Counterfeiting as Private Money in Mechanism Design*

Ricardo Cavalcanti     Ed Nosal
FGV, Brazil      Cleveland Fed

October 11, 2007

Abstract

We describe counterfeiting activity as the issuance of private money, one which is difficult to monitor. Our approach, which amends the basic random-matching model of money in mechanism design, allows a tractable welfare analysis of currency competition. We show that it is not efficient to eliminate counterfeiting activity completely. We do not appeal to lottery devices, and we argue that this is consistent with imperfect monitoring.

1 Introduction

Since the beginning of time, every monetary economy has had to deal with the issue of counterfeit currency. In spite of this robust observation, to date there is no model that provides an understanding of the value of genuine money and the boundaries to counterfeiting. In this paper, we identify the basic ingredients that a theory of counterfeiting requires. In particular, we

*We have benefited from the discussions at the New York Fed Workshop on Payments, the SED Meetings, the SAET Meetings and the Cleveland Fed Workshop on Money, Banking, Payments and Finance.

1Counterfeiting activities are pervasive, having a tendency to become more attractive the more valuable a currency becomes. Adoption of monitoring processes that increase detection rates and counterfeiting costs is the usual response, but it is evident that new technologies also become available for the provision of crime itself.
view counterfeiting as a second-best outcome, which arises because monitoring resources are scarce. A key insight is that a theory of counterfeiting is essentially a theory of private money—albeit a private money that is a rather undesirable substitute for the real thing.

One can think of a counterfeit note as a good that is produced and priced in a competitive environment: we do not take this approach. We believe that a purely technological explanation for counterfeiting—one that uses classic price theory to predict the value of money—is a dead end. For the fundamental value of money—and its substitutes—comes from their ability to overcome exchange difficulties and not from the scarcity of paper and ink needed to produce them. As a result, an attractive theory of private money should describe formally what problem money is solving, what its exchange value is, and to what extent substitutes can be adopted or, in the case of counterfeiting, tolerated. This is what we mean by the boundaries of counterfeiting.

It is important to emphasize that private money, or for that matter credit, is a response to an opportunity that may arise; in the case of credit, the opportunity is the existence of would-be borrowers and lenders and enforcement devices. And counterfeiting is no exception. In modern economies, resources are allocated for some transactions—typically high-value transactions—to ensure that settlement is seamless. And seamless settlement requires a high level of monitoring. For example, stock market purchases are settled through broker accounts and house purchases are settled by escrow accounts. These types of accounts are designed to detect counterfeiting activity almost surely; more to the point, these types of high-value transactions are not the kind of opportunities that counterfeiters can use to successfully put their notes into circulation. Since monitoring is costly, society will allocate its scarce monitoring resources to important, high-value transactions, leaving relatively small-value transactions less protected.

The notion that limited monitoring is necessary for the essentiality of money is a reasonably old one, dating back at least to Ostroy (1973). A more recent formulation appears in Kocherlakota (1998), who emphasizes that holding money constitutes tangible evidence of an earlier socially-desirable activity; that is, it indicates that the current holder of money has in the

---

2 One can think of the use of these accounts as a form of signaling: one has assets of good quality and is willing to wait before trading so that the asset quality in the account can be verified. Obviously, not all transactions can be performed on this trading-account basis.
past traded valuable goods for money. Cavalcanti and Wallace (1999) take it a step further by introducing imperfect monitoring of agents in a way that gives rise to the coexistence of money and credit transactions. In their model, the notion of imperfect monitoring concerns the ability—or inability—to record the actions that people take. In this paper, we pursue another avenue: imperfect monitoring applies not to people but to the attributes of the means of payments. In particular, whether the means of payments are recognized as being either genuine or counterfeit.

A more general model of multiple assets—which may include undesirable ones such as counterfeit money—would allow ranges for monitoring of both people’s actions and assets’ attributes. Society would set up trading institutions, with an eye on monitoring costs, which ultimately generates a spectrum of exchange opportunities, where large transactions will be heavily monitored and small ones will not. Individuals would respond by issuing questionable quality assets for only the small-value transactions. In this paper, we focus exclusively on small-value transactions and assume that it is only possible to imperfectly monitor assets. We tackle this by amending the basic random-matching model of money—which has only one asset—to allow for a mechanism-design analysis of the coexistence of genuine and counterfeit money.

For reasons of tractability, we preserve the anonymity of individuals and restrict attention to pairwise exchanges, where money holdings are restricted to either zero or one units. We ask what is the best incentive-feasible response to counterfeiting? Because we describe an optimum, we can verify whether the coexistence of counterfeits and genuine money is a robust phenomenon. Indeed, we find a necessary condition for robustness: counterfeiting opportunities must be distributed unevenly across individuals.

The rest of the paper is as follows. In section 2, the environment and our concept of implementability are presented. In section 3, we discuss optimality. In section 4, we provide concluding remarks, which include comments on the existing literature and a discussion of why we choose not to use lotteries. The appendix contains all the proofs and additional algebra for solving the model.
2 The environment and the equilibrium concept

Our model modifies the environments of Trejos and Wright (1995), Shi (1995) and Cavalcanti and Wallace (1999) to allow for the costly production of an alternative medium of exchange. We will refer to lawful money or outside money as fiat notes, and the alternative as counterfeit notes.

2.1 The environment

Time is discrete and the horizon is infinite. A unit measure population is divided into $N$ fixed types according to the goods they can produce and consume, where $N \geq 3$. There are $N$ types of perishable goods. An $i$-type individual specializes in consuming only good $i$ and producing only good $i + 1$ (modulo $N$). Individuals maximize discounted expected utility. Period utility for an $i$-type individual who produces a counterfeit note is $u(x) - y - \omega$ and $u(x) - y$ if he does not, where $x$ is the amount of good $i$ consumed, $y$ is the amount of good $i + 1$ produced and $\omega$ is the utility cost of producing a counterfeit note. The function $u$ is continuous, concave, differentiable, with $u(0) = 0$, $u'(0) = +\infty$, and $u'(+\infty) = 0$. The discount factor is $\beta \in (0, 1)$.

Individuals are unable to commit to future actions, and histories of their actions are private information. In order to facilitate trade, a durable object, such as money, is required. We assume that individuals can hold either one unit of money or no money at all. If an individual holds a unit of money, it may be either a fiat note or a counterfeit note. The economy is endowed with a fixed stock of fiat money, $\mu_1 \in (0, 1)$, to be chosen by a planner. Fiat money is perfectly durable and lasts forever. Counterfeit money, however, is not perfectly durable: In each period there is a fixed probability that a counterfeit note disintegrates.

Each period has two subperiods. At the beginning of the first subperiod, each individual draws an idiosyncratic realization for a nonnegative utility cost of counterfeiting $\omega$. The cost of counterfeiting, $\omega$, is modeled as the realization of a random variable, identically and independently distributed over time and across individuals, according to the cumulative distribution function $F$, assumed to have support in a bounded interval of $\mathbb{R}_+$. Except for a degenerate case considered in one part of the analysis, the function $F$ is assumed to be continuous and differentiable, with $F(0) = 0$ and $F'(0) < +\infty$. 

4
After individuals learn their current-period $\omega$, those who are not holding any type of money will either produce a counterfeit note or not. After counterfeit notes are produced, with probability $\pi \in [0, 1]$, a counterfeit note—either an old note or a newly produced one—disintegrates.

In the second subperiod, individuals meet randomly and in pairs. In a single-coincidence meeting—e.g., a meeting between $i$-type and $i - 1$-type individuals—a perishable good may be produced and consumed if the buyer—the $i$-type individual—has a unit of money and the seller—the $i - 1$-type individual—does not. The seller is unable to distinguish between a unit of counterfeit money and fiat money at the time of trade. After trade occurs, the seller learns about the type of money—fiat or counterfeit—that he has just acquired and this information remains private. As a result, buyers are eventually informed about the quality of the money they hold.

2.2 Allocations

The planner's problem is to maximize the average ex ante utility of individuals, by choosing an allocation in some class. We restrict attention to a class with symmetry and stationarity properties. We further restrict attention to allocations in which counterfeit notes trade for the same level of output as a fiat note, so that only “one price” is observed.

The average utility in the planner’s objective is taken with regard to the endogenous distribution of individuals, as indexed by their money holdings. The planner’s maximization problem includes participation constraints, which are dictated by individual rationality. As in Cavalcanti and Wallace (1999), we assume that individuals cannot coordinate on defection, so the participation constraints only reflect the possibility of an individual defection.

The distribution of money at the beginning of the first subperiod is given by $\mu = (\mu_0,\mu_1,\mu_2)$, where $\mu_0$ represents the fraction of individuals holding no money, $\mu_1$ is the fraction of individuals holding a fiat note and $\mu_2$ is the fraction of individuals holding a counterfeit note. The beginning-of-the-first-subperiod value functions associated with the various money holdings are denoted by $w = (w_0, w_1, w_2)$, where the notation should be obvious. The beginning-of-the-second-subperiod value functions—which are evaluated after the counterfeiting costs are incurred and after a fraction $\pi$ of counterfeit notes disintegrate, but before individuals are matched—are denoted by $v = (v_0, v_1, v_2)$. 
We anticipate that a necessary condition for optimality will, in part, be characterized by the existence of a cutoff counterfeiting cost, \( \bar{\omega} \), such that only individuals who do not hold money and draw an \( \omega < \bar{\omega} \) will choose to produce a counterfeit note. As well, we anticipate that in any allocation, holders of money never dispose of their monies in the first subperiod: There is no point for a holder of a counterfeit note to dispose of it at the beginning of the first subperiod only to (possibly) produce another one that has an identical chance of being disintegrated. As well, since a fiat note has no chance of being disintegrated and has value, a holder of a fiat note will never dispose of it. Notice that the realization \( \omega \) for an individual is not a state variable at the second subperiod because counterfeiting decisions have already been made in the first subperiod.

In a steady state, inflows and outflows into the different money holding states “cancel out”; this implies that in a steady state the distribution of money holdings at the beginning of the first subperiod, \( \mu = (\mu_0, \mu_1, \mu_2) \), also describes the distribution of money holdings at the beginning of the second subperiod. The stationarity requirement for \( (\mu, \bar{\omega}) \) is as follows. In the steady state, the amount of counterfeit notes produced at the beginning of the first subperiod, \( \mu_0 F(\bar{\omega}) \), must equal the amount that disintegrates at the end of the first subperiod, \( \pi \mu_0 F(\bar{\omega}) + \pi \mu_2 \). With the understanding that \( \mu \) is a probability measure, \( (\mu_i \geq 0 \text{ and } \mu_0 + \mu_1 + \mu_2 = 1) \), an allocation \( (\mu, y, \bar{\omega}) \) is said to be stationary if it satisfies

\[
\mu_2 = \frac{(1 - \pi) F(\bar{\omega})}{\pi} \mu_0.
\]

In summary, we assume a class of outcomes that are stationary and symmetric across consumption/production types and states. We define an allocation to be a list \( (\mu, y, \bar{\omega}) \) satisfying (1), where money holdings are distributed according to \( \mu \), both monies trade for \( y \)—the level of output produced and consumed in single-coincidence meetings—and counterfeit notes are created according to \( \bar{\omega} \).

### 2.3 Implementability

Allocation \( (\mu, y, \bar{\omega}) \) is implementable if it satisfies individual rationality requirements. Individual rationality for producers of consumption goods takes
the form of the participation constraint
\[ y \leq \beta \left[ \frac{\mu_1}{\mu_1 + \mu_2} (w_1 - w_0) + \frac{\mu_2}{\mu_1 + \mu_2} (w_2 - w_0) \right], \tag{2} \]
since the producer can only use his knowledge about \( \mu \) to infer the probability he is receiving fiat or counterfeit money. The difference \( w_i - w_0 \) represents the increase in expected discounted utility associated with an individual starting the first subperiod with money \( i = 1, 2 \) compared to starting with no money at all. The bracketed term on the right-hand side of (2) represents the increase in expected discounted utility associated with accepting a unit of money in trade in exchange for some output. Since a producer receives this benefit beginning the next period, the value of this benefit (today) must be discounted by \( \beta \). The left-hand side of (2) represents the cost of receiving this benefit, i.e., the cost of producing output \( y \). Hence, the seller will produce \( y \) if the benefit exceeds \( y \). Since we assume that fiat and counterfeit money trade for the same level of output, individual rationality for the consumer is simply
\[ u(y) \geq \beta \max\{w_1 - w_0, w_2 - w_0\}. \tag{3} \]
The consumer knows whether he is holding a counterfeit or fiat note; he will only trade the note if the benefit of surrendering the note, which is given by the left-hand side of (3), exceeds the cost, which is given by the right-hand side (3).³

Notice that there is an individual rationality constraint associated with the production of counterfeit notes, i.e., \( \omega < \bar{\omega} \). We can be more explicit about the critical cutoff \( \bar{\omega} \). The benefit of creating a counterfeit, \( w_2 - v_0 \), can be simplified to read
\[ w_2 - v_0 = (1 - \pi)v_2 + \pi v_0 - v_0 = (1 - \pi)(v_2 - v_0), \]
where \( \pi \) is the probability that a counterfeit note disintegrates. Therefore, the cutoff value for \( \omega, \bar{\omega} \), for which \( \omega \leq \bar{\omega} \) produces a counterfeit note, satisfies
\[ \bar{\omega} = (1 - \pi)(v_2 - v_0). \tag{4} \]

The allocation \((\mu, y, \omega)\) is implementable if there exists nonnegative \((w, v)\) satisfying participation constraints (2)-(4), as well as the standard Bellman equation (which is described in the appendix).

³In the appendix we show that if (2) holds, then so does (3). This is what one would expect. A producer will be only willing to supply output for a unit of money if he expects to spend it on a consumption good when he is a consumer.
2.4 The planner’s objective

It is straightforward to demonstrate that the average utility, $\sum_i \mu_i w_i$, associated to any implementable allocation $(\mu, y, \bar{\omega})$, is proportional to

$$W(\mu, y, \bar{\omega}) = \frac{1}{N} \mu_0 (1 - \mu_0) [u(y) - y] - \mu_0 \int_0^{\bar{\omega}} \omega dF.$$  \hspace{1cm} (5)

Equation (5) defines our *ex-ante welfare criteria*. The term $\frac{1}{N} \mu_0 (1 - \mu_0)$ represents the probability that a good is traded for money. The probability that a particular money holder (buyer) is matched with a seller who produces the good that he desires is $\frac{1}{N}$ times $\mu_0$, and the total measure of potential buyers is $\mu_1 + \mu_2 = 1 - \mu_0$. The term $\frac{1}{N} \mu_0 (1 - \mu_0)$ is sometimes referred to as the *extensive margin*. In each single-coincidence match, total period utility flow is $u(y) - y$; this flow is sometimes referred to as the *intensive margin*. Finally, in each first subperiod, there is a measure of agents without money, $\mu_0 \int_0^{\bar{\omega}} dF$, who choose to produce counterfeit bills, where the total cost of counterfeiting these bills is $\mu_0 \int_0^{\bar{\omega}} \omega dF$.

3 Optimality

We demonstrate in the appendix how the choice of an optimal allocation can be separated into steps: first a *margin* $(\mu_0, y)$ is selected; next a stationary allocation $(\mu, y, \bar{\omega})$ is constructed from $(\mu_0, y)$; finally, if the allocation satisfies the producer’s participation constraint, then it is shown to be implementable. The optimum is the one that maximizes the *ex ante* welfare criteria among the implementable allocations.

There are two attractive features of this algorithm. On the one hand it allows an easy comparison with the optimum of the standard model, where counterfeiting is ignored. This is convenient because the algebra necessary to derive the producer participation constraint with counterfeiting is quite lengthy, and the comparison with the standard case provides a simple check on our algebra. Incidentally, we have chosen the sequence of events so that the $\pi$-shock happens before trade. Hence, even if a subset of individuals have low counterfeiting costs, then, as $\pi$ approaches unity, counterfeiting does not take place, and the optimum is identical to that of the standard model.

On the other hand, under some conditions, we can show that the optimum problem can be solved on the $(\mu_0, y)$-space itself. We can show (in the
appendix) that both the objective (5) and producer participation constraint are well defined for a given margin and $\bar{\omega}$. Moreover, we can also show that when $F$ is uniform, $\bar{\omega}$ is uniquely defined for a given margin. The optimum problem is, however, not necessarily transformed into the maximization of (5) subject to only the producer participation constraint because, unlike the standard model, not all margins yield stationary allocations in our model. Specifically, if, for given $\mu_0$, the level of output $y$ is sufficiently high, then the associated $\bar{\omega}$ may generate, from (1), a $\mu_2 > 1 - \mu_0$, which violates the nonnegativity of $\mu_1$. As a result, in addition to the producer participation constraint, the nonnegativity requirement

$$\frac{(1 - \pi) F(\bar{\omega})}{\pi} \mu_0 \leq 1 - \mu_0$$

must be satisfied to ensure that the margin $(\mu_0, y)$ is implementable.

Notice that if $\mu_0 \leq \frac{1}{2}$ and $\pi \geq \frac{1}{2}$, then inequality (6) is always satisfied. This observation is useful because, as we show, when $\beta$ is sufficiently high, the optimum is characterized by $\mu_0 < \frac{1}{2}$. Hence, inequality (6) is satisfied in a large neighborhood of the standard model (where the standard model has $\pi = 1$) when $\beta$ is high. Note also that reductions in $\mu_0$ relax the nonnegativity constraint (because it also reduces the value of money, it also reduces $\bar{\omega}$).

Since the details associated with solving the model are not crucial to understanding the main results, we leave these derivations to the appendix. The main result is divided into two propositions. The first is

**Proposition 1** Assume that $F(\omega) > 0$ for all $\omega > 0$, and let $y^*$ be the unique maximizer of the intensive margin $u(y) - y$. Then any optimum $(\mu, y, \bar{\omega})$ features $\mu_2 > 0$. Moreover, if $\beta$ is sufficiently high, then $\mu_0 < \frac{1}{2}$ and $y < y^*$.

The proposition thus states a sufficient condition for which the optimum features a second-best allocation with counterfeiting, and it is that for some people, the cost of counterfeiting has to be arbitrarily small. It also states that when there is enough patience—so that the participation constraint for the producer can be ignored—both the extensive and intensive margins will be distorted, in contrast to the standard model (without counterfeiting),

---

4We derive the conditions under which the optimum problem is reduced to the standard case, i.e., the model with no counterfeiting; see claim 3 in appendix A3.

5In Nosal and Wallace (2007) if the cost of counterfeiting is sufficiently small, then there does not exist a monetary equilibrium.
where the optimum would set \( \mu_0 = \frac{1}{2} \) and \( y = y^* \). The deviation from the first-best optimum quantity of money is biased towards inflation.

The basic idea underlying proposition 1 is that the planner is aware that a higher value of money encourages more counterfeiting, and that some counterfeiting should be tolerated up to the point that the distortions it imposes on the economy become too high.

In numerical simulations with \( F \) uniform, we also found that \( y < y^* \) is robust to reductions in \( \beta \). When the participation constraint binds, reductions in \( y \) below \( y^* \) tend to reduce \( \bar{\omega} \) and the value of counterfeits. We are unable to prove the result in general because we cannot establish how changes in \( \bar{\omega} \) affect the participation constraint for arbitrary parameter values.

We now argue that, except in some extreme cases, the planner can and will essentially eliminate counterfeiting when the distribution of shocks is degenerate. The basic idea is as follows. Consider a perturbation of an implementable allocation with positive counterfeiting when the distribution of shocks is degenerate and individuals without money are indifferent between counterfeiting or not. Keeping the margin \( (\mu_0, y) \) constant, consider a perturbation that reduces the measure of those creating counterfeits at all dates, accompanied by a corresponding increase in the measure of those who hold genuine money. Because, as we show, \( \bar{\omega} \) is determined by \( (\mu_0, y) \) alone—and thus invariant to the ratio \( \mu_2/\mu_1 \)—the indifference to counterfeiting is maintained. Indeed, the value of holding one unit of genuine money is not affected because consumption opportunities are fully described by \( (\mu_0, y) \). Likewise, the value of holding a counterfeit does not change. Because the ratio \( \mu_2/\mu_1 \) is reduced, the participation constraint for the producer is weakened and the perturbation is thus implementable. But because \( \mu_2 \) is reduced, the flow of resources lost to counterfeiting is also reduced, and social welfare is, therefore, increased. The argument implies that, under homogenous counterfeiting opportunities, the optimum either has none or all of the individuals without money counterfeiting. We consider the first case—(i) below—to be the relevant one.

**Proposition 2** Assume that the support of \( F \) is the singleton \( \{\bar{\omega}\} \). (i) No Counterfeiting: If \( \pi \) is sufficiently low or sufficiently high, then the optimum has no counterfeiting. (ii) Counterfeiting: \((\mu, y, \bar{\omega})\) with \( \bar{\omega} > \bar{\omega} \) is implementable if and only if \( \mu_0 \leq \pi, \mu_2 = \frac{1-\pi}{\pi} \mu_0, \mu_1 = 1 - \mu_0 - \mu_2 \geq 0, \)

\[
\bar{\omega} = \frac{1 - \beta + \frac{\beta}{N} \mu_0}{1 - \beta + \pi \frac{\beta}{N}} \left\{ \frac{1 - \pi}{N} \left[ \mu_0 u(y) + (1 - \mu_0) y \right] + \beta \bar{\omega} \right\};
\]
\[ y \leq \beta \bar{\omega} + \frac{\beta \mu_1}{\mu_1 + \mu_2 (1 - \pi)(1 - \beta + \frac{\pi}{N})}, \]  
(8)

where \( \hat{\beta} = \beta \left(1 - \pi\right)\left(1 - \frac{1}{N}\right) \), and the associated discounted utilities solving the Bellman equation are nonnegative.

The first part of proposition 2 tells us that if \( \pi \) is sufficiently high, then it is not possible to have \( \bar{\omega} > \bar{\omega} \); hence, there cannot be counterfeiting. Intuitively, if the probability of having a counterfeit bill confiscated is high, then the critical cost of counterfeiting, \( \bar{\omega} \), will be extremely low; in fact, it will be lower than the actual cost of counterfeiting, \( \bar{\omega} \). If, on the other hand, \( \pi \) is sufficiently low, then in order to ensure that \( \mu_1 \) is nonnegative, \( \mu_0 \) must also be sufficiently low. But if both \( \pi \) and \( \mu_0 \) are low, there will be very few trading opportunities and, hence, it will not be possible to have \( \bar{\omega} > \bar{\omega} \). The second part of proposition 2 describes implementable allocations when all individuals without money choose to counterfeit. In particular, the proposition suggests that an allocation with \( \mu = (\pi, 0, 1 - \pi) \) and \( y = \beta \bar{\omega} \) can be implementable if \( \bar{\omega} \) is not too high, so that discounted utilities are nonnegative.

4 Final remarks

There is a small literature on counterfeiting,\(^6\) and the papers closest to ours are Green and Weber (1996) and Nosal and Wallace (2007). In the former, the supply of counterfeit notes is essentially exogenous. Both papers assume that the cost of counterfeiting is homogeneous and neither emphasizes optimality. The focus of the latter paper is on the emergence of separating prices when traders can use lotteries. We conclude by defending our assumptions of no lotteries and 0-1 money holdings.

Recall that lotteries were introduced in monetary theory to approximate divisible money. Absent lotteries, models with indivisible money holdings could predict inefficient trade outcomes—too much output is produced—and such a result is a direct implication of the indivisibility. But this is

\(^{6}\)In the monetary models of Kultti (1996) and Williamson (2002), small levels of counterfeiting is a possibility but not an equilibrium outcome. Quercioli and Smith (2007) develop a static model in which individuals also choose monitoring efforts on different denominations, but the lack of general equilibrium makes statistics about “money” value and circulation difficult to interpret.
not a relevant issue here. First, in our model, output production is never inefficiently high. Second, and more importantly, counterfeiting money, as opposed to wine, is about the attempt to misrepresent the quality of a given \textit{indivisible} object.

One can think of Nosal and Wallace (2007) as presenting a fundamental result about signaling in pairwise trades, where signaling requires the use of lotteries. Although we do believe that signaling is a relevant issue for counterfeiting—for example, see footnote 2—we have, as a first step, focused on lightly monitored exchanges and have “ignored” the more heavily monitored transactions where signalling is relevant.

In practice, there are large costs associated with distributing counterfeit bills in an economy. For example, one would not purchase a $600 suit with 30 counterfeit $20 bills; instead, one would use a very small number of counterfeits at any one time to avoid detection. For this reason, we feel comfortable with our 0-1 money holding structure. More generally, the existence of monitoring costs will restrict the scope of counterfeit or private money usage, relegating it to relatively small trades. Because trading opportunities are bounded, as are profits, private money can coexist with outside money. This idea has been laid out in a random-matching model by Cavalcanti et al. (1999).

Future research could expand our model in the direction of allowing small and large transactions, as well as the allocation of scarce monitoring technologies across them. Such an extension should yield predictions about the formation of markets where large transactions become safer. At the same time, it could offer insights about the supremacy of money issued by the government for settling large transactions as a result of its recognizability.
References


The possibility of counterfeiting forces a higher dimensionality of stationary allocations with respect to those of the standard monetary environment. Implementable distributions of money $\mu$ are now tied to the level of output, $y$, insofar as changes in the value of (fiat) money—which can be brought about by changes in $y$—affect the supply of counterfeits. By contrast, in the standard model with holdings in $\{0, 1\}$, the distribution of money holdings can be set independently of $y$. Here we have to address the admissibility of the distribution of money holdings, $\mu$, relative to $y$.

After describing admissibility, we show how to convert the current problem with counterfeiting into a standard one. The optimal problem can be stated in various ways, as different series of sub-problems, where, for example, the planner could choose $\mu_1 \in [0, 1]$, the quantity of fiat money, and then optimize with respect to $y$, and then “market forces” choose $\mu_2$ according to incentive constraints. We prefer to present the problem as choosing first the intensive and extensive margins—that is, the point mass $\mu_0$ and the “price” $y$—and then look for equilibrating $\mu_1$ and $\mu_2$. We then examine if the implied allocation satisfies the participation constraints (2) and (3).

We say that a margin $(\mu_0, y)$ is admissible if there exists $(\mu_1, \mu_2)$ and $\bar{\omega}$ such that $[(\mu_0, \mu_1, \mu_2), y, \bar{\omega}]$ satisfies the stationarity requirement (1) and implementability requirement (4), where $\mu_1, \mu_2 \geq 0$. In the standard monetary environment, $(\mu_0, y)$ defines a unique allocation. In our problem, the inclusion of “$\bar{\omega}$” in the description of outcomes is important because $\bar{\omega}$ helps to describe the inflow of counterfeits and their costs. One has to check, therefore, whether an arbitrary $(\mu_0, y)$ is admissible. In particular, there must exist an $\bar{\omega}$ such that (i) $\mu$ is a probability measure that satisfies (1); and (ii) $(w, v)$ satisfy both (4) and the Bellman equation associated with $(\mu_0, y)$. The following lemmas explains how the critical cost of producing a counterfeit, $\bar{\omega}$, can be determined for a given $(\mu_0, y)$. We shall present the Bellman equation in a way that facilitates proving these lemmas.

**Lemma 1** A margin $(\mu_0, y)$ is admissible if and only if there exists $\bar{\omega}$ satisfying

$$
\frac{(1 - \pi) F(\bar{\omega})}{\pi} \mu_0 \leq 1 - \mu_0
$$

(9)
and
\[ \bar{\omega} F(\bar{\omega}) a(\mu_0) + \bar{\omega} b(\mu_0) = c(\mu_0, y) + \int_0^{\bar{\omega}} \omega dF, \] (10)

where \( a, b, \) and \( c \) are continuous and increasing functions described below.

The inequality (9) assures that the \( \mu_2 \) given by (1) implies a nonnegative \( \mu_1 \), where \( \mu_1 = 1 - \mu_0 - \mu_2 \), while (10) is the condition (4) after the Bellman equation is solved for \((w, v)\), given \((\mu_0, y)\) and the stationarity assumption (1).

**Lemma 2** There exists at least one solution to (10). If \( F \) is uniform, with support \([0, \omega_H]\), and either \( \omega_H \) is sufficiently high or \( y \) is sufficiently low, then this solution is unique in \([0, \omega_H]\).

Let admissible \((\mu_0, y)\) be the margin of allocation \((\mu, y, \bar{\omega})\). We say that \((\mu_0, y)\) is implementable if \((\mu, y, \bar{\omega})\) is implementable. Hence implementability of a margin requires existence of an allocation that, in addition to the admissibility requirements, yields nonnegative discounted utilities and satisfies the producer’s and consumer’s participation constraints (2) and (3). We now show how to verify the satisfaction of these participation constraints in a simple way.

**Lemma 3** Suppose that an allocation \((\mu, y, \bar{\omega})\) yields nonnegative discounted utilities and its margin is admissible. This allocation is implementable if and only if

\[ y \leq d(\mu_0, \bar{\omega}), \] (11)

where \( d \) is a continuous function to be described below.

The Bellman equation has the standard structure satisfying a contraction-mapping property: when an allocation \((\mu, y, \bar{\omega})\) is fixed, current discounted utilities are given by the utility flow corresponding to the fixed allocation, and by future discounted utilities, discounted by \( \beta \). A unique solution \((w, v)\) is therefore associated to the fixed \((\mu, y, \bar{\omega})\). Instead of solving for \((w, v)\) we choose below a more tractable algebra that solves the Bellman equation for the differences \( \Delta_i \equiv v_i - v_0 \), for \( i = 1, 2 \), which suffices for checking admissibility and implementability.
We start by showing how the solution for $\Delta_2$ is used to produce the requirement (10), expressing the consistency between $\Delta_2$ and $\bar{\omega}$ by the way of (1) and (4).

The following notation will be useful: $k \equiv \int_0^{\bar{\omega}} \omega dF$ and $f \equiv F(\bar{\omega})$. The discounted utility $w_2$ is related to $v_2$ and $\Delta_2$ according to

$$w_2 = (1 - \pi) v_2 + \pi v_0 = v_2 - \pi (v_2 - v_0) = v_2 - \pi \Delta_2.$$  \hspace{1cm} (12)

The discounted utility $w_0$ is related to $k$, $f$, $v_0$ and $\Delta_2$ according to

$$w_0 = \int_0^{\bar{\omega}} (w_2 - \omega) dF + (1 - f) v_0$$
$$= -\int_0^{\bar{\omega}} \omega dF + w_2 \int_0^{\bar{\omega}} dF + (1 - f) v_0$$
$$= -k + v_0 + f (w_2 - v_0)$$
$$= -k + v_0 + f (1 - \pi) \Delta_2.$$ \hspace{1cm} (13)

We now derive explicit expressions for the value function $v$. Starting with the discounted utility $v_1$,

$$v_1 = \frac{\mu_0}{N} (u + \beta w_0) + \left(1 - \frac{\mu_0}{N}\right) \beta v_1.$$  \hspace{1cm} (14)

Since $w_1 = v_1$,

$$(1 - \beta) v_1 = \frac{\mu_0 u}{N} + \beta \frac{\mu_0}{N} (w_0 - v_1)$$
$$= \frac{\mu_0 u}{N} + m_0 (-k - \Delta_1 + f (1 - \pi) \Delta_2),$$

where $m_i \equiv \beta \mu_i / N$. Note that $\beta / N = m_0 + m_1 + m_2$. The discounted utility $v_2$ satisfies

$$v_2 = \frac{\mu_0}{N} (u + \beta w_0) + \left(1 - \frac{\mu_0}{N}\right) \beta v_2$$
$$= \frac{\mu_0}{N} (u + \beta w_0) + \left(1 - \frac{\mu_0}{N}\right) \beta (v_2 - \pi \Delta_2)$$
or

\[(1 - \beta) v_2 = \frac{\mu_0 u}{N} - \beta \pi \Delta_2 + m_0 \left( -(v_2 - w_0) + \pi \Delta_2 \right) \]

\[= \frac{\mu_0 u}{N} - \beta \pi \Delta_2 + m_0 \left[ - (v_2 - (-k + v_0 + f (1 - \pi) \Delta_2)) + \pi \Delta_2 \right] \]

\[= \frac{\mu_0 u}{N} - \beta \pi \Delta_2 + m_0 (-k - \Delta_2 + f (1 - \pi) \Delta_2 + \pi \Delta_2) \]

\[= \frac{\mu_0 u}{N} - \beta \pi \Delta_2 + m_0 (-k - (1 - f) (1 - \pi) \Delta_2). \] \hspace{1cm} (15)

Finally, the discounted utility \( v_0 \) can be represented as

\[v_0 = \frac{1 - \mu_0}{N} \left( -y + \beta \left( \frac{\mu_1}{\mu_1 + \mu_2} w_1 + \frac{\mu_2}{\mu_1 + \mu_2} w_2 \right) \right) + \left( 1 - \frac{1 - \mu_0}{N} \right) \beta w_0 \]

\[= \frac{1 - \mu_0}{N} y + m_1 w_1 + m_2 w_2 + \left( 1 - \frac{1 - \mu_0}{N} \right) \beta (-k + v_0 + f (1 - \pi) \Delta_2) \]

or

\[(1 - \beta) v_0 = -\frac{1 - \mu_0}{N} y - \beta k + \beta f (1 - \pi) \Delta_2 + m_1 w_1 + m_2 w_2 \]

\[+ (m_1 + m_2) w_0 \]

\[= -\frac{1 - \mu_0}{N} y - \beta k + \beta f (1 - \pi) \Delta_2 + \]

\[m_1 (w_1 - w_0) + m_2 (w_2 - w_0) \]

\[= -\frac{1 - \mu_0}{N} y - \beta k + \beta f (1 - \pi) \Delta_2 + m_1 (k + \Delta_1 - f (1 - \pi) \Delta_2) \]

\[+ m_2 (k + (1 - f) (1 - \pi) \Delta_2). \] \hspace{1cm} (16)

Subtracting (16) from (14), and recognizing that \( (1 - \beta) v_1 - (1 - \beta) v_0 = (1 - \beta) \Delta_1 \), we get
\[
(1 - \beta) \Delta_1 = \frac{\mu_0 u}{N} + m_0 (-k - \Delta_1 + f (1 - \pi) \Delta_2) - \\
\left\{ -\frac{1 - \mu_0}{N} y - \beta k + \beta f (1 - \pi) \Delta_2 + m_1 (k + \Delta_1 - f (1 - \pi) \Delta_2) + \\
m_2 (k + (1 - f) (1 - \pi) \Delta_2) \right\}
\]
\[
= \frac{\mu_0 u}{N} + \frac{1 - \mu_0}{N} y + \beta k + m_0 (-k - \Delta_1 + f (1 - \pi) \Delta_2) - \\
-\beta f (1 - \pi) \Delta_2 - m_1 (k + \Delta_1 - f (1 - \pi) \Delta_2) - \\
\left( \frac{\beta}{N} m_0 - m_1 \right) (k + (1 - f) (1 - \pi) \Delta_2)
\]
\[
= \frac{\mu_0 u}{N} + \frac{1 - \mu_0}{N} y + \beta k - \frac{\beta}{N} k - \beta f (1 - \pi) \Delta_2 - \\
(m_0 + m_1) \Delta_1 + \frac{\beta}{N} f (1 - \pi) \Delta_2 - m_2 (1 - \pi) \Delta_2
\]
\[
= M + \frac{\beta}{N} f (1 - \pi) \Delta_2 - \beta f (1 - \pi) \Delta_2 - (m_0 + m_1) \Delta_1 - m_2 (1 - \pi) \Delta_2,
\]
where
\[
M = \frac{\mu_0 u}{N} + \frac{1 - \mu_0}{N} y + \beta k \left( 1 - \frac{1}{N} \right).
\]

Similarly, subtracting (16) from (15), we get
\[
(1 - \beta) \Delta_2 = -\beta \pi \Delta_2 + \frac{\mu_0 u}{N} + m_0 (-k - (1 - f) (1 - \pi) \Delta_2) - \\
-\frac{1 - \mu_0}{N} y - \beta k + \beta f (1 - \pi) \Delta_2 + \\
m_1 (k + \Delta_1 - f (1 - \pi) \Delta_2) + m_2 (k + (1 - f) (1 - \pi) \Delta_2)
\]
\[
= -\beta \pi \Delta_2 + M - \beta f (1 - \pi) \Delta_2 - \frac{\beta}{N} (1 - f) (1 - \pi) \Delta_2 + \\
m_1 (1 - \pi) \Delta_2 - m_1 \Delta_1.
\]

Now, subtracting (19) from (17) we get
\[
(1 - \beta) (\Delta_1 - \Delta_2) = -(m_0 + m_1) \Delta_1 + \frac{\beta}{N} f (1 - \pi) \Delta_2 - m_2 (1 - \pi) \Delta_2 - \\
\left\{ -\beta \pi \Delta_2 - \frac{\beta}{N} (1 - f) (1 - \pi) \Delta_2 + m_1 (1 - \pi) \Delta_2 - m_1 \Delta_1 \right\}
\]
\[
= -m_0 \Delta_1 + m_0 (1 - \pi) \Delta_2 + \beta \Delta_2 \pi.
\]
Rearranging terms we get

\[(1 - \beta + m_0) \Delta_1 = [1 - (1 - \pi) \beta - m_0)] \Delta_2, \quad (20)\]

which implies that \(\Delta_1 > \Delta_2\) for \(\Delta_2 > 0\). Substituting for \((1 - \beta) \Delta_1\) from equation (17), into the above, we get

\[\left[ M + \frac{\beta}{N} f (1 - \pi) \Delta_2 - \beta f (1 - \pi) \Delta_2 - (m_0 + m_1) \Delta_1 - m_2 (1 - \pi) \Delta_2 \right] + m_0 \Delta_1 = [1 - (1 - \pi) (\beta - m_0)] \Delta_2\]

or

\[\left[ 1 - \beta + \pi \beta + \beta f (1 - \pi) + m_2 (1 - \pi) - \frac{\beta}{N} f (1 - \pi) + (1 - \pi) m_0 \right] \Delta_2 = M - m_1 \Delta_1\]

or

\[\left[ 1 - \beta + \pi \beta + \beta f (1 - \pi) + \frac{\beta}{N} (1 - f) (1 - \pi) - m_1 (1 - \pi) \right] \Delta_2 = M - m_1 \Delta_1\]

If we multiply both sides of the above equation by \(1 - \beta + m_0\) and use equation (20) to substitute out for \(\Delta_1\) on the right-hand side, we get

\[f \Delta_2 (1 - \beta + m_0) \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) + \]

\[\Delta_2 (1 - \beta + m_0) \left\{ 1 - \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) - m_1 (1 - \pi) \right\} = ((1 - \beta + m_0)) M - m_1 (1 - (1 - \pi) (\beta - m_0)) \Delta_2\]

or

\[f \Delta_2 (1 - \beta + m_0) \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) + \]

\[\Delta_2 (1 - \beta + m_0) \left( 1 - \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) \right) - \Delta_2 m_1 (1 - \pi) = ((1 - \beta + m_0)) M - m_1 \Delta_2\]

\[19\]
or
\[
f \Delta_2 (1 - \beta + m_0) \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) + \\
\Delta_2 (1 - \beta + m_0) \left( 1 - \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) \right) - \Delta_2 m_1 \pi \\
= (1 - \beta + m_0) M
\]

or

\[
f \Delta_2 \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) + \Delta_2 \left[ 1 - \beta \left( 1 - \frac{1}{N} \right) (1 - \pi) + \frac{m_1 \pi}{(1 - \beta + m_0)} \right] = M. 
\]

In order to derive equation (10), let us use the notation.

\[\hat{\beta} \equiv \beta \left( 1 - \frac{1}{N} \right) (1 - \pi)\]

Since, from (4)

\[\Delta_2 = \frac{\omega}{1 - \pi},\]

we can rewrite the above equation as

\[\omega f \hat{\beta} + \omega \left( 1 - \hat{\beta} + \frac{m_1 \pi}{(1 - \beta + m_0)} \right) = \frac{1 - \pi}{N} (\mu_0 u + (1 - \mu_0) y) + \hat{\beta} \int_0^\omega \omega dF\]

or

\[\omega f \hat{\beta} + \omega \left( 1 - \hat{\beta} + \frac{m_1 \pi}{(1 - \beta + m_0)} \right) = \hat{c} + \hat{\beta} k, \quad (22)\]

where

\[\hat{c} = \frac{1 - \pi}{N} [\mu_0 u + (1 - \mu_0) y].\]

Note that (22) is a function of both \(m_0\) and \(m_1\). We will be able to use (1) and the restriction \(\mu_0 + \mu_1 + \mu_2 = 1\) to eliminate \(m_1\) from this equation, provided that the nonnegativity condition (9) holds. Recognizing that \(m_i = \beta \mu_i / N\), from equation (1), \(m_1\) can be written as

\[m_1 = \frac{\beta}{N} m_0 - \frac{(1 - \pi)}{\pi} f m_0. \quad (23)\]
Substituting (23) into (22) and rearranging, we get

$$\tilde{\omega} \hat{\beta} f + \tilde{\omega} \left\{ 1 - \hat{\beta} + \pi \left( \frac{\beta}{N} - m_0 \right) \right\} - \tilde{\omega} f \frac{(1 - \pi) m_0}{1 - \beta + m_0} = \hat{c} + \hat{\beta} k.$$  

We can rewrite the above equation as

$$\tilde{\omega} f \left\{ \hat{\beta} - \frac{(1 - \pi) m_0}{1 - \beta + m_0} \right\} + \tilde{\omega} \left\{ 1 - \hat{\beta} + \pi \left( \frac{\beta}{N} - m_0 \right) \right\} = \hat{c} + \hat{\beta} k. \tag{24}$$

Letting now

$$a \equiv 1 - \frac{(1 - \pi) m_0}{\beta (1 - \beta + m_0)};$$

$$b \equiv \frac{1 - \hat{\beta}}{\beta} + \frac{\pi \left( \frac{\beta}{N} - m_0 \right)}{\beta (1 - \beta + m_0)};$$

$$c = \frac{\hat{c}}{\beta};$$

equation (10) follows. The proof of Lemma 1 is thus complete. Note that (10) was derived using only value functions and equations (1) and (4).

We now prove Lemma 2. If \( y = 0 \), then a nonnegative solution to the Bellman equation exists only if \( \tilde{\omega} = 0 \) (and thus \( w_i = v_i = 0 \) for all \( i \)). Since \( c = 0 \) for \( y = 0 \), the proof is trivial in this case. Let us assume now that \( y > 0 \), so that \( c > 0 \). Letting

$$h(\tilde{\omega}) \equiv a F(\tilde{\omega}) \tilde{\omega} + b \tilde{\omega} = \tilde{\omega} (a F(\tilde{\omega}) + b);$$

$$g(\tilde{\omega}) \equiv c + \int_0^{\tilde{\omega}} s dF;$$

equation (10) can be compactly represented as \( h(\tilde{\omega}) = g(\tilde{\omega}) \).

**Claim 1** \( h \) is increasing, continuous, \( h(0) = 0 \), \( h(\omega) \geq 0 \) for \( \omega \geq 0 \), and \( h(+\infty) = +\infty \).
Proof. The proof follows from the assumed continuity of $F$ in the benchmark case, from the definition of $h$, and from the fact that $a + b > 1$, since $\beta < 1$ and

$$
\hat{\beta}(a + b) = \frac{1 - \beta + m_0 + \pi \left( \frac{\beta}{N} - m_0 \right) - (1 - \pi) m_0}{1 - \beta + m_0} = \frac{1 - \beta + \pi \beta}{1 - \beta + m_0} > 1.
$$

Claim 2 \text{ \textit{g is nondecreasing, with $g(0) > 0$ and $g(+\infty) < +\infty$.}}

Proof. The proof follows from the fact that $c > 0$ for $y > 0$, and that $F$ has finite mean.

The properties of $h$ and $g$ demonstrated above implies that the equation $h(\omega) = g(\omega)$ has at least one solution on $\mathbb{R}_+$, and, generically, an odd number of solutions.

If $F$ is uniform then $h(\omega) - g(\omega) = (a - \frac{1}{2}) \frac{\omega^2}{\omega_H} + b\omega - c$ for $\omega \in [0, \omega_H]$ and $h(\omega_H) - g(\omega_H) = \omega_H(a + b - \frac{1}{2}) - c$. Since $a + b > 1$, it follows that $h(\omega_H) - g(\omega_H) > 0$ if $\omega_H$ is sufficiently high, or if $y$ (and thus $c$) is sufficiently low. But then, for such values of $\omega_H$ and $y$, the quadratic form of $h - g$ and the fact that $h(0) - g(0) < 0$, imply that there cannot exist two solutions in $[0, \omega_H]$, as stated. This proves Lemma 2.

We now prove Lemma 3. We first derive the expression (11) for the participation constraint for producers (2), and then show that it implies the participation constraint for consumers. Let us write down the participation constraint for producers in terms of $(\mu_0, y, \bar{\omega})$.

For convenience, let $x \equiv \beta - m_0$; recall that

$$
\Delta_1 = \frac{1 - (1 - \pi) x}{1 - x} \Delta_2
$$

and $k \equiv \int_0^\bar{\omega} \omega dF$. We know that

$$
\begin{align*}
w_2 - w_0 &= k + (1 - f) (1 - \pi) \Delta_2 \\
w_1 - w_0 &= k + \Delta_1 - f (1 - \pi) \Delta_2 \\
          &= k + \frac{1}{1 - x} \{1 - (1 - \pi) x - f (1 - \pi) (1 - x)\} \Delta_2
\end{align*}
$$
Consider now the weighted sum
\[ m_1 (w_1 - w_0) + m_2 (w_2 - w_0) = (m_1 + m_2) k + \]
\[ \frac{1}{1 - x} \left\{ \left( \frac{\beta}{N} - m_1 - m_0 \right) (1 - f) (1 - \pi) (1 - x) + m_1 [1 - (1 - \pi) (x + f (1 - x))] \right\} \Delta_2. \]

Let us take a closer look at the \{ \} term;

\[
\{ \} = \left( \left( \frac{\beta}{N} - m_0 \right) (1 - f) (1 - \pi) (1 - x) + m_1 [1 - (1 - \pi) (x + f (1 - x)) - (1 - f) (1 - x) (1 - \pi)] \right) \Delta_2
\]

Now we have
\[ m_1 (w_1 - w_0) + m_2 (w_2 - w_0) = (m_1 + m_2) k + \]
\[ \frac{1}{1 - x} \left\{ \left( \frac{\beta}{N} - m_0 \right) (1 - f) (1 - \pi) (1 - x) + m_1 \pi \right\} \Delta_2 \]

Since \( \left( \frac{\beta}{N} - m_0 \right) = m_1 + m_2 \), the participation constraint becomes
\[ y \leq \beta k + \beta \left\{ (1 - f) (1 - \pi) + \pi \frac{m_1}{m_1 + m_2} \frac{1}{1 - \beta - m_0} \right\} \Delta_2. \quad (25) \]

Note that
\[ m_1 + m_2 = \frac{\beta}{N} - m_0; \quad (26) \]
from the flow equation \( \pi m_2 = (1 - \pi) m_0 f \),
\[ m_1 = \frac{\beta}{N} - m_0 - \frac{1 - \pi}{\pi} m_0 f; \quad (27) \]
and
\[ \Delta_2 = \frac{\omega}{1 - \pi}. \quad (28) \]
Substituting (26), (27) and (28) into (25) we get, after some algebra,

\[ y \leq \beta k + \beta \left\{ \pi \left( \frac{\beta}{N} - m_0 \right) + (1 - \pi) \left( 1 - \beta + m_0 \right) \left( \frac{\beta}{N} - m_0 \right) \right\} \]

\[ - (1 - \pi) f \left[ \frac{\beta}{N} - \left( \frac{\beta}{N} - m_0 \right) \left( \beta - m_0 \right) \right] \cdot \frac{\bar{\omega}}{1 - \pi} \frac{1}{1 - \beta + m_0} \frac{1}{\frac{\beta}{N} - m_0}, \]

which defines the expression for inequality (11). The proof of Lemma 3 is now complete.

We now show that satisfaction of (11) implies satisfaction of the participation constraint for consumers, (3). These inequalities assure that trade provides a nonnegative flow of expected utility. In the case of the producer, it is equivalent to \( v_0 \geq \beta w_0 \), as the Bellman equation shows. That is, \( v_0 \) is bounded below by the option of not producing, which provides discounted utility \( w_0 \). Likewise, buyers holding genuine money, face the lower bound \( v_1 \geq \beta w_1 \), while those holding counterfeits face \( v_2 \geq \beta w_2 \). In addition, as demonstrated above,

\[ w_1 - w_0 = k + \Delta_1 - f(1 - \pi) \Delta_2 \geq k + \Delta_2 - f(1 - \pi) \Delta_2 - \pi \Delta_2 = w_2 - w_0 \]

so that, the relevant inequality in (3) is \( u \geq \beta(w_1 - w_0) \), which is equivalent to \( v_1 \geq \beta w_1 \). Since \( w_1 = v_1 \), it suffices to show that \( v_0 \geq \beta w_0 \)—which is the producer participation constraint—implies \( v_1 \geq 0 \)—which would satisfy the consumer participation constraint. Since \( w_2 - v_0 = (1 - \pi)(v_2 - v_0) = \bar{\omega} \geq 0 \), then if \( v_0 \geq 0 \), then \( v_2 \geq 0 \); but this implies, by \( v_1 - v_0 = \frac{1 - (1 - \pi)(\beta - m_0)}{1 - \beta + m_0} (v_2 - v_0) \), that if \( v_0 \geq 0 \), then \( v_1 \geq 0 \). So if \( w_0 \geq 0 \)—which implies \( v_0 \geq 0 \) by the producer participation constraint—then \( v_1 \geq 0 \)—which means the consumer participation constraint is satisfied. So all we need to demonstrate is that \( w_0 \) is nonnegative. From the Bellman condition (13), we have \( w_0 = \int_0^\omega (w_2 - \omega) dF + (1 - f) v_0 \); but since \( \bar{\omega} = w_2 - v_0 \), this Bellman condition can be rewritten as \( w_0 = \int_0^\omega (\bar{\omega} - \omega) dF + v_0 \). Therefore, the producer participation constraint, \( v_0 \geq \beta w_0 \), implies \( (1 - \beta)w_0 \geq \int_0^\omega (\bar{\omega} - \omega) dF \geq 0 \), which means that \( w_0 \geq 0 \). We are now ready to complete the proof of Proposition 1.

Since \( F(\omega) > 0 \) for all \( \omega > 0 \), then any implementable allocation must feature \( \mu_2 > 0 \) unless \( y = \bar{\omega} = 0 \). Since \( u \) is concave and \( u'(0) = +\infty \) then, for a fixed \( \mu_0 \), any \( y \) positive but sufficiently small satisfies the producer participation constraint (29). (Note that \( \frac{\bar{\omega}}{1 - \pi} = \Delta_2 \) and \( \Delta_2 \) is a function of \( u(y) \).
see equations (18) and (21). Since \( F'(0) < +\infty \) and \( u'(0) = +\infty \) then, for \((\mu_0, y)\) admissible and \( y \) sufficiently small, \( \frac{1}{N}\mu_0(1-\mu_0)[u(y)-y] > \mu_0 \int_0^{\tilde{\omega}} \omega dF \), demonstrating that any optimum must feature positive counterfeiting.

We now investigate the relationship between \( \bar{\omega} \) and \((\mu_0, y)\). Since

\[
\hat{\beta} a = \hat{\beta} - \frac{1 - \pi}{\frac{1 - \beta}{m_0} + 1}; \quad \frac{\partial a}{\partial \mu_0} < 0;
\]

since

\[
\hat{\beta} b = 1 - \hat{\beta} + \pi \left( \frac{\beta}{N} - m_0 \right); \quad \frac{\partial b}{\partial \mu_0} < 0;
\]

since

\[
\hat{\beta} c = \frac{1 - \pi}{N} (\mu_0 u(y) + (1 - \mu_0) y), \quad \frac{\partial c}{\partial \mu_0} > 0 \text{ and } \frac{\partial c}{\partial y} > 0.
\]

For a fixed \((\omega, y)\) then \( \frac{\partial h}{\partial \mu_0} = \frac{\partial (\omega (aF + b))}{\partial \mu_0} < 0 \) and \( \frac{\partial a}{\partial \mu_0} = \frac{\partial (c+k)}{\partial \mu_0} > 0 \). For a fixed \((\omega, \mu_0)\) then \( \frac{\partial h}{\partial y} = 0 \) and \( \frac{\partial a}{\partial y} > 0 \). Now, if \( \beta \) is sufficiently high and the participation constraint for producers can be ignored then optimality requires that for a fixed \((\mu_0, y)\) the smallest solution (if there is more than one) to \( h(\omega) = g(\omega) \) is chosen. For such \( \omega \), the function \( h \) cuts the function \( g \) from below so that, given these derivatives, \( \frac{\partial \omega}{\partial \mu_0} > 0 \) and \( \frac{\partial \omega}{\partial y} > 0 \). Since \( \mu_0(1-\mu_0)[u(y)-y] \) is maximized by \((\mu_0, y) = (\frac{1}{2}, y^*)\) and \( \mu_0 \int_0^{\tilde{\omega}} \omega dF \) decreases with reductions in \((\frac{1}{2}, y^*)\), the proof follows.

A2 Degeneracy and the proof of Proposition 2

We now assume that all individuals draw the same \( \omega = \tilde{\omega} > 0 \). There are three kinds of candidates for the optimum. The first candidate optimum has \( \omega < \tilde{\omega} \), so that \( \mu_2 = 0 \) and \((\mu_0, y)\) solves a “relaxed problem,” which is described in appendix A3, below. The second candidate optimum has \( \omega = \tilde{\omega} \) with \( \mu_2 = 0 \). The \((\mu_0, y)\) from the second candidate will (probably) not solve the relaxed problem; that is, the solution to the relaxed problem will generate a \( \omega > \tilde{\omega} \), which implies that everyone without money will counterfeit. The best way to think about the second candidate is that it solves the relaxed problem subject to the additional constraint that \( \omega \leq \tilde{\omega} \).

The third candidate optimum has \( \omega > \tilde{\omega} \) and \( F(\tilde{\omega}) = 1 \). This is possible provided that the nonnegativity constraints for \( \mu \) and \((w, v)\) are satisfied. The former imposes an upper bound on \( \mu_0 \) given by \( \mu_0 \leq \pi \) to ensure a nonnegative \( \mu \), (e.g., it follows that if \( \mu_0 = \pi \), then \( \mu_2 = 1 - \pi \) and \( \mu_1 = 0 \),
and that $\mu_1 > 0$ only if $\mu_0 < \pi)$. The latter imposes lower bounds on the values for $(\mu_0, y)$ to ensure that $\bar{\omega} > \tilde{\omega}$. Although the allocations of the third-candidate kind must satisfy additional constraints that the two other kinds of candidates do not and there is a welfare loss associated with counterfeiting, one would think that the planner would never choose this kind of allocation. We are, however, unable to prove that the third candidate optimum is dominated in welfare terms by either of the first two candidate optima. The difficulty here is due to the fact that, as shown below, for a fixed $\mu$, the right-hand side of the participation constraint for producers, (2), increases by $\beta \bar{\omega}$ when counterfeiting is introduced. An increase in $\bar{\omega}$ beyond $\tilde{\omega}$ has two effects that cannot, in principle, be unambiguously ranked. Thus it is possible that the constraint (2) changes from active to inactive as $\bar{\omega}$ is raised from $\tilde{\omega}$.

Using the characterization provided by the remark (on the relaxed problem) below, a sufficient condition for optimality of allocations of the first kind is that the restriction $\bar{\omega} \leq \tilde{\omega}$ is not binding, where $\bar{\omega}$ is computed according to the Bellman equations with $\mu_0 = \mu_1 = \frac{1}{2}$ and $u'(y) = 1$. Due to continuity, reductions in the value of $\bar{\omega}$ below a threshold imply that the optimum becomes of the second kind.

The value for $\bar{\omega}$ in equation (7) follows from the condition $h(\bar{\omega}) = g(\bar{\omega})$ derived above, say equation (22), after $f = 1$ and $k = \tilde{\omega}$ are imposed. Likewise, the inequality for $y$ results from (29).

The proof for the first part of Proposition 2 is straightforward. If $\pi \to 1$, then from (7), $\bar{\omega} \to 0$, and, hence, $\bar{\omega} > \tilde{\omega}$, for $\pi$ sufficiently high. Similarly, if $\pi \to 0$, then, from (8), $y \to \beta \tilde{\omega}$ and from (7), $\bar{\omega} \to \beta \tilde{\omega}$; hence, $\bar{\omega} > \tilde{\omega}$.

To complete the proof (for the second part of proposition 2), note that the restrictions on $\mu$ follows from (1) and (9). Provided that $\bar{\omega} > \tilde{\omega}$, the necessity and sufficiency of the conditions for implementability follows from Lemma 3.

A3 The relaxed planner’s problem

Claim 3 Assume that $F(\omega) = 0$ for $\omega \leq \omega_L$ for a sufficiently large $\omega_L$. Then an implementable $(\mu, y, \bar{\omega})$ is optimal if and only if its margin $(\mu_0, y)$ maximizes the relaxed problem

$$\frac{1}{N} \mu_0(1 - \mu_0)[u(y) - y]$$ (30)
subject to

\[ y \leq \frac{u(y)}{1 + \frac{1 - \beta}{\beta} N}. \] (31)

If, in addition, \( \beta \) is sufficiently high, then the optimum features \( \mu_0 = \frac{1}{2} \) and \( u'(y) = 1 \).

The representation (30) of the planner’s objective function, as well as the derivation of the constraint (31), have been derived elsewhere (see Cavalcanti and Wallace (1999) and Cavalcanti and Nosal (forthcoming)) for the basic case without counterfeiting. The derivation of inequality (31) for the basic case follows by imposing \( \mu_2 = 0 \), using the Bellman equations to solve for \( w_1 - w_0 \) as a function of \( \mu_0 \) and \( y \), and then obtaining an expression for the producer’s constraint (11) precisely as (31). The sufficiency of (31) for implementability follows because \( F(\omega) = 0 \) is a necessary condition for optimality when \( \omega_L \) is large, as discussed below, and because (11) implies (2) and (3), according to Lemma 3.

Since \( u'(+\infty) = 0 \) then \( u(y) < y \) and welfare falls below zero if \( y \) is sufficiently large. Since the welfare associated to autarky is zero, the planner’s problem can be restricted, without loss of generality, to bounded output and bounded discounted utilities. As a result, if \( F(\omega_L) = 0 \) for \( \omega_L \) sufficiently high, then an optimum must feature a \( \omega \) such that \( F(\omega) = 0 \). We now show that (29) is equivalent to (31) when \( F(\omega) = 0 \).

Imposing \( F(\omega) = 0 \) in the condition \( h(\omega) = g(\omega) \) above provides a solution \( \omega = \frac{c}{b} \), which can be substituted in (29), together with \( k = f = 0 \), yields (31). The following algebra steps are useful for the task. With \( k = f = 0 \), (29) reads

\[ y \leq \beta \frac{\pi + (1 - \pi)(1 - \beta + m_0)}{(1 - \pi)(1 - \beta + m_0)} \omega. \]

Before substituting for \( \omega = \frac{c}{b} \), it is useful to write
\[ \hat{\beta} = 1 - \hat{\beta} + \frac{1}{N} \pi \beta (1 - \mu_0) \]
\[ = \frac{1 - \beta + \frac{\beta}{N} \mu_0 + \pi \frac{\beta}{N} (1 - \mu_0)}{1 - \beta + m_0} - \hat{\beta} \]
\[ = \frac{1 - \beta + \frac{\beta}{N} (1 - \pi) \frac{\beta}{N} \mu_0}{1 - \beta + m_0} - \hat{\beta} \]
\[ = \frac{\pi (1 - \beta) + (1 - \pi) (1 - \beta + m_0) + \frac{\pi \beta}{N}}{1 - \beta + m_0} - \hat{\beta} \]

so that

\[ \hat{\beta} (1 - \beta + m_0) = \pi + (1 - \pi) (1 - \beta + m_0) - \pi \beta \left( 1 - \frac{1}{N} \right) - \]
\[ (1 - \pi) \beta \left( 1 - \frac{1}{N} \right) (1 - \beta + m_0) \]
\[ = \pi + (1 - \pi) (1 - \beta + m_0) - \beta \left( 1 - \frac{1}{N} \right) [\pi + (1 - \pi) (1 - \beta + m_0)] \]
\[ = \left[ 1 - \beta \left( 1 - \frac{1}{N} \right) \right] [\pi + (1 - \pi) (1 - \beta + m_0)] \]

and then proceed with the substitution in order to derive (31).

It remains to be shown a description of the optimum when \( \beta \) is sufficiently large. It follows that (31) is slack when \( \mu_0 = \frac{1}{2} \), \( y \) is such that \( u'(y) = 1 \) and \( \beta \) approaches one. The proof of the claim is now complete.

As a final remark, consider the relaxed problem together with the additional constraint \( \bar{\omega} \leq \bar{\omega} \). Consider also the solution \( \bar{\omega} \) for \( h(\bar{\omega}) = g(\bar{\omega}) \), i.e., equation (10), when \( F(\bar{\omega}) = 0 \), derived above,

\[ \bar{\omega} \]..
Since $\pi$ does not appear in the objective or in the participation constraint, and since $\tilde{\omega} > 0$, it follows that the solution attains a welfare level that is positive and bounded away from zero.

By contrast, by Proposition 2, implementable allocations with $\tilde{\omega} > \tilde{\omega}$ and positive counterfeiting features $\mu_0 \leq \pi$, and thus their extensive margins are driven towards zero when $\pi$ is reduced. Allocations of this kind either fail to become implementable or yield welfare below the allocation without counterfeiting (the relaxed problem with $\tilde{\omega} \leq \tilde{\omega}$) when $\pi$ is reduced.