\psi[q(i)]\} & \text{otherwise},
\end{cases} \quad (13)
\]

6
Figure 1: An incentive-feasible mechanism

where $q(i)$ is the positive solution to $\psi(q) = \frac{\sigma}{\bar{t} + \sigma} u(q)$. 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.

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

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

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 in the case where money holdings are private information.

**Proposition 1** There is

\[ \bar{i} = \frac{\sigma [u(q^*) - \psi(q^*)]}{\psi(q^*)} > 0 \] (14)

such that

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

2. For all \( i > \bar{i} \), \( \frac{\partial q^p}{\partial i} < 0 \), \( \frac{\partial D^p}{\partial i} < 0 \), and \( \frac{\partial W}{\partial i} < 0 \).

The quantity \( \bar{i} \) is the highest nominal interest rate below which the first-best level of output is incentive-feasible.\(^6\) 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. If the nonpecuniary rate of return on money is larger, then there is a larger range of inflation rates that are consistent with the first-best allocation. The first part of Proposition 1 also shows that 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. When the buyer’s participation constraint binds, \( i > \bar{i} \), 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.\(^7\)

### 4 The irreducible cost of inflation

Following the methodology of [20] and LW, I construct the aggregate money demand and check whether there are parameter values for which it fits the data. The model can then be used to

---

\(^6\) The results in Proposition 1 hinge on some specific assumptions of the Lagos-Wright model. The assumption of pairwise meetings is crucial for the first part of the proposition according to which the first best is implementable for a range of inflation rates. Because the core in small group meetings is non-degenerate, agents can be punished if they do not bring enough real balances. Such punishment would not be available in a model with a degenerate core. The assumption of quasi-linear preferences makes the distribution of money holdings degenerate, which eliminates the incentive-compatibility constraints that would emerge with heterogeneous money holdings.

\(^7\) Proposition 1 provides a testable implication of the model. Inflation does not affect output for low inflation rates but it does reduce output for large inflation rates. While the evidence on the superneutrality of money is mixed, there are studies showing nonlinear effects of inflation on output.
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.\footnote{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),

$$L(i) = \left\{ \begin{array}{ll}
\frac{\psi(q^*)}{A + \sigma \psi(q^*)}, & \text{if } i \leq \bar{i} \equiv \frac{\sigma [u(q^*) - \psi(q^*)]}{\psi(q^*)} \\
\frac{\psi[q(i)]}{A + \sigma \psi[q(i)]}, & \text{if } i > \bar{i} \equiv \frac{\sigma [u(q^*) - \psi(q^*)]}{\psi(q^*)}.
\end{array} \right.$$

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, $(A, a, b, \sigma)$, 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 [15], for this parametrization the first-best allocation is implementable for all the interest rates observed in the data. However, as revealed by Figure 2, 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.

The next step is to recalibrate the model. Figures 3 and 4 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. In Figure 3, 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
value of CM output, $A$, to adjust the level of money demand. In Figure 4 I set $b = 0$ but choose $(\sigma, A, a)$ 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.

To understand how the model can match the empirical money demand, consider the following two moments: $L = \min(L_t)$ is the minimum real balance and $\bar{L} = \max(L_t)$ 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 (15) as

$$L(i) = \left[ \frac{1}{\psi(q^\tau)} + \sigma, \frac{1}{\psi(q^\tau)} \left( 1 + \frac{i}{\sigma} \right) \frac{A}{\psi(q^\tau)} + \sigma \right].$$
To match these two targets it is sufficient to choose parameter values such that $\frac{1}{\psi(q^{*})} \leq \bar{L}$ and $\frac{1}{(1+i_{\text{max}})^{\frac{1}{\psi(q^{*})}} + \sigma} \geq \bar{L}$, where $i_{\text{max}}$ is the highest interest rate in the data. For any given $\sigma$, one can always find $\frac{A}{\psi(q^{*})}$ 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 $\frac{A}{\psi(q^{*})}$, one can make $\frac{A}{u(q^{*})} = \frac{A}{\psi(q^{*})} \cdot \frac{\psi(q^{*})}{u(q^{*})}$ and $\sigma$ sufficiently low 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.

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 [20]. 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 [9]) 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.

5 Extensions

In this section I check the robustness of the results to three different extensions. First, I show that if money holdings are partially-proveable or fully observable, as assumed in [15], the set of incentive-feasible allocations expands, which makes it even easier to implement good allocations. Second, I introduce idiosyncratic preference shocks that are private information and show that, in contrast to [13] under price posting, low inflation imposes no cost on society. Third, I introduce participation decisions and search externalities and find that inflation is neither costly nor beneficial, in contrast to [24] under bargaining and [22] under different trading mechanisms.

5.1 Private vs partially-proveable money holdings

Throughout the paper I assume that money holdings are private information. In contrast, [15] assumes that while agents can hide their money balances, they cannot overstate them. In order to discuss the importance played by this assumption, I derive the set of stationary, symmetric, incentive-feasible allocations in the case where money holdings cannot be overstated.

A necessary condition for agents to be willing to accumulate $z^p$ real balances is $W(0) \geq \beta W(0)$, i.e.,

$$-iz^p + \sigma [u(q^p) - \psi(q^p)] \geq 0. \quad (17)$$

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. 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. It should be emphasized from (17) that the money holdings of an agent, $z^p$, need not coincide with the transfer of money in the DM, $d^p$.

9For readers familiar with the literature, the analysis of the set of implementable allocations in [15] is erroneous as they impose (5) as a necessary condition for an implementable allocation instead of (17), which is the relevant condition when money holdings cannot be overstated.
Lemma 2 Any allocation \((q^p, d^p, z^p)\) such that \((q^p, d^p) \in C(z^p, z^p)\) that satisfies \(d^p \leq z^p\), (17) and (6) can be implemented by the following coalition-proof trading mechanism:

\[
(q, d) = \arg \max_{q, d \leq z^b} [d - \psi(q)] \quad \text{s.t.} \quad u(q) - d \geq u(q^p) - d^p \quad \text{if} \quad \min(z^b, z^s) \geq z^p \quad (18)
\]

\[
(q, d) = \arg \max_{q, d \leq z^b} [d - \psi(q)] \quad \text{s.t.} \quad u(q) - d = 0 \quad \text{if} \quad z^b < z^p \quad (19)
\]

\[
(q, d) = \arg \max_{q, d \leq z^b} [u(q) - d] \quad \text{s.t.} \quad -\psi(q) + d = 0 \quad \text{if} \quad z^s < z^p \text{ and } z^b \geq z^p. \quad (20)
\]

The novelty is (20) according to which if the seller holds less than \(z^b\) and the buyer holds at least \(z^p\), then the mechanism proposes the preferred trade of the buyer in the pairwise core.

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

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

If money holdings are private information,

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

It is easy to check that for all \(i \geq 0\), \(\mathcal{A}^u(i) \subset \mathcal{A}^o(i)\). Intuitively, when money holdings cannot be overstated, there is more leverage to punish an agent who does not carry enough real balances. Therefore, our calibration of the model and our measures of the cost of inflation are robust if money holdings cannot be overstated or are observable.

5.2 Match-specific heterogeneity

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 \(\varepsilon u(q) - \psi(e) + U(e) - h\), where \(\varepsilon \in \mathcal{E} \subset [0, \bar{\varepsilon}]\) is a match-specific component drawn from a distribution \(G(\varepsilon)\). The preference shock, \(\varepsilon\), is private information to the buyer.

An allocation rule in the DM maps a triple, \((\varepsilon, z^b, z^s)\), 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. 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.

A necessary condition for a buyer to have incentives to reveal truthfully his preference shock is

\[ \varepsilon u(q^p_\varepsilon) - d^p_\varepsilon \geq \varepsilon' u(q^p_{\varepsilon'}) - d^p_{\varepsilon'} \quad \text{for all } \varepsilon' \neq \varepsilon. \]  \hspace{1cm} (21)

According to (21), the buyer will achieve a higher surplus by reporting his true preference shock, \( \varepsilon \), instead of some other value, \( \varepsilon' \). Individual rationality in a match requires

\[ \psi(q^p_{\varepsilon}) \leq d^p_{\varepsilon} \leq \varepsilon u(q^p_{\varepsilon}) \quad \text{for all } \varepsilon \in [0, 1]. \]  \hspace{1cm} (22)

Finally, necessary conditions for an agent in the CM to accumulate \( z \) units of real balances are:

\[ -iz^p + \sigma \int_0^1 [\varepsilon u(q^p_\varepsilon) - d^p_\varepsilon] dG(\varepsilon) \geq -iz' + \sigma \int_0^1 \max [\varepsilon u(q^p_{\varepsilon'}) - d^p_{\varepsilon'}] I_{(d^p_{\varepsilon'} \leq z')} dG(\varepsilon), \]  \hspace{1cm} (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 feasible offer such that \( d^p_{\varepsilon'} < z' \) must not be profitable.

It is easy to see that any allocation \( \{(q^p_\varepsilon, d^p_\varepsilon) : \varepsilon \in \mathcal{E}, z^p = d^p_\varepsilon \} \) that satisfies \( z^p = d^p_\varepsilon \), (21), (22), and (23) is implemented by the following mechanism:

\[
\begin{align*}
q(z^b, z^s, \varepsilon), d(z^b, z^s, \varepsilon) & = (q^p_{\varepsilon}, d^p_{\varepsilon}) \quad \text{if } z^b \geq d^p_{\varepsilon}, \\
& = (0, 0) \quad \text{otherwise}.
\end{align*}
\]  \hspace{1cm} (24) (25)

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 \( \pi_h \), while the probability of the low preference shock is \( \pi_\ell = 1 - \pi_h \). The next Lemma establishes the conditions under which the first-best allocation is implementable. The first best requires \( q_{\varepsilon_\ell} = q^*_h \) and \( q_{\varepsilon_h} = q^*_\ell \), 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 \( \bar{\varepsilon} \in (0, \infty) \) such that the first-best allocation is implementable for all \( i \in [0, \bar{\varepsilon}] \).

There is a range of inflation rates, including the Friedman rule, that implement the first-best allocation.\(^{10}\) So, as in the case with homogenous matches, the first-best allocation can be obtained

\(^{10}\)In the working paper, [23], I consider the case of a continuous idiosyncratic preference shock. I show that unless \( i = 0 \) the first-best is not implementable and the cost of inflation is bounded above by the cost obtained under buyers-take-all bargaining. However, if money holdings cannot be overstated, then there is an interval of inflation rates above the Friedman rule that are consistent with the implementation of the first best.
even with inflation rates above the Friedman rule. (From the proof of Proposition 2, it can also be checked that for sufficiently large inflation rates, the first best is not implementable).

Finally, there are calibrated examples that fit money demand and that implement the first-best allocation for all inflation rates in the data. Under the parameter values \( \varepsilon_h = 1, \varepsilon_t = 0.5, \pi_h = 0.5, \sigma = 0.5, \) and \( a = 0.9 \) the welfare cost of (moderate) inflation is 0, in contrast to the large welfare costs of inflation found by [13] under price posting.

### 5.3 Endogenous participation

So far I have considered economies where the measure of trades in the DM is exogenous. A key insight from [24] is that endogenous participation decisions mitigate or amplify the cost of inflation due to search externalities.\(^{11}\) For some parametrizations, the cost of small inflation is negative and the Friedman rule is suboptimal. In the following, we check the robustness of this insight to the adoption of an optimal trading mechanism.

In order to endogenize participation, I follow the approach of [24].\(^{12}\) I assume that there is a unit measure of ex ante identical agents who choose to be either buyers or sellers in the DM. 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 \( W = na \left( \frac{1-n}{n} \right) [u(q) - \psi(q)] + U(c) - c. \) Let \( n^* \) denote the composition of the market that maximizes the number of trades, \( n^* = \arg\max na \left( \frac{1-n}{n} \right). \) The first-best allocation is such that \( q = q^*, n = n^*, \) and \( c = c^*. \)

A symmetric, stationary allocation is represented by the 5-tuple \((q^p, d^p, n^p, z^p_b, z^p_s)\), where \( z^p_b \) denotes the buyer's real balances and \( z^p_s \) the seller's real balances. Under any incentive-compatible trading mechanism, sellers do not carry real balances in the DM, \( z^p_s = 0, \) and buyers hold no more money than they intend to spend, \( d^p = z^p_b. \) Moreover, for all \( n^p \in (0,1), \) an agent must be indifferent between being a buyer or a seller in the DM,

\[
-iz^p_b + \alpha(\theta) \left[ u(q^p) - d^p \right] = \frac{\alpha(\theta)}{\theta} \left[ d^p - \psi(q^p) \right] \geq 0, \tag{26}
\]

\(^{11}\)See [22] for a similar argument in the context of the LW model

\(^{12}\)There are many ways one can endogenize participation decisions. I adopt an approach that does not add parameters to calibrate.
where $\theta = \frac{1-n^p}{n^p}$. The left side of (26) is the expected surplus of a buyer, net of the cost of holding real balances. The right side of (26) is the expected surplus of a seller.

An optimal, incentive-feasible allocation is given by

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

subject to $-\psi(q) + d \geq 0$.

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

**Proposition 3** There is $\tilde{i} = \frac{\alpha(1-n^p/(q^*))[u(q^*)-\psi(q^*)]}{\psi(q^*)}$ such that

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

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

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

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

Under an optimal mechanism the first-best allocation can be implemented for low inflation rates. According to (30), 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).\(^{13}\) If the cost of holding money is larger than the threshold $\tilde{i}$, 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.

In the working paper, [23], I calibrate the model and I show that it fits the data well for the whole sample, 1900-2006, as well as the subsample 1981-2006.\(^{14}\) For both periods, a 10 percent inflation imposes no cost on society.

\(^{13}\)This result provides another testable implication according to which markups should be decreasing with inflation (at least for low inflation rates).

\(^{14}\)I adopt the same functional forms as before, $\psi(e) = e$, $u(q) = \frac{(q+b)^{1-a}-q^{1-a}}{1-a}$, but I set $b = 0$ and $a = 0.95$. The matching function is $\alpha(\theta) = \frac{\alpha}{1+a}$. I look for the pair $(\sigma, A)$ that minimizes squared residuals. For the whole sample a good fit requires a low frequency of trade, $\sigma = 0.06$. For the subsample 1981-2006 a good fit is obtained for a much larger frequency of trades, $\sigma = 0.47$. 

16
6 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.
References


[21] F. Mattesini, C. Monnet, R. Wright, Banking: A mechanism design approach, Rev. Econ. Stud. (Forthcoming).


APPENDIX

Proof of Proposition 1.

Part 1. From (12), the threshold $\bar{i}$ 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, $\psi(q^*) = \frac{\sigma}{i+\sigma} u(q^*)$, which gives (14). From (9) $W = \sigma [u(q^*) - \psi(q^*)] + U(c^*) - c^*$ is independent of $i$. Hence, $\frac{\partial W}{\partial i} = 0$ for all $i < \bar{i}$. Finally, $D^p(i) = \left[ \psi(q^*), \frac{\sigma}{i+\sigma} u(q^*) \right] \subset D^p(i') = \left[ \psi(q^*), \frac{\sigma}{i+\sigma} u(q^*) \right]$ for all $i' < i < \bar{i}$.

Part 2. From (13), for all $i > \bar{i}$, $q(i)$ 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)

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

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

Proof of Proposition 2. Denote $A^*(i)$ the set of pairs $(z, d_{\varepsilon_{\ell}}) \in \mathbb{R}_{2+}$ such that $z = d_{\varepsilon_{h}}$, and (21), (22), and (23) hold with $q_{\varepsilon_{h}} = q_{h}^*$ and $q_{\varepsilon_{\ell}} = q_{\ell}^*$.

Consider first the case $i = 0$. The constraints (21) and (22), which imply (23), can be reexpressed as:

$$\varepsilon_{\ell} \left[ u(q_{h}^*) - u(q_{\ell}^*) \right] \leq d_{\varepsilon_{h}} - d_{\varepsilon_{\ell}} \leq \varepsilon_{h} \left[ u(q_{\varepsilon_{h}}^*) - u(q_{\varepsilon_{\ell}}^*) \right]$$

$$\psi(q_{\varepsilon_{h}}) \leq d_{\varepsilon_{h}} \leq \varepsilon u(q_{\varepsilon_{h}}) \text{ for all } \varepsilon \in \{\varepsilon_{\ell}, \varepsilon_{h}\}.$$

The set $A^*(0)$ is illustrated in the figure below.
The measure of the set $A^*(0)$ is
\[
\mu [A^*(0)] = \int_{\tilde{z}^p} x + \varepsilon_h [u(q^*_h) - u(q^*_f)] - \psi(q^*_h) dx + \int_{\tilde{z}^p} (\varepsilon_h - \varepsilon_\ell) [u(q^*_h) - u(q^*_f)] dx,
\]
where $\tilde{z}^p = \min \{\psi(q^*_h) - \varepsilon_\ell [u(q^*_h) - u(q^*_f)], \varepsilon_\ell u(q^*_f)\} > \psi(q^*_f)$. Moreover, $x + \varepsilon_h [u(q^*_h) - u(q^*_f)] - \psi(q^*_h) > 0$ for all $x > \psi(q^*_f)$. So $\mu [A^*(0)] > 0$.

Consider next the case $i > 0$. The individual-rationality constraints in the CM, (23), can be expressed as
\[
\begin{align*}
\bar{z}^p &\leq \frac{\sigma\pi_\ell [\varepsilon_\ell u(q^*_f) - d^p_{\varepsilon_\ell}] + \sigma\pi_h \varepsilon_h u(q^*_h)}{i + \sigma\pi_h}, \quad (31) \\
\bar{z}^p &\leq \frac{\sigma\pi_h \varepsilon_h [u(q^*_h) - u(q^*_f)] + d^p_{\varepsilon_\ell}}{i + \sigma\pi_h}, \quad (32)
\end{align*}
\]
where I used that $d^p_{\varepsilon_h} = z^p$. Define $\bar{z}(i, d^p_{\varepsilon_\ell})$ as the upper bound on real balances consistent with (21), (31) and (32) when $q_{\varepsilon_h} = q^*_h$ and $q_{\varepsilon_\ell} = q^*_f$, i.e.,
\[
\bar{z}(i, d^p_{\varepsilon_\ell}) = \min \left\{ \frac{\sigma\pi_\ell [\varepsilon_\ell u(q^*_f) - d^p_{\varepsilon_\ell}] + \sigma\pi_h \varepsilon_h u(q^*_h)}{i + \sigma\pi_h}, \frac{\sigma\pi_h \varepsilon_h [u(q^*_h) - u(q^*_f)]}{i + \sigma\pi_h} + d^p_{\varepsilon_\ell}, d^p_{\varepsilon_\ell} + \varepsilon_h [u(q^*_h) - u(q^*_f)] \right\}.
\]
Then,
\[
\mu [A^*(i)] = \int_{\tilde{z}^p} \bar{z}(i, x) - \psi(q^*_h)]^+ dx + \int_{\tilde{z}^p} [\bar{z}(i, x) - \varepsilon_\ell [u(q^*_h) - u(q^*_f)]^+ dx,
\]
where $[x]^+ = \max(x, 0)$. It follows that $\mu [A^*(i)]$ is nonincreasing and continuous with $i$. Finally, from (21) and (32),
\[
\varepsilon_\ell [u(q^*_h) - u(q^*_f)] \leq d^p_{\varepsilon_h} - d^p_{\varepsilon_\ell} \leq \frac{\sigma\pi_h \varepsilon_h}{i + \sigma\pi_h} [u(q^*_h) - u(q^*_f)].
\]

A necessary condition for $A^*(i) \neq \emptyset$ is

$$\frac{\sigma \pi h}{i + \sigma \pi h} \geq \frac{\varepsilon i}{\varepsilon h}.$$ 

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

**Proof of Proposition 3.** The solution to (27)-(29) is

$$q^p = q^*$$

$$n^p = n^*$$

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

(33)

if $i \leq \tilde{i} \equiv \frac{\alpha \left( \frac{1 - n^*}{n^*} \right) \psi (q^*) - \psi (q^*)}{\psi (q^*)}$, where $\tilde{i}$ is obtained from (28) at equality with $q^p$ and $d^p$ as defined above. If $i > \tilde{i}$, then (28) holds at equality and

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

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

(34)

(35)

From (35)

$$n = \frac{1}{1 + \alpha^{-1} \frac{i \psi (q)}{u(q) - \psi (q)}}.$$ 

(36)

Substituting (36) into (34), the problem becomes

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

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

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

(37)

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

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

22
Using (37),
\[
\frac{\partial W}{\partial i} = -\frac{\psi(q^p)}{i\alpha'(1-n^p)} \left[ u'(q^p)\right] \left[ u'(q^p) - 1 \right].
\]

Next, we prove by contradiction that \( q^p < q^* \). Suppose \( q^p > q^* \). A profitable deviation consists in:
(i) reducing \( q^p \) to \( q^* \) in order to increase \( u(q^p) - \psi(q^p) \); (ii) adjusting \( d^p \) so that the buyers’ and sellers’ incentives to participate are unaffected. To see this denote \( d_0 \) and \( d_1 \) the values for \( d^p \) such that the participation constraints hold with equality, i.e.,
\[
-id_1 + \alpha \left( \frac{1-n^p}{n^p} \right) [u(q^*) - d_1] = 0
\]
\[
n^p (1-n^p) \alpha \left( \frac{1-n^p}{n^p} \right) [d_0 - \psi(q^*)] = 0.
\]
The assumption \( q^* < q^p \) implies \( \psi(q^*) < \psi(q^p) \) and \( u(q^*) - \psi(q^*) > u(q^p) - \psi(q^p) \). Hence,
\[
-id_1 + \alpha \left( \frac{1-n^p}{n^p} \right) [u(q^*) - d_1] = 0 < -id_0 + \alpha \left( \frac{1-n^p}{n^p} \right) [u(q^*) - d_0],
\]
and
\[
d_0 - \psi(q^*) = 0 < d_1 - \psi(q^*).
\]
By continuity, there is a \( d \in [d_0, d_1] \) such that \( \Gamma(d) = 0 \) where
\[
\Gamma(d) \equiv -id + \alpha \left( \frac{1-n^p}{n^p} \right) [u(q^*) - d] - \frac{n^p}{1-n^p} \alpha \left( \frac{1-n^p}{n^p} \right) [d - \psi(q^*)].
\]
This establishes that the deviation is incentive feasible and it raises \( W \). Consequently, \( q^p \leq q^* \). Moreover, from (37), \( q^p \neq q^* \) since otherwise \( n^p = n^* \), which is inconsistent with \( i > \bar{i} \). Consequently, \( q^p < q^* \) and, from (38), \( \frac{\partial W}{\partial n} < 0 \).

To establish that \( n^p < n^* \), notice first from (37) that \( n^p \neq n^* \). From (36) \( n \) is a decreasing function of \( q \). So if \( n^p > n^* \), one can reduce \( n \) and increase \( q \), which raises welfare.

Finally, I establish that aggregate real balances are decreasing with inflation. From (28) at equality and (35), \( W(i) = in^p d^p + U(c^*) - c^* \), with \( n^p d^p = \frac{M_i}{p_t} \). The result from above according to which \( \frac{\partial W}{\partial n} < 0 \) implies \( \frac{\partial (M_i/p_t)}{\partial n} < 0 \).