A Resolution of the National Banking System Note-issue Puzzle*

Bruce Champ and Neil Wallace†

November 5, 2001

1 Introduction

Under the National Banking System (NBS), in effect from 1863–1914, National Banks that deposited sufficient collateral could issue notes provided they paid a tax on notes in circulation: 1% per year prior to 1900 and $\frac{1}{2}$% thereafter (see, for example, Friedman and Schwartz [5], pp 20–23). The simple and predominant view of this system uses the observation that note issue was far below the allowed maximum and an arbitrage argument to predict that the system should have produced a low and constant upper bound on short-term nominal interest rates—a bound equal to the tax rate on notes outstanding (see, in particular, Champ, Wallace, and Weber [3]).[1] Unfortunately, as is well known and documented in [3], data on interest rates contradict that prediction: the data suggest the presence of arbitrage profits. That is the note-issue puzzle.

*This is a preliminary version. Please do not cite without permission of the authors.

†Champ, Research Department, Federal Reserve Bank of Cleveland <bruce.a.champ@clev.frb.org>; Wallace, Department of Economics, Penn State University <neilw@psu.edu>.

[1] The easiest way to understand the arbitrage argument is by way of an analogy to a central bank discount window that makes loans on demand at an interest rate equal to the tax rate subject to a collateral requirement. So long as eligible collateral exceeds the amount used as collateral, the collateral requirement does not impose a cost. Therefore, such a system produces a perfectly elastic supply of money at an interest rate (on safe loans) equal to the tax rate—perfectly elastic up to the quantity at which all eligible collateral is used as collateral.

Champ, Wallace, and Weber suggested a route to resolving the puzzle. They said that the questionable ingredient in the arbitrage argument is the assumption that banks could earn the market rate of interest on any quantity of notes they chose to issue. In particular, the rules concerning note issue and redemption imply that issuing notes to buy securities in organized markets was unprofitable because such notes would be presented for redemption in a matter of days and the redemption processing costs, which were charged to the issuing bank, would more than offset the interest earned even at interest rates far in excess of the tax rate.[2] In this paper, we pursue that suggestion.[3]

Our resolution of the note-issue puzzle takes the form of a model, a variant of the Cavalcanti-Wallace model of inside money (see [1] and [2]), in which notes play a role, but not in the centralized (intertemporal) market of the model. We show that the model is consistent with a steady state in which note issue is not maximal even though the tax rate on notes is lower than the interest rate in the centralized market.

Although the model contains many extreme assumptions and omits many aspects of the NBS economy, we intentionally keep it close to the original Cavalcanti-Wallace model. By doing so, we are demonstrating that minimal changes in an off-the-shelf model that was designed for other purposes can go quite far in resolving the puzzling features of the NBS economy.

2 The model

The Cavalcanti-Wallace model uses a version of the background environment of Shi [6] and Trejos and Wright [7]. Time is discrete and the horizon is infinite. There are $S > 2$ perishable goods at each date and a $[0, 1]$ continuum of each of S types of people. A type s person consumes only good s

[2]Friedman and Schwartz hint at aspects of this when they say "An issuing bank ... had no way of identifying banks that returned its notes to the Treasury for redemption; hence its New York City correspondents could do so with impunity." ([5], footnote 8, page 21.)

[3]Redemption costs are also the subject of [4]. However, they do not build on the central idea of this paper, which is that a bank could influence how quickly notes would be presented for redemption by the use to which it put newly issued notes. Their approach, which delivers a version of the arbitrage argument adjusted for *average* costs and rates of redemption, is not consistent with the conclusion that any direct attempt to arbitrage by using notes to buy securities in financial markets was almost certain to generate a loss.

and is able to produce only good $s+1$ (modulo S). Each person maximizes expected discounted utility with discount parameter $\beta \in (0,1)$. The period utility function is $u(x) - y$, where x is consumption of the relevant good and y is production of the relevant good. The function u is strictly concave and increasing, and satisfies $u(0) = 0$, $u'(0) = \infty$, and $u'(\infty) = 0$.

Cavalcanti and Wallace split the $[0,1]$ continuum of each type into two intervals. The interval $[0,A]$ consists of those whose previous actions are perfectly monitored, and, therefore, are common knowledge. We call them the "known" people. The rest, the interval $(A,1]$, are not monitored at all so their previous actions are private. We refer to them as the "unknown" people. The parameter A can be interpreted as society's monitoring capacity. A person's specialization type and known-unknown status are common knowledge and people cannot commit to future actions.

In [1], each known person has a printing press. The printing press turns out uniform, indivisible, and perfectly durable objects called notes. The notes of any person can be distinguished from those of any other person. These notes are the only durable assets. Finally, each person can carry from one date to the next at most one note. We maintain these assumptions. In the original version of the model, each person meets one person per date at random. So, for example, an unknown person meets another such person with probability $1-A$ and meets a known person with probability A. Here, we change the meeting pattern and timing.

We divide each date into two subdates and assume additive separability of preferences and discounting between one subdate and the next. At the first subdate, unknown people meet others at random as in the original model, but known people never meet each other. This last assumption simplifies the model a bit and seems innocuous for our purposes. At the second subdate, unknown people meet no one, while known people are all together in a centralized meeting. This second subdate serves two functions. First, at this subdate the taxes on outstanding notes are paid by those who issue notes and redistributed lump-sum among all the known people. Second, competitive trade among the known people at this subdate provides an interest rate.

We limit consideration to pure strategy steady states and to allocations which do not depend on the specialization type, s, and which are simple in other respects that we now describe.

2.1 The steady-state stock of notes

Let λ denote the fraction of known people who participate in the note-issuing scheme. All conditions are written for $\lambda > 0$. The non-participants, the fraction $1-\lambda$ of the known people, use notes of participants when they meet unknown people in the first subdate. Aside from the tax on notes, this requirement is the main conseqence of being a non-participant. A participant can issue a new note in any meeting with an unknown person, but a non-participant must have acquired a note in an earlier trade. Also, participants must redeem notes, but non-participants do not. A steady state has a constant stock of notes in the hands of non-participants and a constant stock in the hands of unknown people. We let x_{kj} denote the fraction of people who are non-participants and have j notes and we let x_{uj} denote the fraction of people who are unknown and have j notes, $j = 0, 1$. Because $x_{k0} + x_{k1} = (1 - \lambda)A$ and $x_{u0} + x_{u1} = 1 - A$, it is enough to equate net inflows and net outflows into one of the states.

One of our simplifying assumptions is that trade occurs in any single-coincidence meeting in which the producer can be given a note without violating the upper bound on holdings. Subject to the condition that there are no gifts, such a trade pattern trade gives rise to as much trade as is possible given the assumptions about production-consumption opportunities and the unit bound on note holdings.

For non-participants, the implied inflow-equal-outflow condition for state 1 is equivalent to

$$x_{k0}x_{u1} = x_{k1}x_{u0}. \tag{1}$$

The inflow arises in single-coincidence meetings with unknown consumers who have a note and the outflow in single-coincidence meetings with unknown producers who do not have a note. (Recall that known people do not meet each other at the first subdate.)

For unknown people, the inflow-equal-outflow condition for state 1 is equivalent to

$$x_{u0}(\lambda A + x_{k1}) = x_{u1}(\lambda A + x_{k0}). \tag{2}$$

Unknown people without a note get a note in two kinds of meetings: whenever they produce for a participant and whenever they produce for a non-participant with a note. Outflows are analogous.

It we subtract (1) from (2), then we get $x_{uo} = x_{u1}$. Then (1) implies $x_{k0} = x_{k1}$. That is, given that notes are traded in all those meetings, the unique steady state is that half of unknown people have a note and half do not and similarly for non-participants.

2.2 The first subdate

Cavalcanti and Wallace study somewhat general allocations in which notes can have one value in terms of goods when issued, another value when traded among unknown people, and still another value when redeemed. They also allow for the possibility that known people are forced to give gifts of output or notes. Here, we simplifiy matters by building in the observation that NBS notes always traded at par and were redeemed at par and that it was not part of the system to force known people to give gifts. Therefore, we assume a single value of a note in terms of output, a value denoted y.

2.3 The second subdate

As noted above, the purpose of the second subdate is twofold: to allow taxes on notes outstanding to be collected and redistributed and to permit the operation of a centralized competitive market from which an interest rate can be identified. The base for the tax is the value in terms of goods of the notes outstanding. Per specialization type, the above steady-state conclusions imply that the amount of notes outstanding is one half of $1 - A + (1 - \lambda)A$ or $\frac{1-\lambda A}{2}$. Because each note trades for y, the total value of notes, the tax base, is $y\frac{1-\lambda A}{2}$. We let τ denote the tax rate which, as in the actual NBS, has the units of an interest rate. Thus, each participant pays $\tau y\frac{1-\lambda A}{2\lambda A}$ at each subdate. We assume that the tax is distributed equally to all known people. (We could as well have assumed that tax revenue disappears.) Hence, each known person gets a transfer equal to $\tau y\frac{1-\lambda A}{2A}$. Given these taxes and transfers, there is competitive trade at each subdate.

The symmetry across goods and the stationarity imply that all relative prices are unity and that the market-clearing interest rate on loans from one subdate to the next, a rate at which there is no borrowing and lending, is $\beta^{-1} - 1$. Let $y^* = argmax_y z(y)$, where $z(y) \equiv u(y) - y$. Provided the transfer does not exceed y^*, the competitive allocation has each known person consuming y^*, and has each person producing y^* plus the person's *net* tax. That is, the subdate return for each participant is $z(y^*) - \tau y\frac{1-\lambda A}{2\lambda A} +$

$\tau y \frac{1-\lambda A}{2A}$ and for each non-participant is $z(y^*) + \tau y \frac{1-\lambda A}{2A}$. Only the difference between the two ends up playing a role.

2.4 The model and the NBS economy

Before proceeding to the rest of the analysis, we want to discuss how we associate the main objects in the model with features of the NBS economy. Fundamental to this association is how we regard the note issuers of the model, the participants.

The notes in the model look like payable-to–the-bearer, trade-credit instruments. In the model, notes get into circulation only when an unknown person without a note produces for a participant and receives a note. They get redeemed when an unknown person receives goods from a participant in exchange for a note. Thus, notes do not get into circulation through the granting of loans by participants and are not redeemed by way of the repayment of loans.

Given those aspects of notes in the model, in order to identify them with NBS notes, we interpret the set of participants in the model as an entity which represents a consolidation of note-issuing activities of NBS banks and their customers who borrow in the form of notes. If all NBS banks had balance sheets consisting of nothing more than notes as liabilities and loans in the form of notes as assets, then we would simply be consolidating across those banks and their borrowers. Several comments are in order about this interpretation.

First, in the actual NBS economy, note-issuing banks were sometimes asked to redeem notes for lawful money. Although lawful money (outside money) could be added to the model, it seems to be an unnecessary complication. Instead, the model has redemption for goods. In terms of unconsolidated entities, we can associate redemption in the model with the following two-part transaction involving a banker: when a note is presented to a banker for redemption in goods, the banker calls in a loan which the borrower repays in goods. Second, there is no collateral requirement for note issue in the model. This omission is justified by the fact, mentioned above, that the collateral restriction was not binding in the aggregate: the amount of eligible collateral exceeded the amount used as collateral. In any case, this omission makes note issue in the model more attractive. Third, note issue in the model is like a credit line with a nonbinding limit. This, again, tends to make note issue in the model more attractive than it probably was in the NBS economy. Fourth, this interpretation obviates the need to be concerned about how NBS note-issuing banks shared the gains from note issue with those who borrowed in the form of notes. Loans are complicated contracts and borrowers who accepted loans in the form of notes rather than in the form of lawful money may have gotten more favorable terms on their loans. On our interpretation, these loans net out and do not need to be considered. Finally, and most important, if note issue was profitable for NBS banks, then it should have been profitable for the kind of consolidated entity that the model describes because those who borrowed in the form of notes from NBS banks should not have incurred losses by engaging in such borrowing.

3 Conditions for an internal steady-state

We now set out the conditions for existence of a steady state with $\lambda \in (0, 1)$, one with both participants and non-participants. We begin by describing expected discounted utilities at the start of a date as functions of λ and y. Although the crucial conditions for an internal steady state depend on those for participants and non-participants, we begin with those for the unknown people. They are relevant for participation constraints which limit the set of possible y's.

We let v_{uj} denote expected discounted utility of an unknown person with j notes at the start of a date, immediately prior to meetings at the first subdate, and let $\Delta_u \equiv v_{u1} - v_{u0}$. These values satisfy

$$v_{u0} = \beta v_{u0} + \left[\frac{\lambda A}{S} + \frac{1 - \lambda A}{2S}\right](-y + \beta\Delta_u) \tag{3}$$

$$v_{u1} = \beta v_{u1} + \left[\frac{\lambda A}{S} + \frac{1 - \lambda A}{2S}\right][u(y) - \beta\Delta_u], \tag{4}$$

where $\frac{\lambda A}{S}$ is the frequency of meetings with participant producers and of meetings with participant consumers and where $\frac{1-\lambda A}{2S}$ is the frequency of meetings with other producers without a note and of meetings with other consumers with a note.

For the above kind of allocation, a participant has no state. Thus, we let v_k denote expected utility of a participant. This satisfies

$$v_k = \beta v_k + \frac{1-A}{2S}[u(y) - y] + \beta^{1/2}\left[z(y^*) - \tau y \frac{1-\lambda A}{2\lambda A} + \tau y \frac{1-\lambda A}{2A}\right]. \quad (5)$$

The second term on the right-hand side represents payoffs at the first sub-date: with probability $\frac{1-A}{2S}$ the participant meets an unknown producer without a note and gets to consume, and with the same probability he meets an unknown consumer with a note and redeems it for y.

Finally, we describe expected discounted utilities for non-participants. We let $\tilde{v}_{kj}$ denote the discounted expected utility of a non-participant with j notes and let $\tilde{\Delta}_k = \tilde{v}_{k1} - \tilde{v}_{k0}$. These satisfy

$$\tilde{v}_{k0} = \beta \tilde{v}_{k0} + \frac{1-A}{2S}[-y + \beta\tilde{\Delta}_k] + \beta^{1/2}\left[z(y^*) + \tau y \frac{1-\lambda A}{2A}\right] \quad (6)$$

and

$$\tilde{v}_{k1} = \beta \tilde{v}_{k1} + \frac{1-A}{2S}[u(y) - \beta\tilde{\Delta}_k] + \beta^{1/2}\left[z(y^*) + \tau y \frac{1-\lambda A}{2A}\right]. \quad (7)$$

In each case, the second term on the right-hand side represents the expected payoff coming from meetings with unknown people at the first subdate.

For the purpose of comparing the value of participating to that of not participating and consistent with a steady state analysis, we take the payoff to being a non-participant to be the expected value of $\tilde{v}_{k0}$ and $\tilde{v}_{k1}$. Because the steady state has half of non-participants with a note and half without a note, that expected value, denoted $\tilde{v}_k$, is the simple average of $\tilde{v}_{k0}$ and $\tilde{v}_{k1}$. Therefore, from (6) and (7), we have

$$(1-\beta)\tilde{v}_k = \frac{1-A}{4S}[u(y) - y] + \beta^{1/2}\left[z(y^*) + \tau y \frac{1-\lambda A}{2A}\right]. \quad (8)$$

Then, from (5) and (8), we have

$$(1-\beta)(v_k - \tilde{v}_k) = \frac{1-A}{4S} z(y) - \beta^{1/2}\tau y \frac{1-\lambda A}{2\lambda A}. \quad (9)$$

The first term on the right-hand side represents the gain to participating. That gain is more frequent trade in the first sub-period: the ability of known people to issue notes makes trade more frequent because it makes their current trades less dependent on their recent trades.[4] The cost to participating is the tax. It follows that the tax rate that implies $\tilde{v}_k = v_k$, denoted $\tau^*(y, \lambda)$, satisfies

$$\tau^*(y, \lambda) = \frac{G(y)A}{2S\beta^{1/2}} \frac{(1 - A)}{(\lambda^{-1} - A)} \tag{10}$$

where $G(y) \equiv \frac{z(y)}{y}$. This last condition is one of the requirements for an internal solution for λ.

The fact that $\tau^*(y, \lambda)$ is increasing in λ seems strange. It implies that if we start at an internal steady state and if there is a once-for-all increase in the tax rate, then the new steady state has more participation. This perverse association arises for the following reason. The higher is λ, the lower the tax levied on each participant. (The total outstanding stock is decreasing in λ and the per-participant share of a given stock is decreasing in λ.) That seems reasonable. It is less reasonable that the benefit of being a participant does not depend on the outstanding stock per participant, and, therefore, does not depend on λ. But that is what the model implies. According to the model, a participant makes the same kind of trades as a non-participant in the first subperiod, but makes on average exactly twice as many per period. Thus, the implied gain does not depend on the outstanding stock of notes per participant. We suspect that different specifications of meetings and of trades would make the gain depend on the outstanding stock per participant in a way that would reverse the positive association between τ^* and λ.[5]

Finally, we set out the constraints coming from the requirement that no one in the model can commit to future actions. In keeping with the NBS rules, we assume that trades by unknown people and by non-participants are voluntary, but that a participant who fails to redeem a note is severely punished. These are expressed by imposing that unknown people and non-participants always get non negative gains from trade, but that participants are punished by permanent autarky for failing to redeem a note. These

[4] This is the basis for the result in [1] that inside money achieves strictly more outcomes than outside money.

[5] Another weakness of our model is that we have not described the underlying system whose steady state is the one we are describing. A law of motion for note issue is implied by our assumptions. However, dynamics for entry into and out of participation is not implied.

participation restrictions are

$$y \leq \beta \Delta_u \leq u(y), \tag{11}$$

$$y \leq \beta \tilde{\Delta}_k \leq u(y), \tag{12}$$

and

$$y \leq \beta^{1/2} \left[z(y^*) - \tau y \frac{1 - \lambda A}{2\lambda A} + \tau y \frac{1 - \lambda A}{2A} \right] + \beta v_k, \tag{13}$$

where the first pertains to unknown people, the second to non-participants, and the third to participants. Because the v's are uniquely determined by y and λ, these are restrictions on y and λ. In general, they are more stringent the smaller is β (the higher is the market rate of interest).

We now have all the requirements for existence of an internal steady state. Any (y, λ) with $\lambda \in (0, 1)$ that implies v's that satisfy (3)–(13) is such a steady state. In what follows, we analyze some examples while requiring that $y = y^*$, the best level of output from an ex ante perspective.

4 Examples

Although the quantitative implications of the model should not be taken seriously, we pick some parameters and compare the implied τ^* to the interest rate. First, we set $u(x) = x^{1/2}$, only because this is a simple function that satisfies the assumptions and has been used before in related models. It implies $y^* = \frac{1}{4}$ and $G(y^*) = 1$. Next, we set $S = 3$—the smallest S consistent with no double coincidences. (A large $G(y)$ and a small S tends to make it more difficult to have the tax rate be lower than the interest rate (see (10)). Then, we compute $\tau^*(y^*, \lambda)$ as a function of λ for $\lambda \in (0, 1]$ for each of 4 cases: $(A, \beta) \in \{0.1, 0.2\} \times \{0.96, 0.99\}$. (Small values of A reflect the idea that most people in the NBS economy would not have qualified for credit lines from banks.) We also verify that the participation constraints are satisfied and that the implied transfer to non-participants falls short of y^*.

The results for τ^* in each case are very closely approximated by a linear function of λ that goes through the origin. Therefore, we summarize the results by presenting the slope of that approximation in Table 1.

Table 1. The slope b in $\tau^* \approx b\lambda$

	$\beta = 0.96$	$\beta = 0.99$
$A = 0.1$	0.017	0.017
$A = 0.2$	0.034	0.034

Because the interest rate is approximately equal to $1 - \beta$, in three of the four cases in the table, there is an interval around $\lambda = 0.5$ consistent with an interest rate higher than the tax rate. The exception is $A = 0.2$ and $\beta = 0.99$.[6]

5 Concluding Remarks

This paper rests on one crucial detail about how the the NBS operated: a newly issued note that was used in the organized securities markets of a large financial center would end up being redeemed in a way that imposed net costs on the issuing bank. That implies that issuing banks had to be concerned about the uses to which their newly issued notes would be put. It seems obvious, then, that notes would get into circulation mainly through loans made by a bank to its customers for the purpose of working capital. That, in turn, has several implications. Because the quantity of notes issued by way of loans is not easily controlled, the decision to be an issuer becomes a once-for-all choice about whether to participate. Also, we can think about the profitability of participating in the note-issue scheme by thinking about the benefits to the consolidated unit consisting of the NBS note-issuing bank and its customers who borrow in the form of notes. That this is valid seems obvious. The borrowers would not be incurring losses by way of their credit arrangements. Therefore, if the NBS bank is gaining, then the consolidated entity must gain. But what are those gains? The gains must have been fairly subtle. Thus, if one believes that the note-issue scheme did nothing but lead to the replacement of specie by notes, then there could not have been permanent gains from participating. The presence of other forms of inside money is also relevant. The quantity of deposits grew rapidly during the period of the NBS. For notes to have been profitable, they must have had a role even if deposits were available.

[6] We mention intervals around $\lambda = 0.5$ because we are skeptical about the model for extreme values of λ.

We have adapted an existing model, one which is best viewed as a model of circulating trade credit, so that it displays analogues of two main features of the NBS system. One is that newly issued notes are not used in the organized securities markets of financial centers. The other is that a tax is levied on outstanding notes. The virtue of the model is that it illustrates the possible discrepancy between the kind of profit calculation done by previous investigators and true measures of profitability. According to the model, the potential benefit of participation in the note-issue scheme is like the benefit of a credit line; the ability to make current transactions is made less dependent on recent realizations of receipts and expenditures. According to the model, this benefit could be balanced by a tax on notes, even though the market interest rate substantially exceeds the tax rate.

The model, which is very special, seems to overstate the benefits of participation in the note-issue scheme. In the model, notes are the only vehicle for getting a credit line. Therefore, even though the non-participants have known histories, if they do not participate, then they have to trade exactly as do people with unknown histories. In the actual NBS economy, such people could have had a credit line against which they could write checks. If the model allowed for that possibility, then it would be harder to find a benefit of the note-issue scheme. Also, the value of the credit line seems, if anything, to be enhanced by the assumptions that notes are indivisible and that people can hold at most one unit. Finally, the fact that issuing banks had to be ready to redeem notes—and, therefore, may at times have had to make costly portfolio adjustments—is handled in a very simple way in the model: current production finances note redemption.

References

[1] Cavalcanti, R., and N. Wallace, Inside and outside money as alternative media of exchange. *Journal of Money, Credit, and Banking*, 31 (August 1999, part 2), 443–57.

[2] Cavalcanti, R., and N. Wallace, A model of private banknote issue. *Review of Economic Dynamics*, 2 (1999), 104–36.

[3] Champ, B., N. Wallace, and W. Weber, Interest rates under the U.S. National Banking System. *Journal of Monetary Economics*, 34(3), (December 1994), 343–58.

[4] Champ, B., S. Freeman, and W. Weber, Redemption costs and interest rates under the U.S. National Banking System. *Journal of Monetary Economics*, 31(3), Part 2 (August 1999), 568–89.

[5] Friedman, M., and A. Schwartz, *A Monetary History of the United States*. 1963, Princeton: Princeton University Press.

[6] Shi, S., Money and prices: a model of search and bargaining. *Journal of Economic Theory* 67 (December 1995), 467–98.

[7] Trejos A., and R. Wright., Search, bargaining, money and prices. *Journal of Political Economy* 103 (January 1995), 118–41.