Faculdade de Economia,
Administração e Contabilidade
de Ribeirão Preto
Universidade de São Paulo

Texto para Discussão

Série Economia

TD-E 02 / 2015

A Lucas critique surmounted by inside money

Prof. Me. Fernando Antônio de Barros Júnior

Av. Bandeirantes, 3900 - Monte Alegre - CEP: 14040-905 - Ribeirão Preto - SP
Fone (16) 3315-3884 - e-mail: posgrad@fearp.usp.br site: www.fearp.usp.br

Universidade de São Paulo

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

A série TEXTO PARA DISCUSSÃO tem como objetivo divulgar: i) resultados de trabalhos em desenvolvimento na FEA-RP/USP; ii) trabalhos de pesquisadores de outras instituições considerados de relevância dadas as linhas de pesquisa da instituição. Veja o site da Comissão de Pesquisa em **www.cpq.fearp.usp.br**. Informações: e-mail: cpq@fearp.usp.br

Noname manuscript No.
(will be inserted by the editor)

A Lucas critique surmounted by inside money

Fernando Barros Jr. · Ricardo Cavalcanti · Caio Teles

5/11/15 - Preliminary and incomplete

Abstract The critique is that the literature on the optimum quantity of money, including that using random-matching models of outside money to measure welfare gains of expansionary policies, has not accounted for the negative effect of inflation on the velocity of money, or the alternative of subsidizing poor traders in particular meetings with superior effects on the allocation of risk. Simple extensions of random-matching models in which savings are sufficiently general to establish the critique indicate that outside money cannot provide a satisfactory explanation for expansionary policies but inside money can.

Keywords exchange risk · expansionary policies · the velocity effect of inflation

JEL: E52, E58.

1 Introduction

The old question of whether monetary expansions can lead to welfare gains, in addition to inflation, has not received to date a satisfactory answer. Having money as the solo device for intertemporal trade — a model of pure currency — Levine (1991), and Kehoe et al. (1992), deliver a formal study of the effects of simple policies on consumption levels and risk sharing. In principle, flat transfers cannot improve risk sharing without aggravating the problem that consumption is constrained by a low value of money, reduced further when inflation rises. Given this tension, Wallace (2014) argues that more sophisticated interventions should provide a robust, affirmative answer.

In this paper, after introducing small changes to matching models that unveil delicate interactions between incentives and risk sharing, we reach however a different assessment. We find that the outside money framework cannot provide a satisfactory explanation for expansionary policies. We state a kind of Lucas critique with two basic properties. The first is that the literature on the optimum quantity of money has ignored that money holders react to inflation by reducing savings and generating a more disperse distribution of money. As a result, when risk sharing is more important,

Barros, Cavalcanti and Teles
EPGE/FGV
Praia do Botafogo 190, Rio de Janeiro, RJ 22250-900, Brazil

as we show numerically to be the case for a large set of parameters, adverse effects of inflation on distributions push policies into a corner of inaction. The second is that in these economies, differently from those featuring centralized markets, social planners can improve risk sharing by dividing trade surpluses according to money holdings. Consequently, rich consumers can be asked to give away some surplus to poor producers without creating much distortions, in terms of adverse externalities on other meetings found with widespread inflation.[1]

In summary, something else other than typical models of outside money is needed. One possibility is to relate expansionary policies with business cycles. Alternatively, our exercises suggest taking the generality proposed by Wallace (2014) in the direction of sectorial creation of liquidity, using some monitoring of economic activity. We find that it is indeed useful to control spending by having newly issued money directed first to intermediaries, which in turn should finance credit operations more selectively. These are characteristics that take the models analyzed below in the direction of inside-money inflation.

The rest of the paper is divided as follows. In section 2, we provide more perspective on the complex issue of risk sharing in matching models of money. In section 3, we explain how a tax on consumers taking the form of financial profits can help. In section 4, we describe the mechanism-design problem with fiat money, lotteries and persistence of intermediation occupation. In section 5, we display computed optima for a variety of cases, completing our critique to the outside-money literature and demonstrating that inside-money inflation has a different nature. In Section 6, we further relate findings with previous work. In particular, we report the emphasis placed by Bagehot on how concentration of money in the hands of the bank sector is important for welfare. Section 7 concludes. The appendix presents supporting findings about pairwise models, including an example of optimal inflation used as a test case, and also describes auxiliary objects for the numerical approach.

2 Subtle features of exchange risk

Relatively little is known about the efficient allocation of risk in matching models (in macroeconomics, more broadly, fiscal and monetary policies are often studied separately). The same cannot be said about effects of inflation on the return of money (sometimes called hot-potato effects). But incomplete markets and thus exchange risk must be an integral part of the study of the optimal quantity of money. Since our critique is covering a lot of ground, it is important to break down the analysis of numerical output into simple parts.

One difficulty with exchange risk is that it is in part endogenous insofar trades are constrained by the distribution of money and therefore by savings behavior (many, including Levine (1991), have resorted to particular assumptions in order to avoid this complication). Another is that suitable terms of trades, that is, meeting-specific prices, can be 'distorted' to provide some insurance. Both issues are delicate as different dynamic effects are generated depending on how distortions are set. We have assembled, in Table 1, selected features of optimal allocations, for a basic case, to provide additional motivation.

[1] According to our critique, support for expansionary policies in the literature can be roughly classified as follows: linear utilities (Levine 1991) or low discount factors (Deviatov 2006) can turn

Table 1 Insurance when optimal inflation is zero: taxing consumers in pairwise meetings

β	0.9	0.8	0.7	0.6	0.5
			core on		
i	v / tax	v / tax	v / tax	v / tax	v / tax
0	0.4514 / -	0.0263 / -	0.0089 / -	0.0018 / -	0.0013 / -
1	1.2638 / 0.0000	0.5940 / 0.0000	0.3948 / 0.0000	0.2282 / 0.0000	0.2253 / 0.0000
2	1.5692 / 0.4067	0.7824 / 0.0000	0.5166 / 0.0000	0.3663 / 0.0502	0.2738 / 0.0698
3	1.7538 / 0.2271	0.8794 / 0.0000	0.5675 / 0.0000	0.3892 / 0.0000	0.2861 / 0.0000
4	1.9027 / 0.0552	0.9291 / 0.3964	0.5851 / 0.2921	0.4107 / 0.0000	0.3010 / 0.0000
			core off		
i	v / tax	v / tax	v / tax	v / tax	v / tax
0	0.7894 / -	0.2035 / -	0.0706 / -	0.0229 / -	0.0068 / -
1	1.2715 / 0.7270	0.5723 / 0.6670	0.3489 / 0.5905	0.2394 / 0.4873	0.1754 / 0.3521
2	1.5437 / 0.6512	0.0745 / 0.4713	0.4745 / 0.2932	0.3351 / 0.0976	0.2489 / 0.0000
3	1.7387 / 0.2268	0.8610 / 0.0000	0.5527 / 0.0000	0.3932 / 0.0000	0.2951 / 0.0000
4	1.8896 / 0.0000	0.9339 / 0.0000	0.5934 / 0.0000	0.4201 / 0.0000	0.3141 / 0.0000

v represents the expected discounted utility for each level of holdings i before meetings take place. In meetings where the consumer holds i and the producer holds 0, 'tax' is computed as $1-y/x$, where y is actual output and x is the cuttoff level determining zero surplus for the producer, keeping fixed the optimal payment.

In the economies underlying these simulations, people meet randomly in pairs, trading perishable goods for fiat money as in Shi (1995) and Trejos and Wright (1995). Expected utilities associated with money holdings, v, must be consistent with payments and output produced in no-coincidence meetings (explained in detail in incoming sections). A social planner chooses from stationary allocations according to average utility. In order to reach an affordable numerical task we assume that money is indivisible and that goods are traded for lotteries on money holdings, whose support is a subset of $\{0, 1, 2, 3, 4\}$. There is then a large set of possible stationary distributions of money and terms of trade — output and lotteries for each type of meeting, as indexed by traders' wealth — to choose from. Exchanges must be better than autarky in meetings for both traders (individual rationality). We also consider a more restrictive notion, the core requirement, that gives consumers room to propose themselves alternative trades that keep producers indifferent (a threat that the planner must anticipate as a 'core on' constraint).

We compute efficient allocations with core on and off, reaching a numerical output too large to be fully addressed at this introductory level. Also a key parameter, the discount factor β, must be allowed to vary if we want to understand the relationship between risk and returns. For now, some interesting insights are revealed by measures of maximum output that poorest producers could be asked to hand out, given the expected utilities they are actually receiving as payments (not displayed). The relative differences between these ceilings and quantities actually delivered, labeled *tax*, represents a loss to consumers that varies with their holdings of money. According to Table 1, for all β, the poorest producer is always receiving a surplus in some meetings (identified by positive taxes). The precise wealth i of whom is taxed depends on β and whether the core requirement is on or off.

off negative effects of inflation on the distribution of money, while giving consumers all trade surpluses (Molico 2006) removes an important tool for risk sharing.

To add to the significance of Table 1 we make now two comments on previous work. First, in attempts to reproduce Levine (1991) findings by Molico (2006) and Deviatov (2006), corresponding tax statistics are always zero. In the former case due to a bargaining assumption and in the latter, we claim, because the support used was too small (a 2-unit upper bound), leading to a flatter v (little affected by Inada conditions). Second, we did allow for a lump-sum transfer of money followed by inflation in the same fashion as Deviatov (2006) did, but no improvements were found (regardless of β or the core requirement).

As it will become clear, expansionary policies create unwanted distortions on risk sharing. Regarding these examples of taxation, notice a positive expected utility for those holding zero, although there are no government transfers (we are normalizing the utility of autarky to be zero). Section 3 is devoted to understanding how taxation can reduce a possible wedge between social and private savings. In order to avoid complicated dynamic effects, we resort there to commodity money and quasi-linear preferences. Relevant histories of savings are entirely captured by exogenous shocks to utilities. When people trade in pairs we do not find any use for taxation: it is efficient to give consumers all surpluses from trades. But when intermediation is added, and meetings include a third trader that can lend to consumers, we find that savings by intermediaries are important and that taxation can provide a better allocation of risk. This notion of credit makes expansionary policies more powerful and capable of improving the distribution of fiat money in other sections.

What fiat and commodity money have in common is a sort of externality: consumers may spend too much fiat money, or traders may save too little commodity money, because they ignore that their actions may facilitate trades of others. In the case of the commodity-money model they can help lending money. In the case of the fiat-money model this too can be important but, in addition, there is the problem that a large payment to someone makes that person less willing to produce in the future (capturing this requires dropping the quasi-linearity assumption). This fiat-money externality, which we call the velocity effect, is important also to understand Table 1. We show that, as β increases, the wedge between private and social savings is reduced, so that taxing richer consumers promotes more risk sharing.[2]

We are aware that some jargon, specific of exchange models, may distract readers that would prefer to see first generic punch lines. One attempt is as follows. When trying to provide insurance, the monetary authority has to combine distortionary taxation (inflation) with simple transfer schemes. As inflation arises, a first consequence of money losing value is a reduction in the intensive margin of consumption. But even if policymakers have confidence in some sort of statistical prediction of how the 'demand' for money varies with inflation rates, they will still predict poorly the effects of lump-sum transfers on consumption smoothing because traders will insure themselves less (they will trade more consumption in the present for less insurance in the future). In addition, a correct assessment of welfare gains of expansionary policies depends on the way insurance can be promoted through the selection of (prices and) taxes in trade meetings.

After investigating these issues in more detail we want to conclude that (i) the level of exchange risk in matching models is just too high to be ignored; (ii) there is some sort of externality caused by the private reaction to inflation, to the extent

[2] Taxation changes a little with the core off and consequent improvements in risk sharing, giving the problem of low return a higher priority. This may explain why poor consumers become more heavily taxed.

that traders increase payments ignoring that those receiving money are less inclined to produce in the future; and (iii) although for a large set of parameters it is best to just avoid expansionary policies altogether, society can provide better insurance with less distortions by supporting inside money with some level of monitoring.

3 Taxing commodity money

In this section, we build on the model of Cavalcanti and Puzzello (2010) (hereafter C&P).

3.1 The environment

Time is discrete and each period is divided into two subperiods. The economy has a large population living forever and experiencing random meetings in the first subperiod, and preference shocks in the second. Preference shocks are realizations of an *iid* process. There is a durable good called money that can be consumed and produced in the second subperiod, according to an idiosyncratic marginal utility θ drawn every date from an uniform distribution. For simplicity, we assume a discrete support $\{\theta_1, ..., \theta_n\}$ and let F, such that $F(\theta_i) = \frac{i}{n}$ for all i, denote the cumulative distribution of θ. In addition, we normalize its mean, setting $\sum_i \frac{1}{n}\theta_i = 1$.

Money is hence a commodity, produced and consumed when people are by themselves, according to linear utility that is the realization of a preference shock. Money balances are planned in order to reach ideal savings, for each θ, for use as a medium of exchange in the next period, first subperiod, when random meetings take place. Money holdings are observable in meetings but trade histories are private information and people cannot commit to future actions.

There is no discounting between first and second subperiods, but there is discounting at the common factor β across dates. We assume θ_i is increasing in i with $\theta_1 > \beta$, so that savings are always costly. There is also the standard specialization of production and consumption in meetings. We assume that every meeting is formed by three people: a producer, an intermediary and a consumer. We assume that a person has equal probability of taking part in a meeting in any of these three occupations. And that the meeting is a single-coincidence meeting, when the first person can produce a perishable good for the third one, with probability 3α, where $\alpha \leq \frac{1}{3}$. With probability $1 - 3\alpha$ there are no potential gains from trade. The utility of consuming $c \in \mathbb{R}_+$ units of the perishable good is $u(c)$, and the utility of producing c units of the perishable good is $-c$. We assume that $u(0) = 0$ and that u is continuous, concave, differentiable and such that $u'(0) = +\infty$ and $u(c) < c$ for c sufficiently large.

We assume that the only feasible trade in a meeting has the intermediary transferring money to the producer, as loan to the consumer, in exchange for goods produced. Then, after production takes place and the producer leaves the meeting, the consumer is able to receive goods and to pay out the loan with the intermediary.

In this economy, the planner's problem is to maximize the present value of average utility by choice of incentive-compatible allocations that provide a suitable level of insurance against shock θ and exchange risk. Following C&P, we restrict attention to stationary allocations. We also anticipate that, due to the quasi-linear structure, optimal allocations are not functions of past histories. A meeting is a vector $m = (m_1, m_2, m_3)$ describing holdings of money of the producer, m_1, the intermediary, m_2 , and the consumer, m_3. An allocation is a list (s, x, y, z) describing saving plans s in the second subperiod, as a function θ, and trade plans (x, y, z) in the first subperiod, as a function of m. Saving plans say how much money people will take with them when leaving the second subperiod, according to the realization of idiosyncratic shocks. Trade plans describe loan size x, output level y, and payment amount z. That is, z is the reduction in holdings of money suffered by the consumer, x is how much the producer receives, and $z - x$ is the intermediation profit. We require money transfers to be feasible in the sense of $x(m) \leq m_2$ and $z(m) \leq m_3$.

A plan $s : \{\theta_1, ..., \theta_n\} \to \mathbb{R}_+$ generates a distribution of money μ on $\mathbb{R}_+$. It is convenient to denote by μ^3 the distribution of meetings on $\mathbb{R}^3_+$ generated by μ, and by μ^2 it marginal distribution on $\mathbb{R}^2_+$ when one coordinate of m is fixed. The welfare criteria corresponds to the utility of an ex-ante representative agent and can be written as

$$w(s, y) = -\int (\theta - \beta) s(\theta) dF(\theta) + \alpha\beta \int (u(y(m)) - y(m)) d\mu^3(m). \tag{1}$$

Notice that the welfare function w does not depend on monetary payments. This is so because expected utility, when leaving meetings, as a function of after-trade holdings, is the same for all traders. Hence, no matter how money is divided by trade, the average discounted value attached to after-trade holdings is $\beta \sum_i \frac{1}{n} s(\theta_i)$.

3.2 Implementable allocations

We also follow the notion of implementability adopted by C&P, that is, that traders agree with (s, x, y, z), given μ associated to s, if autarky in meetings would not make them better off, and if there are no other saving choices that could improve individual utility given (x, y, z) and μ. We shall leave the discussion of group deviations for the fiat money environment of following sections. But while in C&P it is optimal to give all surplus to consumers, here this is not so due to an externality associated to savings decisions of people who end up in position to make loans.

In order to be implementable, an allocation must satisfy incentive constraints. Trade incentive constraints are given by

$$y(m) \leq x(m),\ x(m) \leq z(m) \text{ and } z(m) \leq u(y(m)). \tag{2}$$

These inequalities ensure that trade surpluses are nonnegative in all meetings. The saving incentive constraint is that $s(\theta)$ must solve the problem of maximizing $-(\theta - \beta)k + \alpha\beta v(k)$ by choice of money holdings k , where the expected gain from trade $v(k)$ is defined by

$$v(k) = \int (u(y(a, a', k)) - z(a, a', k) + z(a, k, a') - x(a, k, a') + x(k, a, a') - y(k, a, a')) d\mu^2(a, a'). \tag{3}$$

Because the distribution of money μ in turn must be generated by s, an incentive-compatible savings plan is a fixed point for each (x, y, z). Allocations that are feasible and incentive compatible are called implementable.

3.3 Welfare bounds

With intermediation, part of money held by relatively rich consumers cannot be used to weaken incentive constraints of producers. For a given stock of money, average utility in meetings depends only on consumption and production, not on how money is divided among traders. But if intermediation activity receives no compensation, incentives to save can be suboptimal. Some allocations that are easy to compute can provide informative welfare bounds.

Let (s^*, x^*, y^*, z^*) denote the solution of the planner's problem of maximizing $w(s, y)$ in the set of implementable allocations. Let us first turn off the intermediation constraint (the cash-in-advance requirement $x(m) \leq m_2$), denoting by $(\hat{s}, \hat{x}, \hat{y}, \hat{z})$ a solution of the corresponding relaxed problem. Notice that, in this case, the incentive constraints $y(m) \leq x(m)$ and $x(m) \leq z(m)$, together with the feasibility constraint $z(m) \leq m_3$, imply the inequality $y(m) \leq m_3$. It turns out that $w(\hat{s}, \hat{y})$ is the optimal welfare of a random-matching model without intermediaries.

In the following proposition, a comparison is made with another relaxed problem, obtained by imposing $x(m) \leq m_2$ but ignoring saving incentive constraints, as if s can be imposed on individuals. If $(\tilde{s}, \tilde{x}, \tilde{y}, \tilde{z})$ denotes the solution of this second problem, the following holds.

Proposition 1 *For m in the support of distributions of meetings, output is $\hat{y}(m) = m_3$ when intermediation is relaxed, and $\tilde{y}(m) = \min\{m_2, m_3\}$ when savings need not be incentive compatible. In these relaxed problems, moreover, welfare satisfies $w(\hat{s}, \hat{y}) \geq w(\tilde{s}, \tilde{y}) \geq w(s^*, y^*)$, with inequalities replaced by equalities when there is a single type of trader.*

Proof See appendix.

3.4 Taxes and financial profits

In the absence of intermediation (see C&P), if the consumer extracts all surplus in meetings then $g(k) = \beta[u'(k) - 1]$ is the marginal private gain from bringing an additional unit of money to meetings when savings is k. The next proposition explores the fact that, with such terms of trade and intermediation, incentive-feasible savings satisfy $\theta - \beta = \alpha F(\theta) g(s(\theta))$.[3] From a social perspective, however, each additional unit saved also affects, with probability α, the volume of resources lent to rich consumers, so that if savings could be imposed to satisfy $\theta - \beta = 2\alpha F(\theta) g(s(\theta))$ a welfare gain would follow. This insight is explored in the proposition to demonstrate that a perturbation of the rule of giving all surpluses to consumers improves welfare.

When all surpluses are given to consumers, output in meetings is given by $y(m) = \min\{m_2, m_3\}$ and idle holdings $m_3 - m_2$, when positive, remain with consumers. Taxes can improve savings without reducing intensive margins of consumption for a fixed m. To see this, consider the following perturbation with transfers that increase the incentives to save for all types, except the richest one.

[3] In the proposition, the first-order condition for the savings problem is written in terms of left derivatives. For numerical examples, incentive-compatible saving s_1, for instance, is found assuming that all other type-1 people are saving a bit more than s_1 and then finding the interior solution $\theta_1 - \beta = \frac{\alpha}{n} g(s_1)$. More generally, savings are found independently for all grid points.

We let $y(m) = x(m) = \min\{m_2, m_3\}$ for all m but allow part of $m_3 - \min\{m_2, m_3\}$ to be transferred to intermediaries. Let $(\bar{s}_1, ..., \bar{s}_n)$ denote incentive-compatible savings for the no-taxation allocation. It is straightforward to show that the saving problem is convex and that $\bar{s}_i > \bar{s}_j$ whenever $\theta_i < \theta_j$. The new (called *profit*) allocation is constructed as follows. First a quantity limit $\varepsilon > 0$ and an interest rate $r > 0$ are fixed. Then, when consumer with m_3 holdings meets an intermediary with m_2, for $m_3 > m_2$ and $\mid m_3 - m_2 \mid > \varepsilon$, then some interest rx is paid to this intermediary if he or she is providing $x \in (\bar{s}_j, \bar{s}_j + \varepsilon)$ in loans. Hence, in the profit allocation, $z(m) = y(m) + rm_2$ in meeting m such that m_2 is discretely lower than m_3. The values of ε and r are chosen sufficiently small so that idle money $m_3 - m_2$ in such meetings is greater than the extra payment rm_2, and also to insure that each type does not envy savings designed for another type.

Proposition 2 *When there is more than one type of trader, welfare is increasing in the profit rate r in a neighborhood of zero, so that it is not optimal to give all surpluses to consumers.*

Proof See appendix.

A numerical illustration of welfare gains promoted by taxation is as follows.[4] Table 2 displays basic statistics of allocations as r varies. We find that, for a broad range of values of r, tax payments never exceed idle holdings $m_3 - m_2$ in meetings.

Table 2 Savings according to profit rates

r	s_1	s_2	s_3	s_4	s_5	Welfare
0%	8.5235	4.5027	1.9184	1.0917	0.7832	3.4733
5%	8.5235	4.5621	1.9361	1.0991	0.7876	3.4758
10%	8.5235	4.6228	1.9541	1.1067	0.7921	3.4783
15%	8.5235	4.6850	1.9724	1.1143	0.7966	3.4807

(1) Values multiplied by 100.
(2) If r is increased to 16% then the richest saver is willing to change behavior (to avoid paying next-type profits).

The proof of proposition 2 can be strengthened. Uniform distributions are not needed but we have omitted a more general treatment for ease of exposition. The fact that, with quasi-linear preferences, a third trader is needed to show that taxation has a role may explain why this point has not appeared formally in the literature. The proof, of course, is simplified by the absence of wealth effects: due to the quasi-linear structure, money transferred to intermediaries (or producers) do not change adversely their incentives to produce in the future. In order to allow for such effects we have unfortunately to resort to numerical methods when we discuss economies with fiat money.

4 Taxing fiat money

We build on the concept of three-party meetings, used in the last section. In addition, exchange risk can also be mitigated with money creation by the government.

[4] We set $u(y) = \sqrt[4]{y}$, $\beta = .6$ and $\alpha = .2$. We let the support of θ be $\{.614, .675, .877, 1.215, 1.619\}$ and set $\varepsilon = 1.854 \times 10^{-3}$.

Except in extreme cases that we interpret as perfect monitoring, money is in exogenous supply (it is *outside* money) and is fiat (it does not provide direct utility). We restrict attention to steady states by choice of a law of motion for holdings that resembles an inflationary process. We also allow for persistent occupation in sectors, group deviations and a limit case with inside money.

As anticipated in section 2, expansionary policies are not needed in the particular case of pairwise trades (and the upper bound on holdings of 4 units). To understand this better, we resort to a smaller support (an upper bound of 2) that reduces the role of consumer taxes described in Table 1, creating more room for expansionary policies with inflation (which affects the whole population). By adding intermediation with different persistence parameters or removing the core requirement, we uncover the role of velocity of money that is not found in the quasi-linear structure of the previous section, and which introduces dynamic effects justifying the zero-inflation result of Table 1. In addition, intermediation allow us to predict that inside-money inflation is robust.

4.1 The environment

A steady-state allocation is now $(\mu, y, \lambda, \tau, \pi)$, where μ is a distribution of money, y defines output for each meeting $m \in M$, λ defines payments in terms of lotteries, also for each meeting m, $\tau = (\tau^n, \tau^b)$ describes occupation-dependent transfers, and π is a measure of inflation. People start each period carrying 0, 1 or 2 units of money, so that $M = \{0, 1, 2\}^3$, either in the *bank* (intermediation) sector or in the complement, the *nonbank* sector. The cases of pairwise meetings with bounds 2 and 4 (leading to Table 1 in Section 2) are straightforward modifications of the specification with intermediation and have detailed results presented in the appendix. More comments about pairwise meetings are made when numerical results are discussed.

After trades occur, holdings of money evolve according to a stochastic process reflecting inflationary transfers. Bank and nonbank occupations are idiosyncratic shocks evolving according to a first-order Markov process. In particular, the probability that bank people keep their occupation in the next period is ρ, and that for nonbank people is $\frac{1+\rho}{2}$. As a result, in a steady state, the bank sector is always composed by one-third of the population.

We let $m = (m_1, m_2, m_3)$ to denote that money holdings are m_1 for the producer, m_2 for the intermediary, and m_3 for the consumer. The ex ante probability that a nonbank person becomes a consumer or a producer in a meeting is $\frac{\alpha}{2}$. Intermediaries, like nonbank people, take part in a no-coincidence meeting with probability $1-\alpha$. We denote by μ_i^b the fraction of people starting a period in the bank sector and holding i, and by μ_i^n that in the nonbank sector and also holding i, where $i \in \{0, 1, 2\}$. In what follows, we often omit the qualification 'coincidence meeting' about m whenever it is clear from the context.

Consumer and producer utilities are again $u(c)$ and $-c$, respectively, and the discount factor is also β. In meeting m, output is deterministic and often denoted by $y(m)$, while there is a probability distribution $\lambda(m)$ defining transfers of money among the three traders. More specifically, for $i = 1, 2, 3$, we let $\lambda_i^j(m)$ denote the (marginal) probability that 'person i' (the person starting with m_i) leaves the meeting holding $j \in \{0, 1, 2\}$ units of money. Hence $\lambda_1^j(m)$ denotes the probability that the producer leaves the meeting holding j units of money. In what follows, Bellman equations are more easily expressed by having $\lambda(m)$ written as a vector, so that $\lambda_i^j(m)$ as a particular coordinate of $\lambda(m)$ (see appendix for more details).

We assume initially that no money can be created or destroyed in meetings, and there are physical restrictions on money flows in meetings dictated by intermediation (this assumption is eventually modified when inside money is discussed later). For now, we say that an (outside-money) allocation is *feasible*, reflecting intermediation frictions of the previous section, if for all $m \in M$ two flow conditions are satisfied. As a first condition, we require that $\lambda_1^{m_1+p}(m) = \lambda_2^{m_2-p}(m)$ for all $p \in \{0,1,2\}$ and, moreover, if $p > \min\{m_2, 2-m_1\}$ then $\lambda_1^{m_1+p}(m) = \lambda_2^{m_2-p}(m) = 0$. That is, if payment to producer has mass on p then the intermediary transits to state $m_2 - p$ with the same probability that the producer transits to $m_1 + p$. Likewise, as a second condition, for every realization p for this payment, we require that $\lambda_2^{m_2-p+q}(m) = \lambda_3^{m_2-q}(m)$ for all $q \in \{0,1,2\}$ and, moreover, if $q > \min\{m_3, 2-m_2+p\}$ then $\lambda_2^{m_2-p+q}(m) = \lambda_3^{m_2-q}(m) = 0$. That is, if a payment to an intermediary has mass on q then the consumer transits to state $m_3 - q$ with the same probability that the intermediary transits to $m_2 - p + q$.

After meetings, but before the period ends, money holdings are affected by policy and new occupation draws take place. We describe policy as transition matrices detailed in the appendix. First there is an inflation shock: a matrix with parameter π is constructed to capture the probability that money disappears, regardless of occupation. A person with one unit has holdings transiting to 0 with probability π, and not transiting with probability $1-\pi$. A person with two units has holdings transiting to 1 with probability $2(1-\pi)\pi$, and to 0 with probability π^2. After the π-shock holdings are updated by a transfer matrix with parameter $\tau = (\tau^b, \tau^n)$. After-inflation holdings j transit to state $j+1$ with probability τ^b (τ^n), and remain in state j with probability $1-\tau^b$ $(1-\tau^n)$ if $j < 2$, for people in the bank (nonbank) sector. If $j = 2$, the probability of transition is zero.

We say that an allocation is *stationary* if, given λ and (τ, π) , $\mu = (\mu^b, \mu^n)$ is a time-invariant distribution of money (see details in the appendix).

Notice, for a given λ, the effect on μ of increasing (τ^b, τ^n, π) above $(0,0,0)$ is to reduce the mass of people with holdings in $\{0,2\}$, in exchange for an increase in the mass of people holding one unit. In principle this policy improves extensive margins, although it now has a potentially negative effect on the return of money, that can reduce y. As we shall see, however, one must account for changes in λ that are incentive-compatible with saving/spending decisions and which can worsen extensive margins as well. For this we need to describe incentive constraints, according to continuation values, defined as follows.

4.2 Welfare criteria and rationality constraints

We now present the welfare criteria and incentive constraints, whose details are also included in the appendix. At the beginning of a period, the expected discounted utility of a person with i units of money in bank and nonbank sectors take, respectively, the following form

$$v_i^b = (1-\alpha) w_0^b(i) + \alpha \sum_{\{m: m_2 = i\}} \mu_{m_1}^n \mu_{m_3}^n w_2^b(m),$$

and

$$v_i^n = (1-\alpha)w_0^n(i) + \frac{\alpha}{2}\left(\sum_{\{m:m_1=i\}} \mu_{m_2}^b \mu_{m_3}^n w_1^n(m) + \sum_{\{m:m_3=i\}} \mu_{m_2}^b \mu_{m_1}^n w_3^n(m)\right),$$

where w_0^b and w_0^n results from transitions after a no-coincidence meeting, while w_1^n results from transitions after a meeting as a producer, w_3^n results from transitions after a meeting as a consumer, and w_2^b results from transitions after a meeting as an intermediary. In the appendix it is presented the system defining value functions in detail. In particular it is shown that for $m \in M$, $w(m)$ takes the form

$$\begin{aligned} w_1^n(m) &= -y(m) + \beta\lambda_1(m)A^n v \\ w_2^b(m) &= \beta\lambda_2(m)A^b v \\ w_3^n(m) &= u(y(m)) + \beta\lambda_3(m)A^n v \end{aligned}$$

where A^n and A^b are transition matrices reflecting current occupation. Likewise,

$$\begin{aligned} w_0^b(i) &= \beta A_{0i}^b v \\ w_0^n(i) &= \beta A_{0i}^n v \end{aligned}$$

where A_{0i}^b and A_{0i}^n are particular matrices for those holding i units of money in no-coincidence meetings. For a given (μ, λ, y) and policy (τ, π) this system has a contraction property and features an unique solution v.

The welfare criteria is given by average utility, corresponding to an inner product of $\mu = (\mu^b, \mu^n)$ and $v = (v^b, v^n)$, which amounts to

$$w(y,\mu) = \mu \cdot v = \frac{\alpha}{1-\beta}\sum_{m\in M} \mu_{m_1}^n \mu_{m_2}^b \mu_{m_3}^n [u(y(m)) - y(m)]. \tag{4}$$

Remark 1 Lotteries λ and policy parameters (τ, π) have only indirect effects on w. The same can be said about β, since it does not change preference orders over stationary outcomes from the social perspective.

We assume that individuals can deviate during trades from what is proposed for a particular meeting, taking as given value functions and the law of movement for aggregate variables. They can deviate individually, by choosing autarky in the meeting, or in groups, by seeking a trade bundle that dominates the proposed allocation for the meeting, without making trade partners worse off. Given such notion of rationality, implementable allocations must satisfy inequalities corresponding to individual-rationality and core requirements. Trade weakly dominates autarky in meeting m for an intermediary if

$$w_2^b(m) \geq w_0^b(m_2), \tag{5}$$

and for producer and consumer if

$$w_1^n(m) \geq w_0^n(m_1) \text{ and } w_3^n(m) \geq w_0^n(m_1). \tag{6}$$

Individuals can also consider group deviations in a meeting. One way to define the requirement that trade belongs to the core in meeting m is to allow the consumer to search for an alternative output/lottery pair $(\bar{\lambda}, \bar{y})$, subject to intermediation constraints with preservation of money holdings defined above, so as to find

$$\bar{w}_3^n(m) = \max_{\bar{\lambda}, \bar{y}} u(\bar{y}) + \beta \bar{\lambda}_3 A^n v \tag{7}$$

$$\text{s.t. } -\bar{y} + \beta \bar{\lambda}_1 A^n v \geq w_1^n(m) \text{ and } \beta \bar{\lambda}_2 A^b v \geq w_2^b(m).$$

A feasible and stationary allocation $(\mu, y, \lambda, \tau, \pi)$ is *implementable* if associated values (v, w) satisfy individual-rationality (5-6) and core constraints

$$w_3^n(m) \geq \bar{w}_3^n(m) \tag{8}$$

for all $m \in M$.[5]

Remark 2 An intuitive description of constraint (8) can be given with pairwise meetings (no intermediation), differentiable value functions (divisible money) and degenerate lotteries. First-order necessary conditions for an interior solution to (7) can be shown to imply, in this case,

$$v'(m_3 - p) = u'(y) v'(m_1 + p)$$

where v' is the derivative of the value function, $m_3 - p$ is after-trade consumer holdings of money, $m_1 + p$ is after-trade producer holdings of money, and y is output. Notice that, according to this condition, money payment p is inversely related to output level y when v is concave. In particular, if β is low and, in turn, individual-rationality requires low output, then due to the core requirement p must be high. Average trades therefore feature high spending when β is low.

5 The velocity effect

In this section we report which implementable allocations, among those that are stationary with respect to the inflationary process described above, solve the welfare maximization problem for many specifications. As explained, pairwise meetings is a particular case, which is also discussed below (subsection 5.3). We report in this case solutions for a large range of discount factors, demonstrating that in these economies, when the upper bound is 4 units, expansionary interventions are not optimal. The appendix reproduces findings by Deviatov (2006) (with the upper bound of 2), and a figure in subsection 5.3 makes a contrast with results for the larger upper bound.

Our critique points out that a key feature for understanding these findings is the velocity effect, a private reaction to expansionary policies causing an amplification of distortions created by inflation on the value of money. Intermediation with the upper bound of 2 units is not only within reach numerically but is also easier to interpret (there are fewer output levels and lottery choices, relative to Tables 9 and 10). Hence, it is convenient to leave the discussion of pairwise meetings to the end of this section, after velocity effects are well documented.

To describe the velocity reaction to inflation and how inside money can reintroduce expansionary policies, we compute three sets of simulations for economies with intermediation. The first two concern outside-money economies exactly as described in section 4. The third set describes results for extreme values of occupation persistence, allowing for an inside-money interpretation of the model. In terms of parameters introduced in the previews section, we set $\alpha = 1$ and $u(y) = y^{2/10}$.

[5] Our algorithm (see appendix) is written with a more general formulation for (8).

5.1 Outside-money inflation

In the first set of simulations we put $\beta = .9$, while in the second set we put $\beta = .5$. In both cases, we vary the parameter ρ that determines how persistent the intermediation occupation is. In these outside-money examples we find that at most one unit is transferred and take advantage of this fact, reporting in tables 5 and 6, the probability λ that the consumer pays a unit of money. Also, in almost every meeting, the payment from intermediaries to producers is equal to the payment from the consumers to the intermediaries; an exception may occur in meeting $(1, 1, 2)$. When that happens, consumers pay exactly one unit and we report with entry 'profit $(1, 1, 2)$' the probability that the intermediary is leaving the meeting with two units. Finally, we report y relative to $\arg\max_x\{u(x) - x\}$, which for our specification is $y^* = .1337$.

Table 3 Outside money, $\beta = .9$ and core on

Persistence	iid	Markov low	Markov high
m	y / λ	y / λ	y / λ
(0,1,1)	1.0000 / 0.19	1.0000 / 0.19	1.0000 / 0.24
(0,1,2)	4.4824 / 1.00	4.4211 / 1.00	1.9903 / 1.00
(0,2,1)	1.0000 / 0.19	1.0000 / 0.19	1.0000 / 0.24
(0,2,2)	4.4824 / 1.00	4.4211 / 1.00	1.9903 / 1.00
(1,1,1)	0.2229 / 0.14	0.2266 / 0.14	0.2842 / 0.19
(1,1,2)	1.0000 / 0.71	1.0000 / 0.72	0.3987 / 1.00
(1,2,1)	0.2229 / 0.14	0.2266 / 0.14	0.2842 / 0.19
(1,2,2)	1.0000 / 0.77	1.0000 / 0.67	1.0000 / 0.65
profit (112)	0	0	0.74
μ_0^n / μ_0^b	0.1452 / 0.1452	0.1455 / 0.1455	0.2277 / 0.0311
μ_1^n / μ_1^b	0.5550 / 0.5550	0.5581 / 0.5581	0.4938 / 0.1084
μ_2^n / μ_2^b	0.2998 / 0.2998	0.2964 / 0.2964	0.2785 / 0.8605
v_0^n / v_0^b	0.0910 / 0.0780	0.1167 / 0.0875	0.4589 / 0.5998
v_1^n / v_1^b	0.9087 / 0.7789	0.9489 / 0.7117	1.1372 / 0.6395
v_2^n / v_2^b	1.1150 / 0.9900	1.2024 / 0.9018	1.3683 / 0.6456
π	0	0	0.0385
τ^n	0	0	0
τ^b	0	0	0.7794

Values for ρ are 1/3, 2/3 and .9 for, respectively, iid, Markov-low and -high. π is the inflation rate, τ^k is the transference for sector k and μ_i^k / v_i^k is the fraction / value function of people in sector k holding i units of money.

We notice first that without persistence in intermediation occupation (*iid* case), as in the pairwise economy of Deviatov (2006) (see appendix), inflationary interventions are only optimal when the discount factor β has a low value. Hence this corner condition, with consumers saving zero, is robust to the introduction of intermediation. Corners are easily hit because, with the small support for holdings, value functions are relatively flat and Inada conditions do not help generating positive savings. In these corners, velocity effects are turned off and stop imposing welfare losses when people have a low propensity to save.

By contrast, when $\beta = .9$, as in table 3, the distribution of money can be considered a good one, as about 56% of people have one unit of money, without any redistributive intervention. Hence it becomes a good thing to have zero inflation, and an average

Table 4 Outside money, $\beta = .5$ and core on

Persistence	iid	Markov low	Markov high
m	y / λ	y / λ	y / λ
(0,1,1)	0.2603 / 1.00	0.5146 / 1.00	0.5692 / 1.00
(0,1,2)	0.2603 / 1.00	0.5146 / 1.00	0.5692 / 1.00
(0,2,1)	0.2603 / 1.00	0.5146 / 1.00	0.5692 / 1.00
(0,2,2)	0.2603 / 1.00	0.5146 / 1.00	0.5692 / 1.00
(1,1,1)	0.0845 / 1.00	0.1571 / 1.00	0.1967 / 1.00
(1,1,2)	0.0845 / 1.00	0.1571 / 1.00	0.1967 / 1.00
(1,2,1)	0.0845 / 1.00	0.1571 / 1.00	0.1967 / 1.00
(1,2,2)	0.0845 / 1.00	0.1571 / 1.00	0.1967 / 1.00
profit (112)	0	0	0
μ_0^n / μ_0^b	0.2357 / 0.2357	0.3053 / 0.1221	0.3479 / 0.0366
μ_1^n / μ_1^b	0.3826 / 0.3826	0.3426 / 0.2427	0.3637 / 0.1252
μ_2^n / μ_2^b	0.3816 / 0.3816	0.3521 / 0.6351	0.2885 / 0.8382
v_0^n / v_0^b	0.0149 / 0.0597	0.0096 / 0.0542	0.0385 / 0.0264
v_1^n / v_1^b	0.1471 / 0.0642	0.2067 / 0.0593	0.2732 / 0.0282
v_2^n / v_2^b	0.1639 / 0.0652	0.2413 / 0.0601	0.3166 / 0.0286
π	0.2241	0.1576	0.2042
τ^n	0	0.0128	0.1751
τ^b	1	1	1

Values for ρ are 1/3, 2/3 and .9 for, respectively, iid, Markov-low and -high. π is the inflation rate, τ^k is the transference for sector k and μ_i^k / v_i^k is the fraction / value function of people in sector k holding i units of money.

monetary spending of just .14 in meetings $(1,1,1)$ and $(1,2,1)$ allows this distribution of money to remain stationary.

Now, if $\beta = .5$ then core constraints, together with producer incentive constraints, push allocations towards low savings and negative effects of inflation are reduced, yielding a measure of optimal inflation of about .22, as we can see in table 4. Meetings $(1,1,1)$ and $(1,2,1)$, that are key for keeping a good distribution of money without inflation, feature no savings at all. The .22-inflation expansion prevents a very bad distribution of money from taking place, so that about 38% of people hold one unit of money.

Effects of velocity and discounting on saving rates become evident when the core constraint is turned off. If this is done for $\beta = .9$ then about 77% of the population are always holding one unit of money in a better distribution relative to the case with core on. This is due to a smaller monetary payment of .02 on average becomes implementable in meetings $(1,1,1)$ and $(1,2,1)$. If $\beta = .5$, turning off the core requirement allows spending in these meetings to fall from maximum levels to .02, delivering a good distribution with about 74% of the population holding one unit, without inflation.

Notice that this pattern is robust to specifications with low persistence in intermediation occupation. When persistence parameters is set as 1/3 (*iid* case) or 2/3 (*Markov-low* case), inflation appears only when $\beta = .5$ and the core requirement is on. When $\beta = .9$ or the core requirement is off, low spending in meetings $(1,1,1)$ and $(1,2,1)$ suffices to generate a good extensive margin. We still find, nevertheless, that a small but robust inflation appears when persistence in intermediation occupation is high. Even when $\beta = .9$ and the core is off, a case of good spending limits, we see the necessity of an inflation measure of .028. Here, however, the intermediation friction

Table 5 Outside money, $\beta = .9$ and core off

Persistence	iid	Markov low	Markov high
m	y / λ	y / λ	y / λ
(0,1,1)	0.9387 / 1.00	0.9402 / 1.00	0.9556 / 1.00
(0,1,2)	2.6305 / 1.00	2.6746 / 1.00	1.9684 / 1.00
(0,2,1)	0.9387 / 1.00	0.9402 / 1.00	0.9556 / 1.00
(0,2,2)	2.6305 / 1.00	2.6746 / 1.00	1.9684 / 1.00
(1,1,1)	0.0509 / 0.02	0.0524 / 0.02	0.1297 / 0.08
(1,1,2)	1.0000 / 0.34	1.0000 / 0.33	0.4720 / 1.00
(1,2,1)	0.0509 / 0.02	0.0524 / 0.02	0.1297 / 0.08
(1,2,2)	1.0000 / 0.34	1.0000 / 0.33	1.0004 / 0.58
profit (112)	0	0	0.73
μ_0^n / μ_0^b	0.0637 / 0.0637	0.0634 / 0.0634	0.1709 / 0.0390
μ_1^n / μ_1^b	0.7735 / 0.7735	0.7748 / 0.7748	0.5724 / 0.1703
μ_2^n / μ_2^b	0.1628 / 0.1628	0.1618 / 0.1618	0.2567 / 0.7907
v_0^n / v_0^b	0.5735 / 0.4916	0.6120 / 0.4590	0.7285 / 0.3451
v_1^n / v_1^b	0.9839 / 0.8433	1.0266 / 0.7700	1.1546 / 0.5469
v_2^n / v_2^b	1.4433 / 1.2371	1.4994 / 1.1246	1.6662 / 0.7893
π	0	0	0.0280
τ^n	0	0	0
τ^b	0	0	0.4653

Values for ρ are 1/3, 2/3 and .9 for, respectively, iid, Markov-low and -high. π is the inflation rate, τ^k is the transference for sector k and μ_i^k / v_i^k is the fraction / value function of people in sector k holding i units of money.

is adding a role for expansionary policies that is different from the usual insurance explanation.

To see this, notice that such inflation rate arises but the distribution of money among the nonbank public experiences relatively small changes. If $\beta = .9$ and the core is on then spending in meetings $(1,1,1)$ and $(1,2,1)$ hit .19 and the nonbank sector fraction holding one unit becomes about 49%. It is the distribution of money in the intermediation sector that experiences a significant change: the fraction of intermediaries without money falls from 14% in the low persistence case to about 3% in the high one. Inflation thus appears with high persistence because money transferred to intermediaries stays in the bank sector for a while, solving in a similar way the externality problem addressed with taxation in section 3.[6]

We also notice that a financial profit exists in some cases in meeting $(1,1,2)$. It occurs when persistence is high. It is followed by improvements in the distribution of money in the nonbank sector. Absent profit outcomes in meeting $(1,1,2)$, lottery realizations would leave either the producer or the consumer with two units of money, excluding this person from some trades next period. Although profits make consumption goods more expensive, when persistence is sufficiently high the positive effect on the distribution of money across traders is dominating.[7]

[6] Transfers directed to intermediaries when persistence is low find a quick inflow into the nonbank sector. Giving money first to intermediaries reduce negative effects on producer constraints not seen in the more static analysis of section 3 due to the quasi-linearity assumption.

[7] Giving profits to intermediaries holding one unit in other meetings would hurt the distribution of nonbank money. Having intermediaries with 2 units is not important: meetings $(0,2,2)$ feature

Table 6 Outside money, $\beta = .5$ and core off

Persistence	iid	Markov low	Markov high
m	y / λ	y / λ	y / λ
(0,1,1)	0.5318 / 1.00	0.6395 / 1.00	0.8788 / 1.00
(0,1,2)	0.5318 / 1.00	0.6395 / 1.00	0.8826 / 1.00
(0,2,1)	0.5318 / 1.00	0.6395 / 1.00	0.8781 / 1.00
(0,2,2)	0.5318 / 1.00	0.6395 / 1.00	0.8826 / 1.00
(1,1,1)	0.0097 / 0.02	0.0112 / 0.02	0.0344 / 0.10
(1,1,2)	0.4233 / 1.00	0.4936 / 1.00	0.1997 / 1.00
(1,2,1)	0.0097 / 0.02	0.0112 / 0.02	0.0344 / 0.10
(1,2,2)	0.4233 / 1.00	0.4936 / 1.00	0.3508 / 1.00
profit (112)	0	0	0.43
μ_0^n / μ_0^b	0.0644 / 0.0644	0.0650 / 0.0650	0.2221 / 0.0233
μ_1^n / μ_1^b	0.7412 / 0.7412	0.7404 / 0.7404	0.5246 / 0.0810
μ_2^n / μ_2^b	0.1944 / 0.1944	0.1946 / 0.1946	0.2533 / 0.8957
v_0^n / v_0^b	0.0000 / 0.0000	0.0000 / 0.0000	0.0016 / 0.0276
v_1^n / v_1^b	0.1718 / 0.0711	0.1953 / 0.0488	0.2617 / 0.0323
v_2^n / v_2^b	0.3195 / 0.1278	0.3451 / 0.0865	0.3577 / 0.0325
π	0	0	0.0458
τ^n	0	0	0
τ^b	0	0	1

Values for ρ are 1/3, 2/3 and .9 for, respectively, iid, Markov-low and -high. π is the inflation rate, τ^k is the transference for sector k and μ_i^k / v_i^k is the fraction / value function of people in sector k holding i units of money.

5.2 Inside-money inflation

In our last set of simulations, we consider specifications displaying no transitions for intermediation occupations ($\rho = 1$). In this case, the planner is not constrained by intermediation incentives. In this case, even without an explicit description of how intermediaries can be monitored, it is reasonable to assume that the planner can ask intermediaries to finance any spending levels. The economy then gains an inside-money interpretation in the spirit of Cavalcanti and Wallace (1999).

In tables 7 and 8, we display results for inside-money economies with two kinds of discount rates and according to two scenarios. The first scenario has the core requirement turned off. One view here is that intermediaries are essential for all conceivable transactions. As a result, the ability to perfectly control them implies that producer and consumers cannot deviate as a group (in a sense, therefore, monitoring is removing the core requirement).

The second scenario leaves the producer-consumer pair with the option of not using the intermediary, and preventing this option from being exercised is in fact a constraint imposed to the planner. That is, although the intermediary is perfectly controlled, it is constrained to do financing only in ways that improve what producer and consumer get by themselves. Although this scenario is in a way a change in

just one unit spent and the same output as $(0, 1, 2)$. Hence profits perform the money destruction feature of inside-money economies discussed later.

the environment (since money can flow freely from consumer to producer off the equilibrium path), it is instructive for checking out how robust our conclusions are.[8]

In these two scenarios, each meeting is fully described by a pair $m = (m_1, m_2)$, where m_1 (m_2) denotes money holdings of the producer (consumer). In tables 7 and 8 we report, for each meeting, output y, relative to y^*, as well as a measure of money transferred to the producer λ. In some meetings, the producer is paid 2 units with positive probability, and hence a reported value $\lambda > 1$ indicates that a two-unit payment has probability $\lambda - 1$, while a single-unit payment has probability $2 - \lambda$. We also indicate, using positive values for b_m, the probability that one unit is created by the intermediary in meeting m. When b_m is negative then $|b_m|$ is the probability that a unit of money of the consumer is destroyed (extracted from the consumer but not transferred to the producer).

We find that in simulations leading to tables 7 and 8 there is no use of transfers to the nonbank sector, and hence there is no need to report τ in these tables.

Table 7 Inside money and core off

β	.5	.6	.7	.8	.9
m	y / λ	y / λ	y / λ	y / λ	y / λ
(0,0)	0.06 / 0.25	0.08 / 0.22	0.10 / 0.22	0.14 / 0.21	0.18 / 0.20
(0,1)	0.25 / 1.00	0.35 / 1.00	0.48 / 1.00	0.66 / 1.00	0.92 / 1.00
(0,2)	0.40 / 1.87	0.53 / 2.00	0.72 / 2.00	0.88 / 1.68	0.98 / 1.16
(1,0)	0.01 / 0.04	0.01 / 0.09	0.02 / 0.10	0.40 / 0.12	0.06 / 0.15
(1,1)	0.10 / 0.59	0.18 / 1.00	0.24 / 1.00	0.32 / 1.00	0.40 / 0.97
(1,2)	0.18 / 1.00	0.18 / 1.00	0.24 / 1.00	0.32 / 1.00	0.41 / 1.00
b_{00}	0.25	0.22	0.22	0.21	0.20
b_{01}	0.00	1.00	1.00	1.00	1.00
b_{02}	0.87	1.00	1.00	0.68	0.16
b_{10}	0.04	0.09	0.10	0.12	0.15
b_{11}	-0.41	1.00	1.00	1.00	0.97
b_{12}	0.00	0.00	0.00	0.00	0.00
μ_0	0.7262	0.6424	0.6430	0.6428	0.6412
μ_1	0.2373	0.3156	0.3142	0.3150	0.3182
μ_2	0.0364	0.0421	0.0427	0.0422	0.0406
v_0	0.2295	0.2868	0.3982	0.6435	1.3914
v_1	0.2692	0.3518	0.4989	0.7678	1.5571
v_2	0.2996	0.3932	0.5234	0.8014	1.5945
π	0.1854	0.3537	0.3543	0.3505	0.3445

π is the inflation rate, μ_i is the fraction of people in nonbank sector holding i units of money, b_m is the probability that one unit of money be created in meeting m and v_i is the value function of people holding i units of money.

In table 7 we find that inflationary policies are implemented in all configurations. Since there is the option to create credit with perfect control according to its social impact, the concern with distributions of holdings is less important in comparison with the outside-money case. The mass of nonbank people with two units is reduced by inflation. The high magnitude of inflation, of 35%, is necessary to remove money created in credit operations.

[8] While we thank Neil Wallace for suggesting examination of this second scenario, we did not find previous work discussing how monitoring of a subset of traders can change the set of core allocations.

Table 8 Inside money and core on

β	.5	.6	.7	.8	.9
m	y / λ	y / λ	y / λ	y / λ	y / λ
(0,0)	0.05 / 0.28	0.06 / 0.19	0.09 / 0.22	0.12 / 0.22	0.16 / 0.16
(0,1)	0.16 / 1.00	0.32 / 1.00	0.39 / 1.00	0.54 / 1.00	1.00 / 1.00
(0,2)	0.23 / 2.00	0.35 / 1.22	0.51 / 1.49	0.59 / 1.15	1.00 / 1.00
(1,0)	0.00 / 0.05	0.01 / 0.05	0.01 / 0.06	0.02 / 0.07	0.05 / 0.13
(1,1)	0.07 / 1.00	0.14 / 1.00	0.23 / 1.00	0.32 / 1.00	0.34 / 1.00
(1,2)	0.07 / 1.00	0.14 / 1.00	0.23 / 1.00	0.32 / 1.00	0.38 / 1.00
$b_{0,0}$	0.28	0.19	0.22	0.22	0.16
b_{01}	0.00	0.00	0.00	0.00	1.00
b_{02}	0.00	0.22	0.49	0.15	0.00
b_{10}	0.05	0.05	0.06	0.07	0.13
b_{11}	0.00	0.00	0.00	0.00	0.00
b_{12}	0.00	0.00	0.00	0.00	0.00
μ_0	0.7752	0.7467	0.7395	0.7286	0.6406
μ_1	0.1954	0.2159	0.2198	0.2290	0.3184
μ_2	0.0294	0.0374	0.0407	0.0424	0.0410
v_0	0.2146	0.2821	0.4059	0.6508	1.375
v_1	0.2673	0.3633	0.4931	0.7545	1.5714
v_2	0.2846	0.3921	0.5376	0.8091	1.6073
π	0.1892	0.1210	0.1387	0.1235	0.2433

π is the inflation rate, μ_i is the fraction of people in nonbank sector holding i units of money, b_m is the probability that one unit of money be created in meeting m and v_i is the value function of people holding i units of money.

In table 8, monetary policy becomes less expansionary. When the producer-consumer pair can deviate as a group, spending increases and generates larger distortions on extensive margins, forcing the planner to create less credit. As a result a lower inflation rate emerges. We find that inside-money inflation is associated to more efficient insurance overall, compensating for negative velocity effects. Computed increases in consumption are in line with simulations reported by Deviatov and Wallace (2014) for inside-money economies with pairwise meetings and 0-1 holdings of money.

5.3 Zero inflation with pairwise meetings

We have already anticipated some basic results of economies without intermediation (pairwise meetings) in section 2. We have included in the appendix a test case when the upper bound is 2 units. There is no optimal inflation with the core off, but expansionary policies are welfare improving when the core is on and the discount factor is low, so that in meeting type $(1, 1)$ consumers are spending all their holdings (output is 12% of first-best level or less).

Tables 9 and 10 show the effects of increasing the upper bound of outside money, from 2 to 4. Expansionary policies now are never optimal, regardless of β or the core requirement. In addition to taxes discussed in section 2, when consumers meet the poorest producer, we also find taxation in meeting $(1, 3)$, when the producer has one unit and the consumer has three, but only for very high β (of .9). So the main lesson is that with more divisibility of money traders are more conservative in terms of spending money and consumer taxes prove to be more efficient in terms of providing insurance.

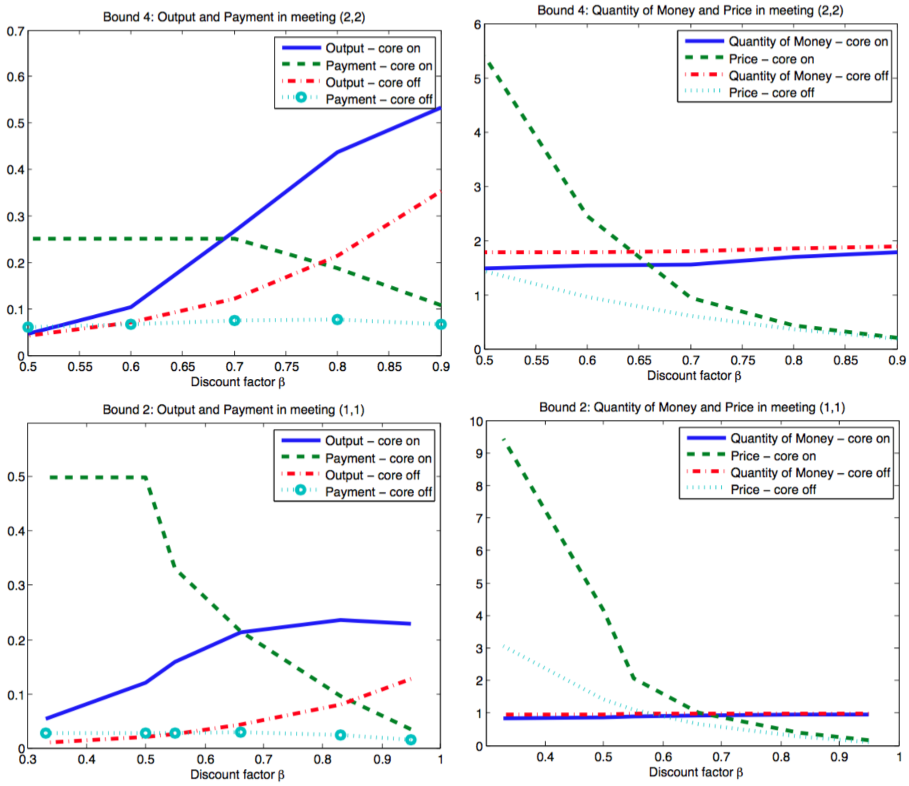

Fig. 1 Salient features of pairwise trades

Some basic effects of changes in the upper bound are displayed in Figure 1. The top two panels refer to the 4 bound, while the bottom two refer to the 2 bound. On the left we can see the effects of β on output and average payment in meeting $(2, 2)$ with the 4 bound (top panel), and in meeting $(1, 1)$ with the 2 bound (bottom panel). On the right, curves represent now the mean of the distribution of holdings and the ratio between average payment and output (called 'price') for the 4 bound (top panel) and the 2 bound (bottom panel).[9]

Meetings $(1, 1)$ and $(2, 2)$ are important meetings in terms of their impacts on the distribution of money for, respectively, bounds 2 and 4. Curves for output, representing consumption relative to first-best levels, indicate that the economy with 4 units has more trade going on. Curves for payments show that people save more for a large set of preferences with the 4 bound, an indication that the Inada condition has produced steeper value functions around 0 holdings. In this sense, money becomes more valuable (output doubles at high β).

With the 2 bound payments hit the ceiling of consumers' holdings in meeting $(1, 1)$ at low β such that positive inflation is optimal. Core effects are important for low β, confirming what has been remarked above. When the core is off, output is smaller but both the meeting's payment and the quantity of money vary little with β.

[9] Payment statistics used in Figure 1 are defined as the average transfer of money in meeting $(2, 2)$ (for bound 4) or $(1, 1)$ (for bound 2), paid by consumers to producers, divided by the corresponding bound.

Table 9 Outside money in pairwise meetings: core on

β	0.9	0.8	0.7	0.6	0.5
m	y / $\lambda(x)$	y / $\lambda(x)$	y / $\lambda(x)$	y / $\lambda(x)$	y / $\lambda(x)$
(0,1)	1.0007 / 0.18 (1)	1.0007 / 0.29 (1)	1.0067 / 0.50 (1)	1.0000 / 0.78 (1)	0.8377 / 1.00 (1)
(0,2)	3.1900 / 1.00 (1)	3.3972 / 1.00 (1)	2.0209 / 1.00 (1)	1.2184 / 1.00 (1)	0.7794 / 1.00 (1)
(0,3)	4.2274 / 1.00 (1)	3.3972 / 1.00 (1)	2.0209 / 1.00 (1)	1.2850 / 0.90 (2)	1.0000 / 0.90 (2)
(0,4)	5.1675 / 1.00 (1)	2.2917 / 0.49 (2)	1.8818 / 1.00 (2)	1.6358 / 1.00 (2)	1.0194 / 1.00 (2)
(1,1)	0.2947 / 0.14 (1)	0.2521 / 0.22 (1)	0.2364 / 0.37 (1)	0.1967 / 0.56 (1)	0.1473 / 0.81 (1)
(1,2)	1.0000 / 0.49 (1)	1.0000 / 0.89 (1)	0.6372 / 1.00 (1)	0.3500 / 1.00 (1)	0.1810 / 1.00 (1)
(1,3)	1.8766 / 1.00 (1)	1.1272 / 1.00 (1)	0.6372 / 1.00 (1)	0.3500 / 1.00 (1)	0.1810 / 1.00 (1)
(1,4)	2.0553 / 1.00 (1)	1.1272 / 1.00 (1)	0.9035 / 1.00 (2)	0.4533 / 1.00 (2)	0.2274 / 1.00 (2)
(2,1)	0.1571 / 0.13 (1)	0.1099 / 0.19 (1)	0.0793 / 0.30 (1)	0.0426 / 0.41 (1)	0.0269 / 0.58 (1)
(2,2)	0.5333 / 0.43 (1)	0.4368 / 0.75 (1)	0.2663 / 1.00 (1)	0.1025 / 1.00 (1)	0.0464 / 1.00 (1)
(2,3)	0.5333 / 0.43 (1)	0.4368 / 0.75 (1)	0.2663 / 1.00 (1)	0.1025 / 1.00 (1)	0.0464 / 1.00 (1)
(2,4)	1.2423 / 1.00 (1)	0.5804 / 1.00 (1)	0.2663 / 1.00 (1)	0.1990 / 1.00 (2)	0.1017 / 1.00 (2)
(3,1)	0.1197 / 0.12 (1)	0.0479 / 0.16 (1)	0.0209 / 0.23 (1)	0.0396 / 0.41 (1)	0.0337 / 0.61 (1)
(3,2)	0.4076 / 0.41 (1)	0.1892 / 0.64 (1)	0.0898 / 0.97 (1)	0.0965 / 1.00 (1)	0.0553 / 1.00 (1)
(3,3)	0.4076 / 0.41 (1)	0.1892 / 0.64 (1)	0.0927 / 1.00 (1)	0.0576 / 0.59 (1)	0.0509 / 0.92 (1)
(3,4)	1.0007 / 1.00 (1)	0.2969 / 1.00 (1)	0.0927 / 1.00 (1)	0.0965 / 1.00 (1)	0.0553 / 1.00 (1)
μ_0	0.0720	0.1171	0.1862	0.2263	0.2695
μ_1	0.3572	0.3708	0.3584	0.3164	0.2801
μ_2	0.3437	0.2806	0.2450	0.2459	0.2415
μ_3	0.1712	0.1605	0.1337	0.1163	0.1083
μ_4	0.0559	0.0710	0.0767	0.0951	0.1006
v_0	0.4514	0.0263	0.0089	0.0018	0.0013
v_1	1.2638	0.5940	0.3948	0.2282	0.2253
v_2	1.5692	0.7824	0.5166	0.3663	0.2738
v_3	1.7538	0.8794	0.5675	0.3892	0.2861
v_4	1.9027	0.9291	0.5851	0.4107	0.3010

$\lambda(x)$ is the optimal probability of transfering x units of money (and paying $x-1$ units with probability $1-\lambda(x)$).

It is fair to say that the curve changing the most with the bound is the one reflecting optimal payments in these ‘critical’ meetings. Based on Figure 1, a reasonable conjecture is that output and payments vary less with the discount factor as the bound is increased further, indicating that more insurance is provided. This perhaps can explain that the quantity of money less than doubles as the bound is increased from 2 to 4: as the insurance problem is alleviated, the planner reduces the quantity of money, relatively, in order to improve the return of money (and the intensive margin of consumption).[10]

We also notice that the value function v, if linearly interpolated, does not become concave with the 4 bound when $\beta = .5$. This is interesting because the assumption of no taxation, when all surpluses are given to consumers, has the feature of generating steady states with concave value functions, which can be attractive for numerical methods with larger state spaces. Unfortunately we are now learning that the optimal allocation of exchange risk can violate concavity.

Confirming what we have seen with intermediation, the core requirement has a strong effect on exchange risk. This is quite evident in the relationship between β

[10] Explaining why turning the core off leads to a higher quantity of money and taxation of poor consumers is more challenging. Focusing on poor consumers instead of richer ones helps with the return of money, which perhaps needs more attention when the quantity of money is increased. Tables 9 and 10 point to other properties that may require investigation of larger bounds, beyond our current capabilities.

Table 10 Outside money in pairwise meetings: core off

β	0.9	0.8	0.7	0.6	0.5
m	y / $\lambda(x)$	y / $\lambda(x)$	y / $\lambda(x)$	y / $\lambda(x)$	y / $\lambda(x)$
(0,1)	0.8856 / 1.00 (1)	0.7345 / 1.00 (1)	0.5969 / 1.00 (1)	0.4981 / 1.00 (1)	0.4084 / 1.00 (1)
(0,2)	1.1316 / 1.00 (1)	1.1668 / 1.00 (1)	1.0299 / 1.00 (1)	0.8773 / 1.00 (1)	0.6305 / 1.00 (1)
(0,3)	2.5093 / 1.00 (1)	2.2064 / 1.00 (1)	1.4570 / 1.00 (1)	0.9895 / 0.04 (2)	0.9058 / 1.00 (2)
(0,4)	3.2446 / 1.00 (1)	2.2745 / 0.07 (2)	2.1144 / 1.00 (2)	1.4009 / 1.00 (2)	0.9058 / 1.00 (2)
(1,1)	0.2266 / 0.12 (1)	0.1249 / 0.12 (1)	0.0718 / 0.11 (1)	0.0419 / 0.10 (1)	0.0254 / 0.09 (1)
(1,2)	1.0007 / 0.55 (1)	1.0000 / 0.97 (1)	0.6574 / 1.00 (1)	0.4293 / 1.00 (1)	0.2745 / 1.00 (1)
(1,3)	1.8325 / 1.00 (1)	1.0337 / 1.00 (1)	0.6574 / 1.00 (1)	0.4293 / 1.00 (1)	0.2745 / 1.00 (1)
(1,4)	1.8325 / 1.00 (1)	1.0337 / 1.00 (1)	1.0000 / 0.84 (2)	0.6904 / 1.00 (2)	0.4473 / 1.00 (2)
(2,1)	0.1137 / 0.09 (1)	0.0516 / 0.07 (1)	0.0254 / 0.06 (1)	0.0135 / 0.05 (1)	0.0082 / 0.05 (1)
(2,2)	0.3530 / 0.27 (1)	0.2132 / 0.31 (1)	0.1227 / 0.30 (1)	0.0711 / 0.27 (1)	0.0419 / 0.24 (1)
(2,3)	1.0000 / 0.76 (1)	0.6941 / 1.00 (1)	0.4099 / 1.00 (1)	0.2610 / 1.00 (1)	0.1728 / 1.00 (1)
(2,4)	1.3126 / 1.00 (1)	0.6941 / 1.00 (1)	0.4099 / 1.00 (1)	0.2610 / 1.00 (1)	0.1728 / 1.00 (1)
(3,1)	0.0464 / 0.05 (1)	0.0157 / 0.04 (1)	0.0060 / 0.03 (1)	0.0030 / 0.02 (1)	0.0015 / 0.02 (1)
(3,2)	0.0965 / 0.09 (1)	0.0396 / 0.09 (1)	0.0172 / 0.08 (1)	0.0082 / 0.07 (1)	0.0045 / 0.06 (1)
(3,3)	0.2378 / 0.23 (1)	0.1369 / 0.31 (1)	0.0868 / 0.41 (1)	0.0606 / 0.50 (1)	0.0411 / 0.57 (1)
(3,4)	1.0000 / 0.98 (1)	0.4361 / 1.00 (1)	0.2132 / 1.00 (1)	0.1212 / 1.00 (1)	0.0711 / 1.00 (1)
μ_0	0.0361	0.0377	0.0365	0.0335	0.0307
μ_1	0.2906	0.3165	0.3366	0.3443	0.3433
μ_2	0.4432	0.4301	0.4341	0.4455	0.4580
μ_3	0.2030	0.1893	0.1692	0.1558	0.1471
μ_4	0.0271	0.0264	0.0236	0.0209	0.0209
v_0	0.7894	0.2035	0.0706	0.0229	0.0068
v_1	1.2715	0.5723	0.3489	0.2394	0.1754
v_2	1.5437	0.0745	0.4745	0.3351	0.2489
v_3	1.7387	0.8610	0.5527	0.3932	0.2951
v_4	1.8896	0.9339	0.5934	0.4201	0.3141

$\lambda(x)$ is the optimal probability of transfering x units of money (and paying $x-1$ units with probability $1-\lambda(x)$).

and the optimal distribution of money across tables. When the core is off (Table 10) the distributions do not change much with β and display a bell shape. When the core is on (Table 9), by contrast, reductions in β (and thus in saving rates) have a remarkable effect on the mass of people with low holdings of money, up to the point that the distribution approaches the uniform case. This is consistent with a strong velocity effect that makes expansionary policies suboptimal. In addition, consumers in meeting $(2,4)$ spend more than one unit of money when β is .5 or .6 and the core is on, but less than one unit when the core is off (which reduces the dispersion of money).

6 More on previous work

We have seen that the velocity effect is very strong in matching models. With a sufficiently large support for holdings, consumption taxes impose less distortions on savings and are more efficient than expansionary policies. Inflation appears with a smaller support when consumers are in a corner, making large payments or when the model approaches an inside-money specification. By contrast, meeting-specific ‘prices’ are not possible in a market setting or when a bargain rule is assumed. Such extreme cases are common, for reasons of tractability of state spaces, producing a biased assessment of expansionary policies.

The major part of the money literature (too large to be reviewed here) has focused on reductions in the return of money. We think that random-matching models requires a broader perspective. An useful fiction is to think that the distribution of money is some sort of intangible capital related to the allocation of exchange risk, and that a high velocity of money subtracts from this capital: inflation produces a change in behavior associated with more dispersed money holdings, and less trade is accomplished when money is distributed poorly. That is why giving all surpluses to consumers may not be a good idea, and monetary policy should target the bank sector. By contrast, there are no dynamic effects on savings in the model of Levine (1991).[11] In that model, the analysis of expansionary policies is simpler since a parameter determines the likelihood that half of the population is without money and facing a relatively high utility of consumption. The ratio of marginal utilities between the two groups is yet another parameter closely connected to the effectiveness of expansionary policies, which can also be specified exogenously.[12]

Cavalcanti and Nosal (2009) have identified another simple way of making flat expansions appealing. They consider random meetings with a seasonal pattern, specifying a utility jump for people specializing in consumption at a particular season and finding welfare gains when expansions target that season. But negative effects of inflation in their setting are also narrowly defined. Intermediation makes expansionary policies more powerful, without the seasonal component of Cavalcanti and Nosal (2009). Since money transferred to nonbank people is spent faster, policies are more effective when they target the bank sector.

Consumers do not take into account that by making large payments they reduce incentives of other traders to produce in the future, depleting the social capital mentioned above. This externality is confirmed by another experiment which consists of turning on and off the core requirement, revealing the importance of exchange risk and the corresponding need to avoid high distortions on savings.

The Friedman (1953, 1969) *rule* is the classic proposition that deflation is optimal. Lucas (1980) and Scheinkman and Weiss (1986) are early references on how incomplete markets give rise to a more sophisticated role for monetary policy. Against the background of a cash-in-advance model, the former notices: “The problem here is not one involving the attractiveness of currency on average, but one of permitting the benefits of gains from trade between differently situated agents.” Aside from literature on incomplete markets and exchange frictions, it is useful to associate the kind of externalities discussed above with introductory passages in Bagehot (1873). In his classic description of money markets and the operation of central banking 150 years ago, the distribution of liquidity plays a central role.

> “Everyone is aware that England is the greatest moneyed country in the world; everyone admits that it has much more immediately disposable and ready cash than any other country. But very few persons are aware how much greater the ready balance—the floating loan-fund which can be lent to anyone or for any purpose—is in England than it is anywhere else in the world. [...]

[11] See also Levine (2015) for a related but nontechnical discussion of Keynesian ‘philosophy’.

[12] Kehoe et al. (1992) shows that positive inflation can still be optimal in Levine (1991) model even if the first best is not attained, while Wallace (2014) proposes more general transfers, finding gains from both regressive and progressive transfers over some no-inflation equilibria. Exceptions noting distribution effects include Imrohoroglu (1992) and Molico (2006): consumption volatility increasing with inflation in a Bewley (1980) model and, respectively, wealth and prices becoming more dispersed with search when consumers extract all surpluses.

> Of course the deposits of bankers are not a strictly accurate measure of the resources of a Money Market. On the contrary, much more cash exists out of banks in France and Germany, and in all non-banking countries, than could be found in England or Scotland, where banking is developed. But that cash is not, so to speak, 'moneymarket money': it is not attainable. [...] But the English money is 'borrowable' money. [...] Concentration of money in banks, though not the sole cause, is the principal cause which has made the Money Market of England so exceedingly rich, so much beyond that of other countries."

This is a simple argument that money needs to be in the right hands to exercise its full potential. We would add that there is also a flip side to this basic concept: when money is spent then this potential is lost, unless some special mechanism is in place to make liquidity flow back to its origin. Bagehot finds that the financial sector has a delicate function, considered fragile to some extent.[13]

Modern wisdom on how changes in the quantity of money affect prices and output are not based on externality arguments. But as emphasized by Bagehot, the distribution of money is important for the availability of credit. We have shown that if a social planner controls private spending well enough then in many cases there are no gains in pursuing liquidity injections. In limit cases with full monitoring of intermediaries, inflation is nonetheless justified to finance intermediation when inside money can be created more efficiently.

7 Conclusion

Our critique applies to a large literature adopting somewhat arbitrary specifications of exchange risk, trade protocols or even quasi-linear preferences for dealing with that risk. Hence there is a delicate matter of deciding what level of exchange risk should be considered, since in simpler models policies can focus on the value of money. While in a cash-in-advance model (without exchange risk) inflation produces essentially a hot-potato effect largely discussed, the velocity effect we are emphasizing works as a multiplier: large money payments make recipients less inclined to produce in the future, amplifying the hot-potato effect.

Dating back to the classic signal-extraction model, simulations (see Wallace (1992)) already showed a rich pattern of savings responses to alternative distributions of nominal and real shocks, making it a good example of how delicate is the job of predicting effects of expansionary policies, in line with the Lucas (1976) critique. Since then, economists have adopted assumptions of low exchange risk or linear preferences not because they like them, but due to tractability. Without such special assumptions, matching models are attractive insofar risks associated with trade are specified in a natural way.

We have seen that with the core requirement removed the optimum would have people spending less, and that this property is robust to the addition of intermediation. This tell us that exchange risk is highly important in random-matching models, so

[13] We have modeled this tension in a rudimentary way: intermediaries provide an important externality, but it is natural for the system to bid down compensation for this service. Bagehot in fact witnessed a gold standard facing episodes of adverse liquidity shocks. His well-known policy recommendation of high interest rates in moments of bank panic, coupled with broad measures of liquidity provision by discounting of bills, are geared at discouraging people from withdrawing funds from the financial sector.

that the velocity effect cannot be ignored. But even if this kind of risk is considered out of proportion for applied work, velocity effects are nevertheless important for understanding inside money. By modeling intermediation in a rudimentary way, we think we have also captured the “ready balance” property emphasized by Bagehot, a fundamental externality likely relevant in modern times.

APPENDIX

A Proofs of propositions

Proposition 1 For m in the support of distributions of meetings, output is $\hat{y}(m) = m_3$ when intermediation is relaxed, and $\tilde{y}(m) = \min\{m_2, m_3\}$ when savings need not be incentive compatible. In these relaxed problems, moreover, welfare satisfies $w(\hat{s}, \hat{y}) \geq w(\tilde{s}, \tilde{y}) \geq w(s^*, y^*)$, with inequalities replaced by equalities when there is a single type of trader.

Proof That optimal welfare $w(s^*, y^*)$ is bounded above by $w(\tilde{s}, \tilde{y})$ is trivial. Consider now optimal allocations with pairwise trades described by C&P. They show that consumers should spend all their holdings and keep all surplus from trade. Such allocations are implementable in our setting when $x(m) \leq m_2$ is relaxed and intermediaries can make loans matching holdings of money by consumers, without profits. Since profits do not affect aggregate welfare (1), we conclude that such allocations solve the relaxed problem with $\hat{y}(m) = m_3$. C&P also show that in their economy, since holdings of money by producers do not matter for output, then individuals do not care about the distribution of money when making saving choices: if they become producers their surplus is zero, and if they become consumers they do not care whether they meet with rich or poor producers. This means that savings are decided on the basis of realizations of idiosyncratic shocks and on output obtained when consumers, without consideration of how others are choosing money holdings. As a result, the distribution of money can be computed residually, after an incentive-compatible savings function associated to $\hat{y}$ is found. This proves the inequality $w(\tilde{s}, \tilde{y}) \leq w(\hat{s}, \hat{y})$. More generally, C&P show that the pairwise optimum solves a relaxed problem with no wedge between private and social savings. Although allocation $(\tilde{s}, \tilde{x}, \tilde{y}, \tilde{z})$ is obtained by ignoring private incentives for saving, these ignored constraints do not bind in the associated pairwise problem. In addition, if the support of λ is a singleton then in all meetings occurring with positive probability consumers and intermediaries have exactly the same money holdings. In this case, the cash-in-advance requirement $x(m) \leq m_2$ is irrelevant and $w(s^*, y^*) = w(\tilde{s}, \tilde{y}) = w(\hat{s}, \hat{y})$ must hold. Finally, $\tilde{y}(m) = \min\{m_2, m_3\}$ follows from an application of the upper-bound construction of C&P: if saving incentives can be ignored, the arrangement that maximizes $w(s, \cdot)$ for a given savings function has all trade surplus going to consumers, and has consumers spending all their holdings up to the bound dictated by intermediation. □

Proposition 2 When there is more than one type of trader, welfare is increasing in the profit rate r in a neighborhood of zero, so that it is not optimal to give all surpluses to consumers.

Proof For sufficiently small r, incentive-compatible savings s_i for type i satisfies

$$\theta_i - \beta = \alpha\beta[u'(s_i) - 1]\frac{i}{n} + \alpha\beta r\frac{i-1}{n}.$$

The term $u'(s_i) - 1$ reflects the utility gain of consumption from bringing an extra unit of money when the consumer receives all the surplus from trade, net of the expected opportunity cost of carrying money, which has been assumed equal to the unity. The expression $\theta_i - \beta = \alpha g(s_i)$, for $g(k) = \beta(u'(k) - 1)$, was obtained by C&P. With intermediation, $\alpha g(s_i)$ must be adjusted by the probability that the consumer is paired with an intermediary carrying at least the same quantity of money, $\frac{i}{n}$ (otherwise there is no marginal effect of an extra unit saved). The term $\alpha r\frac{i-1}{n}$ represents the marginal, expected payment of profits from richer consumers. Now the derivative of the welfare function with respect to r, w', satisfies

$$w' = \sum_{i=1}^{n}\left[-\frac{\theta_i - \beta}{n}s_i' + \frac{\alpha}{n^2}\left(\sum_{j\geq i} g(s_j)s_j' + \sum_{j<i} g(s_i)s_i'\right)\right]$$

where s_i' denotes the derivative of s_i with respect to r. Using now that for $r = 0$

$$g(s_j) = \frac{(\theta_j - \beta)n}{\alpha j}$$

then

$$w'\big|_{r=0} = \sum_{i=1}^{n}\left[-\frac{\theta_i - \beta}{n}s_i' + \frac{\alpha}{n^2}\left(g(s_i)s_i' + \sum_{j>i} g(s_j)s_j' + \sum_{j<i} g_i(s_i)s_i'\right)\right] =$$
$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j>i}\frac{\theta_j - \beta}{j}s_j' + \sum_{j<i}\frac{\theta_i - \beta}{i}s_i'\right).$$

Now, since u is concave then s_i' is positive if $i > 1$ and r is sufficiently small. Therefore, under the assumption that $n > 1$, w' is positive for such r. □

B A test case

In this appendix we present facts of economies without intermediation, studied by Deviatov (2006). Reproducing his table is useful as a test case (see more on our numerical approach in the second part of the appendix), and for showing that inflationary interventions, in such economies, are only used to counterbalance high spending in corner situations. In order to reproduce his tables we need to change the timing of monetary policy in the model presented above. In Deviatov (2006) money transfers occur first, followed by the inflationary process that keeps the quantity of money constant. We should remark, by the way, that we tried other configurations to make sure that described relationship between inflation and corner outcomes it robust to timing specifications (there are small changes in allocations overall).

In order to produce the first table below, we keep Deviatov's specification intact, which means we can use the same utility function used above to study intermediation. It turns out that in his economy the consumer never spends more than one unit of money. Hence we can just report λ_{ij}, defined as the probability that one unit is transferred in meeting (i, j) — henceforth a meeting in which the producer has i and the consumer has j units of money — in addition to reporting output relative to first-best output y^*.

Table 11 Pairwise meetings and core on

β	.95	.83	.66	.55	.50	.33
y_{01}	1.0000*	1.0000*	1.0000*	1.0000*	0.4353*	0.2109*
y_{02}	3.7614	2.4951	1.7038*	1.0486*	0.4353*	0.2109*
y_{11}	0.2296*	0.2356*	0.2124*	0.1593*	0.1204*	0.0531*
y_{12}	1.0000*	1.0000*	0.4928*	0.2415*	0.1204*	0.0531*
λ_{01}	0.0915†	0.2565†	0.5872†	0.9538†	1	1
λ_{02}	1	1	1	1	1	1
λ_{11}	0.0682†	0.1921†	0.4306†	0.6607†	1†	1†
λ_{12}	0.2972†	0.8155†	1	1	1	1
μ_0	0.1932	0.2600	0.3232	0.3679	0.3686	0.3810
μ_1	0.6553	0.5314	0.4302	0.3769	0.3970	0.4016
v_0	1.0169	0.0783	0.0000	0.0000	0.0217	0.0088
v_1	2.5514	0.7040	0.3417	0.2524	0.1890	0.1348
v_1	3.0240	0.9008	0.4405	0.3105	0.2174	0.1476
Inflation	0	0	0	0	0.1763	0.2018
Transfers	0	0	0	0	0.2498	0.2795

* Producer's incentive constraint is binding.
† The core constraints is bimding.

In table 11 we can notice two effects taking place as the discount factor falls: money spent in meeting $(1, 1)$ increases and, consequently, holdings are scattered as the set of people holding one unit loses mass (μ_1 falls). This is relevant because people holding one unit can be both producers and consumers. When discount factors β are very low, expansionary policies are needed, but these are also corner cases in which velocity effects are absent. To see that interventions would not be necessary if spending is sufficiently controlled we turn off the core requirement preventing group defections, as in table 12.

Table 12 Pairwise meetings and core off

β	.95	.83	.66	.55	.50	.33
y_{01}	0.9634	0.9132	0.8751	0.8265	0.7449*	0.3328*
y_{02}	3.6649*	2.3717*	1.3552*	0.9237*	0.7449*	0.3328*
y_{11}	0.1279*	0.0792*	0.0426*	0.0254*	0.0194*	0.0089*
y_{12}	1.0000*	1.0000*	0.7501*	0.4577*	0.3567*	0.1623*
λ_{01}	1	1	1	1	1	1
λ_{02}	1	1	1	1	1	1
λ_{11}	0.0282	0.0458	0.0563	0.0557	0.0551	0.0545
λ_{12}	0.2212	0.5778	1	1	1	1
μ_0	0.1299	0.1619	0.1738	0.1774	0.1800	0.1882
μ_1	0.7484	0.6997	0.6774	0.6782	0.6789	0.6785
v_0	1.9821	0.2728	0.0435	0.0067	0.0000	0.0000
v_1	2.4065	0.6532	0.3154	0.2290	0.1992	0.1336
v_1	3.0414	0.9310	0.4658	0.3392	0.2946	0.1988
Inflation	0	0	0	0	0	0
Transfer	0	0	0	0	0	0

* Producer's incentive constraint is binding.

When the core is off, the planner manages to implement a better distribution of money. In this case, money spend in meeting $(1, 1)$ are low even when a low β tightens up producer constraints and reduces output.

C Auxiliary objects and numerical approach

We describe below in more detail objects used in our simulations. The probability distribution of after-trade holdings, λ, is in fact a 3×3 matrix, $\lambda(m) = (\lambda_1(m); \lambda_2(m); \lambda_3(m))$. In particular, $\lambda_i(m) = (\lambda_i^0(m), \lambda_i^1(m), \lambda_i^2(m))$ is a line vector for $i = 1, 2, 3$, where $\lambda_i^j(m)$ denotes the (marginal) probability that 'person i' (the person starting with m_i) leaves the meeting holding $j \in \{0, 1, 2\}$ units of money. For example, $\lambda_1^j(m)$ denotes the probability that the producer leaves the meeting holding j units of money.

We have six possible states where people are transiting. The state space can be written as $\{n, b\} \times \{0, 1, 2\} = \{(n, 0), (n, 1) \ldots , (b, 1), (b, 2)\}$. In this context, therefore, we have to pile up some three-dimensional objects. The value function can be written in vector notation as $v = (v_0^n, v_1^n, v_2^n, v_0^b, v_1^b, v_2^b)'$. For this configuration of states, monetary policy implies two transition matrices. The inflation matrix P is

$$P = \begin{bmatrix} \Pi & \mathbf{0}_3 \\ \mathbf{0}_3 & \Pi \end{bmatrix}, \tag{9}$$

where

$$\Pi = \begin{pmatrix} 1 & 0 & 0 \\ \pi & 1-\pi & 0 \\ \pi^2 & 2\pi(1-\pi) & (1-\pi)^2 \end{pmatrix} \tag{10}$$

and $\mathbf{0}_3$ is a 3×3 matrix of zeros. Money transfers imply the following matrix

$$T = \begin{bmatrix} \Psi^n & \mathbf{0}_3 \\ \mathbf{0}_3 & \Psi^b \end{bmatrix} \tag{11}$$

where

$$\Psi^k = \begin{pmatrix} 1-\tau^k & \tau^k & 0 \\ 0 & 1-\tau^k & \tau^k \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{for } k \in \{b, n\}. \tag{12}$$

For occupation shocks, let

$$\Lambda = \begin{pmatrix} \frac{1+\rho}{2} & \frac{1-\rho}{2} \\ 1-\rho & \rho \end{pmatrix}. \tag{13}$$

The transition matrix generated by occupation shocks can be written as $S = \Lambda \otimes \mathbf{I}_3$, where $\mathbf{I}_3$ is the 3×3 identity matrix and $\otimes$ represents the Kronecker product.

Now, let $\mathcal{I}^n = [\mathbf{I}_3, \mathbf{0}_3]$ and $\mathcal{I}^b = [\mathbf{0}_3, \mathbf{I}_3]$, then we can write the matrices A^n and A^b describe in the text as:

$$A^n = \mathcal{I}^n PTS \tag{14}$$
$$A^b = \mathcal{I}^b PTS. \tag{15}$$

Finally, let $\mathbf{e}_k$ be a canonical vector in direction k of $\mathbb{R}^3$, $\mathbf{f}_i^n = [\mathbf{e}_{i+1}, (0,0,0)]$ and $\mathbf{f}_i^b = [(0,0,0), \mathbf{e}_{i+1}]$. Then,

$$A_{0i}^n = \mathbf{f}_i^n PTS \tag{16}$$
$$A_{0i}^b = \mathbf{f}_i^b PTS. \tag{17}$$

If we denote by $\sigma^m(i,j)$ the joint probability that after-meeting holdings of the producer-consumer pair is precisely (i,j) then

$$\lambda_1^i(m) = \sum_j \sigma^m(i,j) \tag{18}$$
$$\lambda_3^j(m) = \sum_i \sigma^m(i,j) \tag{19}$$
$$\lambda_2^k(m) = \sum_{(i,j):i+j+k=\bar{m}} \sigma^m(i,j) \tag{20}$$

where $\bar{m} = \sum_\ell m_\ell$.

Now we rewrite participation constraints as

$$\Pi_1(m) = -y(m) + \beta \sum_i \lambda_1^i(m)(\mathbf{f}_j^n - \mathbf{f}_{m_1}^n)PTSv \geq 0 \tag{21}$$
$$\Pi_2(m) = \beta \sum_k \lambda_2^j(m)(\mathbf{f}_k^b - \mathbf{f}_{m_2}^b)PTSv \geq 0 \tag{22}$$
$$\Pi_3(m) = u(y(m)) + \beta \sum_j \lambda_3^j(m)(\mathbf{f}_j^n - \mathbf{f}_{m_3}^n)PTSv \geq 0 \tag{23}$$

and, using σ,

$$\Pi_1(m) = -y(m) + \beta \sum_i \sum_j \sigma^m(i,j)(\mathbf{f}_j^n - \mathbf{f}_{m_1}^n)PTSv \geq 0 \tag{24}$$
$$\Pi_2(m) = \beta \sum_k \sum_{(i,j):i+j+k=\bar{m}} \sigma^m(i,j)(\mathbf{f}_k^b - \mathbf{f}_{m_2}^b)PTSv \geq 0 \tag{25}$$
$$\Pi_3(m) = u(y(m)) + \beta \sum_j \sum_i \sigma^m(i,j)(\mathbf{f}_j^n - \mathbf{f}_{m_3}^n)PTSv \geq 0. \tag{26}$$

As a result, the problem defining the core in meeting m is

$$\underset{y(m),\sigma^m(i,j)}{\text{Max}} \; \Pi_3(m)$$
$$\text{s.t. } \Pi_1(m) \geq \gamma_1(m) \text{ and } \Pi_2(m) \geq \gamma_2(m).$$

for some (meeting-specific) $\gamma_1(m)$ and $\gamma_2(m)$ consistent with participation constraints.

Let $\zeta_1(m)$ and $\zeta_2(m)$ be Lagrange multipliers associated the restrictions in the problem above. It is easy to see that $\zeta_1(m) = u'(y(m))$. Also, let $L_{ij}(m)$ denote the derivative of the Lagrangian with respect to $\sigma^m(i,j)$. As a consequence of the linearity of Π's in σ's, the solution must satisfy

$$\sigma^m(i,j)\left(\operatorname*{Max}_{i',j'} L_{i'j'}(m) - L_{ij}(m)\right) = 0, \qquad \forall i,j. \tag{27}$$

The numerical problem is to find allocations maximizing (4) subject to constraints dictated by rationality, stationarity, core and feasibility, given value-function definitions and bounds on money holdings. In particular, there are bounds necessary to guarantee that measures of people across states add up to one, and transition probabilities defined by lotteries also add up to one, so that in the outside-money case money is not created nor destroyed in meetings.

Our approach is to guess and verify that value functions are increasing and concave (that is, $0 \leq v_0^k < v_1^k < v_2^k$ and $v_2^k - v_1^k < v_1^k - v_0^k$ for $k = n, b$). We also restrict lotteries associated to people with intermediation occupations, due to incentive constraints, and transform (27) into inequality constraints. This way the numerical problem fits in conventional non-linear maximization routines. We then resort to the *KNITRO* solver. Issues related to local optima are handled by considering of many alternative initial conditions.

We now resort to an example of how (27) is handled, and how some lotteries can be eliminated in the outside-money case. Let us fix $\tilde{m} = (0,2,1)$. Since the consumer with never end with two units of money, we put $\sigma^{\tilde{m}}(i,2) = 0$ for $i \in \{0,1,2\}$. Also, we can impose $\sigma^{\tilde{m}}(0,0) = 0$, since money cannot be destroyed. In addition, the intermediary would not entertain an allocation with less than two units after trade, so that $\sigma^{\tilde{m}}(2,0) = \sigma^{\tilde{m}}(1,1) = \sigma^{\tilde{m}}(2,1) = 0$. It remains to be determined just two transition probabilities for this meeting, that is, choices of $\sigma^{\tilde{m}}(1,0)$ and $\sigma^{\tilde{m}}(0,1)$. Hence we can write

$$\begin{aligned} L_{01}(\tilde{m}) &= (\mathbf{f}_1^n - \mathbf{f}_1^n)PTSv + (\mathbf{f}_1^n - \mathbf{f}_1^n)PTSv = 0. \\ L_{10}(\tilde{m}) &= (\mathbf{f}_0^n - \mathbf{f}_1^n)PTSv + \zeta_1(\tilde{m})(\mathbf{f}_1^n - \mathbf{f}_0^n)PTSv \\ &= (\mathbf{f}_0^n - \mathbf{f}_1^n)PTSv + u'(y(\tilde{m}))(\mathbf{f}_1^n - \mathbf{f}_0^n)PTSv. \end{aligned} \tag{28}$$

Given that $\sigma^{\tilde{m}}(1,0) + \sigma^{\tilde{m}}(0,1) = 1$, the core constraint for meeting $\tilde{m}$ becomes

$$\begin{aligned} &\left(u'(y(\tilde{m})) - \frac{(\mathbf{f}_1^n - \mathbf{f}_0^n)PTSv}{(\mathbf{f}_1^n - \mathbf{f}_0^n)PTSv}\right)\sigma^{\tilde{m}}(1,0) \geq 0 \\ &\Leftrightarrow \left(u'(y(\tilde{m})) - 1\right)\sigma^{\tilde{m}}(1,0) \geq 0. \end{aligned} \tag{29}$$

References

Bagehot, W. (1873). *Lombard Street: a description of the money market.* New York: Scribner's.

Bewley, T. (1980). The optimum quantity of money. In J. H. Kareken and N. Wallace (Eds.), *Model of Monetary Economies*. Federal Reserve Bank of Minneapolis.

Cavalcanti, R. d. O. and E. Nosal (2009, May). Some benefits of cyclical monetary policy. *Economic Theory 39*(2), 195–216.

Cavalcanti, R. d. O. and D. Puzzello (2010). Stationarity without degeneracy in a model of commodity money. *Economic Theory 43*(2), 263–280.

Cavalcanti, R. d. O. and N. Wallace (1999). Inside and outside money as alternative media of exchange. *Journal of Money, Credit and Banking 31*(3), pp. 443–457.

Deviatov, A. (2006). Money creation in a random matching model. *Topics in Macroeconomics 6*(3), 1–20.

Deviatov, A. and N. Wallace (2014). Optimal inflation in a model of inside money. *Review of Economic Dynamics 17*(2), 287 – 293.

Friedman, M. (1953). *Essays in Positive Economics.* Business Economics. University of Chicago Press.

Friedman, M. (1969). *The Optimum Quantity of Money and Other Essays.* Business Economics. Aldine.

Imrohoroglu, A. (1992, January). The welfare cost of inflation under imperfect insurance. *Journal of Economic Dynamics and Control 16*(1), 79–91.

Kehoe, T. J., D. K. Levine, and M. Woodford (1992). The optimum quantity of money revisited. In P. Dasgupta, D. Gale, O. Hart, and E. Maskin (Eds.), *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, pp. 501–526. Cambridge and London: MIT Press.

Levine, D. K. (1991). Asset trading mechanisms and expansionary policy. *Journal of Economic Theory 54*(1), 148 – 164.

Levine, D. K. (2015). The keynesian illusion. http://www.dklevine.com/general/illusion.5.htm.

Lucas, Robert E, J. (1980, April). Equilibrium in a Pure Currency Economy. *Economic Inquiry 18*(2), 203–20.

Lucas, R. J. (1976, January). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy 1*(1), 19–46.

Molico, M. (2006). The distribution of money and prices in search equilibrium. *International Economic Review 47*(3), 701–722.

Scheinkman, J. A. and L. Weiss (1986, January). Borrowing Constraints and Aggregate Economic Activity. *Econometrica 54*(1), 23–45.

Shi, S. (1995, December). Money and prices: A model of search and bargaining. *Journal of Economic Theory 67*(2), 467–496.

Trejos, A. and R. Wright (1995, February). Search, Bargaining, Money, and Prices. *Journal of Political Economy 103*(1), 118–41.

Wallace, N. (1992). Lucas's signal-extraction model: A finite state exposition with aggregate real shocks. *Journal of Monetary Economics 30*(3), 433 – 447.

Wallace, N. (2014). Optimal money creation in "pure currency" economies: a conjecture. *The Quarterly Journal of Economics 129*(1), 259–274.