Counterfeiting as Private Money in Mechanism Design

Ricardo Cavalcanti — Getulio Vargas Foundation
Ed Nosal — Cleveland Fed

November 2006
Preliminary and Incomplete

Abstract

We describe counterfeiting activity as the creation of inside money that cannot be recognized and which is costly to produce. We provide a welfare analysis of currency competition which ammends the basic random-matching model of inside money in mechanism design. We show that it is not efficient to eliminate conterfeiting activity completely when the costs of producting counterfeits differ amount agents, and that the optimal return of outside money should be reduced relative to the typical case when counterfeiting is ignored. The existence of counterfeits notes will also depress economy activity.

1 Introduction

The counterfeiting of notes is a pervasive phenomenon in monetary economies and cannot be eliminated by advances in technological progress, such as ones that tend to increase detection rates, or increase the production costs of fake means of payments. Counterfeiting, like theft, is costly to the economy. There are both direct costs and indirect costs. The direct costs represents the diversion of resources from productive activities to counterfeiting and the indirect costs are individual's response to counterfeiting.

In recent models of counterfeiting, for example Nosal and Wallace (2005), it is the *threat* of counterfeiting that imposes costs on the economy. In equilibrium, however, counterfeiting never occurs. This result is problematic and we suspect it may be related to the game theoretic approach taken. That is, by imposing a specific game form that agents must play—and in particular, the form of the bargaining game that buyers and sellers play—one may inadvertently be ruling out the possibility of the emergence of counterfeit notes.

In this paper, we adopt a mechanism design approach. This approach is quite flexible in terms of modeling how agents interact, in that the mechanism selects a game form that maximizes society's welfare. We believe that a mechanism design approach to counterfeiting can be fruitful since private money has been selected an optimal instrument in a number of recent papers, e.g., Cavalcanti and Wallace (1999). We view counterfeit notes as being a form of private money, albeit one that the mechanism may want to limit in its circulation.

We find that in order to have circulating counterfeit notes as part of the optimal mechanism, there must be heterogeneity of opportunities to create and circulate counterfeit among agents. When such heterogeneity exists, we find that counterfeiting creates distortions at both the intensive and extensive margins. That is, output will tend to "low" and the supply of money will tend to be "high," compared to an environment where counterfeiting is not possible. When there is no heterogeneity in opportunities to create and circulate counterfeit notes, then, like in Nosal and Wallace (2005), although the threat of counterfeiting has negative implications for welfare, the optimal mechanism will not allow counterfeit notes to circulate.

The rest of the paper is as follows. Section 2 presents the environment and describes what an allocation is. Some preliminary, but important, results are presented in section 3. All the main results are collected in section 4. Section 5 concludes. All proofs are relegated to an appendix.

2 The benchmark money-competition problem

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 media of exchange. We will refer to this alternative media of exchange 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 is $u(x)-y$ if he does not, where x is the amount of good i consumed, y is the amount of good $i+1$ produced and ω is the utility cost of producing a counterfeit note. The function u is continuous, concave, differentiable, with $u(0)=0$, $u'(0)=+\infty$, $u'(+\infty)=0$. 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 a 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 $\mu_1 \in (0,1)$. Fiat money is perfectly durable. At the time when a counterfeit note is produced, it is indistinguishable from a fiat note. Over time, however, a counterfeit note may be marked or altered in a way that clearly distinguishes it from a fiat note.

Each period has two sub-periods. At the beginning of the first sub-period, individuals draw a realization for the cost of counterfeiting $\omega \in \{\omega_L, \omega_H\}_+$. The cost of counterfeiting, ω, is identically and independently distributed overtime and across individuals according to the cumulative distribution function F. The function F is assumed to be continuous and differentiable, and satisfies $F(\omega_L)=0$. After individuals learn their current period ω, 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—is permanently marked or identified as being a counterfeit note. We assume that individuals who hold marked counterfeit notes destroy them before entering the second subperiod; hence, these individuals enter the second sub-period holding no money. This ends the first sub-period. (In the next section, We will demonstrate below that the assumed destruction of a marked counterfeit note is without loss of generality.)

In the second sub-period, 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. If trade occurs, then the seller learns about the type of money—fiat or counterfeit—that he just acquired and this information remains private.

The distribution of money at the beginning of the first subperiod is given by $\mu = (\mu_0, \mu_1, \mu_2)$, where μ_0 represents the fraction of individuals holding no money, μ_1 is the fraction of individuals holding a fiat note and μ_2 is the fraction of individuals holding a counterfeit note. The beginning of the first sub-period 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 sub-period value functions—which are evaluated after the counterfeiting cost and counterfeiting mark realizations, but before individuals are matched—are denoted by $v = (v_0, v_1, v_2)$.

The planner's problem is to maximize the average ex-ante utility of individuals, where the average taken with regard to the endogenous distribution of individuals as indexed by their money holdings. The planner's maximization problem must respect participation constraints, which are dictated by individual rationality. As in Cavalcanti and Wallace (1999), we assume that individuals cannot coordinate on defection, so participation constraints only reflect the possibility of an individual defection. (We leave to the conclusion some comments about group defection.)

The planner can be interpreted as choosing $\mu_1 \in [0, 1]$, the fraction of individuals who hold fiat money. 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 sub-period, $\mu = (\mu_0, \mu_1, \mu_2)$, also describes the distribution of money holdings at the beginning of the second sub-period.

We will focus on equilibria where fiat and (unmarked) counterfeit notes trade for the same quantity of output, y. 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 equilibrium holders of money never dispose of their monies in the first sub-period: there is no point for a holder of a counterfeit note to dispose of it at the beginning of the first sub-period only to (possibly) produce another one that has an identical chance of being detected; since a fiat note has no chance of being detected as a counterfeit note and, as well, has value, a holder of a fiat note will never dispose of it. We assume that allocations are stationary and symmetric across consumption/production types and states and define a *symmetric, stationary and pooling allocation* to be the list $(\mu, y, \bar{\omega})$, where money holdings are distributed according to μ, both monies trade for y—the level output produced and consumed in single-coincidence meetings—and counterfeit money is created according to $\bar{\omega}$.

2.2 Implementable allocations

Allocation $(\mu, y, \bar{\omega})$ is implementable if it satisfies individual participation or individual rationality constraints. Individual rationality for the producers requires that

$$y \leq \beta[\frac{\mu_1}{\mu_1 + \mu_2}(w_1 - w_0) + \frac{\mu_2}{\mu_1 + \mu_2}(w_2 - w_0)], \tag{1}$$

since the producer can only use his knowledge about μ 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 with money $i = 1, 2$ compared to starting with no money at all. The bracketed term on the right-hand side of (1) 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 β. The left-hand side of (1) 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 ouput, individual rationality for the consumer is simply

$$u(y) \geq \beta \max\{w_1 - w_0, w_2 - w_0\}. \tag{2}$$

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 (2), exceeds the cost, which is given by the right-hand side (2).

Finally, there is an individual rationality constraint associated with the production of counterfeit notes. 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 π is the probability that a counterfeit note will be marked. Therefore, there exists a the cutoff value for ω that satisfies

$$\bar{\omega} = (1 - \pi)(v_2 - v_0) \tag{3}$$

and, assuming that the individual does not have a unit of money at the beginning of the first sub-period, the individual rationality constraint associated with the production of a counterfeit note is,

$$\omega \leq \bar{\omega}; \tag{4}$$

for $\omega > \bar{\omega}$, an individual without money will choose to be a seller in the second subperiod.

The allocation $(\mu, y, \bar{\omega})$ is implementable if there exists (w, v) that satisfy the participation constraints (1), (2), and (4), as well as the standard Bellman equations (which are described in the appendix).

3 Preliminary Results

An environment that admits the possibility of counterfeiting has a higher dimensionality than that of a standard monetary environment. In this section, we show how one is able to convert the current problem with counterfeiting into a standard one.

An allocation is denoted by $(\mu, y, \bar{\omega})$. The inclusion of "$\bar{\omega}$" in the definition of an allocation is non-standard but required because $\bar{\omega}$ describes the amount of counterfeit money that will be produced. In the standard planning problem—one without a counterfeiting technology—the planner chooses the fraction of agents who do not have money, μ_0, and the level of output to be produced in single coincidence meetings, y. In the planning problem with counterfeiting, suppose that the planner selects the pair (μ_0, y). In the steady state, the amount of counterfeit notes that are produced at the beginning of the first subperiodnd, $\mu_0 F(\bar{\omega})$, must equal the amount that is destroyed (marked) at the end of the first subperiod, $\pi \mu_0 F(\bar{\omega}) + \pi \mu_2$. Since $\mu_0 + \mu_1 + \mu_2 = 1$, for a given μ_0, in a steady state we have

$$\mu_2 = \frac{(1 - \pi) F(\bar{\omega})}{\pi} \mu_0 \tag{5}$$

and

$$\mu_1 = 1 - \left[1 + \frac{(1 - \pi) \mu_0 F(\bar{\omega})}{\pi}\right] \mu_0 \tag{6}$$

The following proposition explains how the critical cost of producing a counterfeit, $\bar{\omega}$, is determined for a given (μ_0, y).

Proposition 1 *If $(\mu, y, \bar{\omega})$ is implementable then $\bar{\omega}$ solves*

$$\bar{\omega}F(\bar{\omega})a(\mu_0) + \bar{\omega}b(\mu_0) = c(\mu_0, y) + d\int_{w_L}^{\bar{\omega}} \omega dF \tag{7}$$

where a, b, c and d are positive and continuous functions of parameters, with a and b increasing in μ_0, and c increasing in μ_0 and y. In addition, for a fixed (μ_0, y), condition (7) can be written as $h(\bar{\omega}) = g(\bar{\omega})$, where h and g are continuous, $g(0) > h(0) = 0$, and $h' > g' + e$, where e is a positive function of parameters, but independent of (μ_0, y).

Conversely, if $(\mu, y, \bar{\omega})$ satisfies (7) then there exists (w, v) satisfying the individual Bellman equations and the participation constraint (3). If, in addition, β is sufficiently high then $(\mu, y, \bar{\omega})$ is implementable.

Proof. See appendix. ■

We show that the cut-off value $\bar{\omega}$ associated to (μ_0, y) in an implementable allocation is uniquely defined.

Proposition 2 *Let an arbitrary pair (μ_0, y) of an allocation be fixed. Then there exists on odd number of solutions to equation (7). If, $F(\omega)$ has a uniform distribution, then the solution to (7) exists and is unique.*
Proof. *See appendix.* ■

3.1 Main Results

Before we proceed to the main results, we will describe planner's problem.

3.2 The planner's problem

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

$$W(\mu, y, \bar{\omega}) = \frac{1}{N}\mu_0(1-\mu_0)[u(y) - y] - \mu_0 \int_{\omega_L}^{\bar{\omega}} \omega dF. \tag{8}$$

Equation (8) 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 (consumer) is matched with a seller is who produces the good that he desires is $\frac{1}{N}$ times μ_0 and the total measure of potential consumers 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, $\mu_0 \int_{\omega_L}^{\bar{\omega}} dF$, who is choose produce counterfeit bills, where the total cost of counterfeiting these bills is $\mu_0 \int_{\omega_L}^{\bar{\omega}} \omega dF$.

If allocation $(\mu, y, \bar{\omega})$ is implementable, then it must satisfy the producer's participation constraint (1), be written as a function of (μ_0, y) and $\bar{\omega}$

$$y \leq \beta \int_{\omega_L}^{\bar{\omega}} x dF(x) + \beta \left\{ \pi \left(\frac{\beta}{N} - m_0 \right) + (1-\pi)(1-\beta+m_0) \left(\frac{\beta}{N} - m_0 \right) \right. \tag{9}$$
$$\left. - (1-\pi) F(\bar{\omega}) \left[\frac{\beta}{N} - \left(\frac{\beta}{N} - m_0 \right) (\beta - m_0) \right] \right\} \frac{\bar{\omega}}{1-\pi} \frac{1}{1-\beta+m_0} \frac{1}{\frac{\beta}{N} - m_0},$$

where $m_0 = \beta\mu_0/N$. (Inequality (9) is derived in the appendix.)

3.3 The "Threat" cost of counterfeiting

We define the threat cost of counterfeiting, denote ω_T, as that critical cost such that if $\omega_T = \bar{\omega}$, then no agent would have an incentive to counterfeit. Of course, this can only happen if $\omega_T \leq \omega_L$. From (7), we see that

$$\omega_T(\mu_0, y) = \frac{c(\mu_0, y)}{b(\mu_0)}.$$

One practical use of the threat cost of counterfeiting, ω_T, is that it allows us to compare, in a rather straightforward manner, the solution of the mechanism-design problem with no counterfeiting, i.e., the standard environment, to one with counterfeiting.

3.4 Main Results

Proposition 3 *Suppose $\omega_L \geq \omega_T\left(\frac{1}{2}, y^*\right)$, then for β sufficiently high*

$$y^R = y^* \text{ and } \mu_0^R = \frac{1}{2}.$$

Now suppose $\omega_L < \omega_T\left(\frac{1}{2}, y^*\right)$*, then for* β *sufficiently high*

$$y^C < y^* \text{ and } \mu_0^C < \frac{1}{2}$$

Remark 4 *This proposition states that if the first-best solution is attainable both in environment where counterfeiting does not occur in equilibrium, i.e., when* $\omega_L \geq \omega_T\left(\frac{1}{2}, y^*\right)$*, and where counterfeiting does occur, i.e., when* $\omega_L < \omega_T\left(\frac{1}{2}, y^*\right)$*, then the first best allocation will the choosen by the planner in the economy without counterfeiting and, in the economy with counterfeiting, the level of output and the measure of people with without money will both be lower than the "first best" levels.*

Proof. See appendix. ■

Up to this point we have assume that agent's cost of counterfeiting is iid, where $\omega \in \{\omega_L, \omega_H\}$. We now consider the case where $\omega_L = \omega_H$.

Proposition 5 *Suppose that* $\omega_L = \omega_H$*, then (i) if* $\omega_L \geq \omega_T\left(\frac{1}{2}, y^*\right)$ *and* β *is sufficiently high, then* $y^C = y^*$ *and* $\mu_0^C = \frac{1}{2}$ *and (ii) if* $\omega_L < \omega_T\left(\frac{1}{2}, y^*\right)$ *and* β *is sufficiently high, then* $y^C < y^*$*,* $\mu_0^C < \frac{1}{2}$ *and* $\bar{\omega}\left(\mu_0^C, y^C\right) = \omega_L$*.*

Remark 6 *This proposition states that if the cost of counterfeiting is sufficiently high, i.e.,* $\omega_L \geq \omega_T\left(\frac{1}{2}, y^*\right)$*, then the first best allocation is achievable if* β *is sufficiently high. However, even if* β *is sufficiently high, if the cost of counterfeiting is sufficiently low, i.e.,* $\omega_L < \omega_T\left(\frac{1}{2}, y^*\right)$*, then (i) both output and the measure of people without money will be below first best levels and (ii) there will be no counterfeiting in equilibrium.*

Proof. See appendix. ■

4 Conclusion

To be added.

5 References

To be added.