Costly Banknote Issuance and Interest Rates under the U.S. National Banking System*

Antoine Martin Cyril Monnet Warren E. Weber

March 22, 2000

Preliminary and Incomplete
Please do not quote or cite

1. Introduction

In 1863, the United States established the so-called National Banking System. Under this system private commercial banks (national banks) were permitted to issue liabilities (bank notes) that had to be fully collateralized by government bonds deposited with the U.S. Treasury and that had to be redeemed on demand at par in "lawful money" – specie or legal tender notes ("greenbacks"). A bank was subject to a tax on its notes, but the tax was applied only to the quantity of notes actually in circulation.[1] The notes of national banks carried an explicit government guarantee against noteholder loss should the issuing bank fail. It is generally accepted that national bank notes were perfect substitutes for specie and legal tender notes in circulation.

From 1863 until the establishment of the Federal Reserve System in 1913, yields on government bonds were consistently above the tax on bank note circula-

*Federal Reserve Bank of Minneapolis and University of Minnesota. Address correspondence to Weber at wew@res.mpls.frb.fed.us. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

[1]The tax was assessed on January 10 and July 10 on the average amount of notes that a bank had in circulation in the six months previous to the assessment date.

tion.[2] The arbitrage opportunities that would seem to exist because of this spread have been noted by many authors (see Bell 1912; Cagan.1963, 1965; Cagan and Schwartz 1991; Champ 1990; Champ, Freeman, and Weber (CFW) 1999; Champ, Smith, and Williamson (CSW) 1996; Champ, Wallace, and Weber (CWW) 1992, 1994; Friedman and Schwartz 1963; Goodhart, 1965; James 1976; Kuehlwein 1992). The persistence and magnitude of these apparent arbitrage opportunities have led Cagan and Schwartz to dub this phenomena the "National Bank note puzzle".[3]

Several solutions to the puzzle have been proposed. CFW argue that the expected cost that banks have to bear when their notes are presented for redemption should be added to the tax in terms of evaluating the magnitude of the arbitrage opportunity. However, they find that the magnitude of these costs is not sufficiently large to make the arbitrage opportunity go away.

Another explanation of the puzzle is given by CSW. They build an OLG spatial separation model of bank note issuance. When bank note issuance is perfectly elastic, the interest rate on loans, which can be considered to be the yield on government bonds, is equal to the cost of note issue. (The same result is obtained in CWW.) However, when note issuance is perfectly inelastic, the rate on loans is above the cost of note issuance and changes in the demand for bank notes give rise to interest rate fluctuations.

CSW argue that the perfectly inelastic note issuance case more nearly characterizes the National Banking System, at least in the short run. In the short run, CSW argue that there were institutional constraints, what they call the "mechanics of note issue," that prevented national banks from changing the amounts of notes that they had in circulation. We agree that in the aggregate, the supply of national bank notes appears to have been inelastic. However, we disagree that the cause(s) was exogenous for two reasons. First, banks held government bonds beyond that required to collateralize their circulation. Since the tax was only on circulation, it would have been costless for banks to have idle notes in their vaults to meet increased demands. Second, much of the increased demand for notes

[2] Actually, yields on government bonds exceeded the tax on note circulation after the Federal Reserve System was established. However, we consider the National Banking System era as ending in 1913, because the establishment of the Fed signalled that the ability of national banks to issue notes would soon come to an end.

[3] Sometimes the puzzle is stated as, If bank notes issuance was profitable, why were fewer bank notes issued than was permitted by the constraint that notes had to be collateralized by deposit of government bonds? For reasons that will become obvious later, we have chosen the alternative of framing the puzzle in terms of the persistence of an arbitrage opporunity.

was predictable due to its seasonality so the mechanics of note issue would not have prevented national banks from having additional notes on hand in a timely manner.

This paper builds upon the model CSW, but instead of exogenously inelastic note issue, it introduces a form of endogenous inelasticity by assuming that cheap notes can be issued up to some point, but after that point, note issuance carries a higher cost. What we are trying to capture is the idea that banks expect that there is some quantity of notes, the cheap notes, that will stay in circulation for long periods of time. However, they also expect that notes issued beyond that quantity, particularly those issued in terms of high demand, will be returned to the bank (presented for redemption) very quickly. These notes are more expensive because the bank has to pay a cost when they are redeemed, but only earns a small amount of interest on the assets purchased with then.

We will judge the success of our model by the ability of a reasonably parameterized version to explain certain secular and cyclical facts during this period. In particular, we want to explain

- the level of and movements in long term interest rates
- the magnitudes of the discounts on certified checks, the quantity of bank notes issued, and the level of call loan rates during financial panics, and
- changes in the quantity of bank notes issued

The paper proceeds as follows. In section 2, we present the National Banknote Puzzle in more detail and present some facts about the elasticity of national bank note circulation. In section 3, we present the model. In section 4, we discuss some data issues and describe our parameterization (calibration) of the model. In section 5, we discuss how well the calibrated model explains the facts. In section 6, we conclude.

2. National bank notes

2.1. The puzzle in more detail

The National Bank Note Puzzle is that during this banking regime there existed an apparent arbitrage opportunity between the cost of issuing bank notes and the yield on U.S. government bonds that was left unexploited. National banks could

issue notes, which could be used to purchase government bonds or other assets. Profit-maximization suggests that banks should have bid the price of assets up (the yield on assets down) to the point where the yield was equal to the cost of issuing notes. Since these notes were considered perfect substitutes for "lawful money," specie and legal tender notes (greenbacks), they carried a nominal interest rate of zero. The tax rate on note issues was one percent, one-half percent after 1900. The other costs associated with note issue also should have been relatively constant. Thus, the yield on government bonds, or other safe assets should have been constant up to 1900, then fallen by one-half percent and remained constant at that new level thereafter.

The data on government yields shows that government bonds did not follow this pattern. During the period we are considering, there were only a few issues of government bonds in existence at any point in time. These issues were all long-term and were identified by their coupon rate and the date at which they first became callable. Thus, for example, the 4s of 1907 were four percent coupon bonds that were first callable in 1907. Monthly bond yields for the three government bonds issues that made up the vast majority of the collateral for national bank notes are shown in Figure 1 for the period 1882 to 1913.

Instead of the pattern implied by the elimination of arbitrage opportunities given above, Figure 1 shows the following general characterization of bond yields. Yields declined from 1882 through the middle of 1889. From the middle of 1889 through the middle of 1893 bond yields increased, reaching or slightly exceeding their levels in 1882. Yields were roughly constant between the middle of 1893 and the end of 1895, after which time, they declined sharply until 1898. They then rose sharply in that year, but subsequently declined until the end of 1901. From 1902 through 1913, bond yields generally showed a slight upward trend. These movements of interest rates are one fact that we would like to explain with our model.[4]

[4]The yields shown in Figure 1 are calculations based upon end-of -month bond price data collected from the *Commercial and Financial Chronicle* by Bruce Champ and James Thomson and kindly provided to us. All calculations were done using the "yield" worksheet function in Excel, assuming that coupons are on the first day of the month that they were scheduled to be paid and that bonds were expected to mature on the first date at which they could be called. Since our yield calculation are sensitive to the assumption about when the bonds would be called and since some bonds – the 4-1/2s of 18xx not shown in our figure and not held extensively by national banks – were called early in the 1880s due to the government budget surpluses at the time, our yield calculations should be taken as approximate. However, Macauley's yields on New England municipal bonds and on railroad bonds exhibit movements that are roughly

2.2. Comments on inelasticity of note issue

The national bank note puzzle has also been framed in terms of too few notes being issued, and CSW focus heavily on the supposed inelasticity of bank note issuance. For these reasons, we comment briefly on the empirical evidence regarding how bank note circulation changed over time.

We first present evidence on high frequency movements in the quantity of national bank notes in circulation. To get an idea of high frequency changes in national bank circulation, for the period 1881 to 1910 we plot the percentage changes in total circulation between call report dates for each of the five call reports that national banks were required to file each year. This is done in Figure 2.[5] This figure makes two points. First, there was not much high frequency fluctuation in the quantity of national bank notes in circulation. The average absolute percentage change was only 2.5 percent. Further, there are only 19 times, out of the 150 observations in Figure 3, in which national bank circulation changed by more than five percent in either direction between call reports and only 4 times in which it increased more than ten percent in either direction. Figure 2 is some evidence why national bank circulation was considered to be inelastic by many contemporary observers.

However, we get a different view of the inelasticity of note issue if we look at low frequency movements in bank note we plotted the circulation of national bank notes at each call report date. Although these call reports are not at evenly spaced time intervals, we have treated them as such for purposes of plotting the data. The result is shown in Figure 3.[6] In terms of low frequency movements in national bank circulation, the figure shows that bank note circulation declined by slightly over 60 percent between the end of 1881 and the end of 1890. After that,

in line with our yield calculations. This gives us some confidence that the yields shown in the figure are indicative of general movements in government bond yields during the period.

[5] Weber has compiled the disaggregated call reports by state and reserve city for each call report date. These data can be found at http:\\woodrow.mpls.frb.fed.us\research\economists\wewproj.html. All national bank data used in this paper are aggregates of these data, rather than the aggregates reported by the Comptroller of the Currency. The reason is in many cases the totals given in the Comptroller's aggregates do not agree with the totals from this disaggregated reports. More details are given with the data at the website.

[6] We also examined net circulation, measured as total circulation less holdings of bills of other banks, of national bank notes. The movements in the two series are virtually identical, so we did not plot net circulation in the figure or discuss it in the text.

circulation increased slowly until the end of 1899 after which it increased at a far more rapid pace.

3. The model

We consider an overlapping generations economy with two-period lived households and two locations. Time is discrete, infinite, and indexed by t. There is a single, non-storable consumption good, denoted $x(t)$, in each location in each period. Consumption goods are not transferable across locations. Locations are physically distinct, so that it is not possible for agents or banks to communicate across locations.

A continuum of identical households of measure one is born in each location in each period. Half of the measure of households in each location, "lenders", receive an endowment of the consumption good $z_t(t) > 0$ only in the first period of life. The other households, "borrowers", receive an endowment of the consumption good $y_t^*(t+1) > 0$ only in the second period of life. We assume that although the money in this economy is commodity money, it does not yield any direct utility to the holders.

There are also a large number of banks in each of the two locations. These banks are infinitely lived (although one could think of banks as also two-period lived with a new cohort of banks coming into existence in each period). These banks can take deposits, issue bank notes, and make loans.

Lenders do not necessarily remain in the same location for their entire lives. Some measure of lenders of each cohort spend the first period of their lives at one location and the second period of their lives at the other location. Although whether or not a particular young lender has to move and the measure of young lenders that have to move are public information, this information is not revealed until the end of the agent's first period of life. The probability that a young lender of cohort t will move is $\pi(t, j)$, where $j = g, b$ is the state of the economy at time t. There are two states, good (g) and bad (b), such that $\pi(t, g) < \pi(t, b)$. That is, fewer young lenders have to move in the good state than have to move in the bad state. Thus, the bad state can be thought of as a bank run. The probability that the good state occurs at time t is $\alpha(t, g)$.

It is the combination of impossibility of communication across locations with the fact that some households have to move that gives rise to the need for banknotes. The impossibility of communication means that a household (or bank) in one location cannot verify that the check written on a bank in another location is

good (not counterfeited) or not. Banknotes are not counterfeitable, by assumption. Thus, a young household that has to move will want bank notes, since deposits will be useless in its new location. The assumption that bank notes can be used for transactions at far locations, whereas deposits are restricted to near locations, is an attempt to capture the transactions facts of the national banking period. Specifically, Sprague says, "..." [find quote]

The specific timing of transactions and when households move is the following:

- *Lenders.* At the beginning of the first period of life, they receive their endowments. They can then deposit part or all of their endowment in one of the banks in the economy. After making the deposit, each young lender consumes and then finds out whether or not she will stay in the same location. Lenders who stay in the same location do nothing. Movers go to their bank and redeem their deposits for specie (lawful money) or banknotes. In the second period of life, old nonmovers go to their bank, withdraw their deposits, sell these deposits to old lenders for goods and consume. Old movers use their banknotes or specie to buy goods from old lenders at the new location and then consume.

- *Borrowers.* In the first period of life, borrowers go and take out a loan from the bank at a nominal gross interest rate of $R(t)$ after lenders have deposited. After getting the loan, the borrowers consume the proceeds. At the beginning of the second period of life, borrowers receive their endowments. They sell the part of the endowment to old lenders either for deposits, specie, or banknotes. They then use the deposits, specie, or banknotes to pay back their loans from banks. After paying off their loans, old borrowers consume what is left of their endowments.

Banks in our model economy face a cost per note issued, just as did national banks. A national bank had to pay a tax on the amount of its notes in circulation. In addition, it was charged for costs of sorting and shipping any of its notes that were presented for redemption through the redemption facility at the U.S. Treasury. Further, national banks incurred fixed costs, the wages of tellers and the cost of having a vault, associated with note issue.

With these considerations in mind, we model the costs of note issuance as follows. Banks can issue two kinds of notes. There are "cheap" notes, denoted $B(t, j)$, which it can issue in state j in time t. These notes carry a redemption cost of $c_B \geq 0$ per note, payable in period $t + 1$. There are also "expensive"

notes, denoted $N(t,j)$, which it can issue in state j in time t. These notes carry a redemption cost of $c_N > c_B$ per note, also payable in period $t+1$. The reason that banks would ever choose to issue expensive notes is that there is a maximum amount of cheap notes, $\beta(t)$, that a bank can issue in any period. What we are trying to capture with the maximum on cheap note issuance is the idea that banks expect that there is some quantity of notes, $\beta(t)$, that will stay in circulation for a relatively long period of time. However, they also expects that notes issued beyond that quantity, particularly those issued in terms of high demand, will be returned to the bank (presented for redemption) very quickly. It is the presence of the upper bound on cheap note issue that drives the results of the model.

3.1. Household behavior

All households have preferences

$$u[x_t(t), x_t(t+1,j)] = \ln x_t(t) + \delta \ln x_t(t+1,j),$$

where $x_k(t,j) \in \Re_+$ denotes the consumption of a household of cohort k in period t in state j and δ is the rate of time preference. Households are assumed to maximize expected lifetime utility subject to their budget constraints. For borrowers the problem is simple:

$$\max_{x_t(t), x_t(t+1)} [\ln x_t(t) + \delta \ln x_t(t+1)]$$

s.t.

$$x_t(t) + \rho(t) x_t(t+1) = \rho(t) y_t^*(t+1)/R(t)$$

where $p(t)$ is the inverse price level in period t and $\rho(t) = p(t)/p(t+1)$ is the rate of inflation between periods t and $t+1$. Solving, borrowers' demand for loans in period t is

$$\ell(t) = x_t(t) = \rho(t) y(t+1)/R(t)$$

where $y(t+1) = (1+\delta)^{-1} y^*(t+1)$.

The lenders' problem is more complicated. As we discuss below, banks can offer state-contingent payment schedules for demand deposits. Specifically, let $R^m(t,j)$ be the gross nominal rate of return paid to depositors (lenders) who withdraw early (movers) in time period t if state j occurs, $R^n(t,j)$ be the gross nominal rate of return paid to depositors who do not withdraw early (nonmovers) in time period t if state j occurs, and $d(t)$ be the deposits of lenders in period

t. Then the consumption of movers and nonmovers in the second period of their lives will be $x(t+1,j) = R^m(t,j)d(t)/\rho(t)$ and $x(t+1,j) = R^n(t,j)d(t)/\rho(t)$, respectively. Thus, the lender's problem is

$$\max_{x_t(t),d(t)} \left\{ \ln x_t(t) + \delta \sum_{j=g,b} \alpha(t,j)\{\pi(t,j)\ln[R^m(t,j)d(t)] + [1-\pi(t,j)]\ln[R^n(t,j)d(t)]\}/\rho(t) \right\}$$

s.t.

$$x_t(t) + d(t) = z(t)$$

The solution is

$$d(t) = \delta(1+\delta)^{-1}z(t).$$

3.2. Bank behavior

Banks take deposits, make loans, and pay out specie and banknotes. We assume that banks take as given the quantity of deposits, the interest rate at which they can make loans, the cost of bank note issue, and the constraint on the amount of cheap bank notes they can issue. However, we allow them to decide how many loans to make, how much specie to hold, and how many of each type of banknotes to issue. (We will make assumptions so that banks will never want to hold excess reserves, but will always want to issue notes.) Further, we permit banks to choose schedules of interest payments on deposits that depend upon the state of the economy and whether on not an individual is a mover or a nonmover.

Banks face three constraints on their behavior. First, banks must be able to meet the withdrawal demands of movers either out of their specie holdings of by issuing bank notes

$$\pi(t,j)R^m(t,j)d(t)/p(t) = S(t) + B(t,j) + N(t,j) \qquad \forall t,j. \tag{3.1}$$

The righthand side of (3.1) is a bank's obligations in period t, the payments it promises to movers. The lefthand side is a bank's resources to meet these obligations, its holdings of specie, denoted $S(t)$, and its issuance of both kinds of banks notes. Because the interest rate $R^m(j,t)$ is expressed in terms of units of $t+1$ dollars per unit t dollars, we have deflated the lefthand side of (3.1) by $\rho(t)$.

Second, in period $t+1$, banks must be able to meet the withdrawal demand of nonmovers and pay the costs of note issue out of their earnings from the loans they made in period t

$$[1-\pi(t,j)]R^n(j,t)d(t)/p(t) + c_B B(t,j) + c_N N(t,j) = R(t)\ell(t)/p(t) \quad \forall t,j. \tag{3.2}$$

Third, banks face an overall resource constraint at the beginning of each period

$$\begin{aligned} & d(t+1) + R(t)\ell(t)/p(t) + p(t+1)[S(t) + B(t,j) + N(t,j)] \\ = \; & l(t+1) + p(t+1)S(t+1) \\ & +[1-\pi(t,j)]R^n(j,t)d(t)/p(t) + c_B B(t,j) + c_N N(t,j) \\ & +\pi(t,j)R^m(t,j)d(t)/p(t) \qquad \forall t, j. \end{aligned} \tag{3.3}$$

The lefthand side of (3.3) is the resources a bank has in period $t+1$. These consist of the deposits it receives, the payback of loans made in the previous period, the specie that it held from the previous period, and the notes that it can issue, although technically these are issued at the very end of period t. The righthand side is the expenditures of the bank. It makes loans, buys specie, and pays off the obligations associated with the deposits in took in the previous period. Substituting (3.1) and (3.2) into (3.3) yields

$$d(t) = l(t) + p(t)S(t) \qquad \forall t. \tag{3.4}$$

The equilibrium concept that we use is Nash equilibrium. Given that there are a large number of banks, this choice has two immediate implications. The first is that in equilibrium banks will earn zero profits, because if banks are earning profits, then there is an incentive for an individual bank to deviate and offer higher interest rates on deposits and thus attract all depositors. The second is that the deposit contracts offered by banks in equilibrium will maximize the utility of households, because if they do not, once again there are incentives for an individual bank to deviate.

Thus the banks' optimization problem is

$$\max \sum_{t=0}^{\infty} \delta^t \left[\sum_{j=g,b} \alpha(t,j) \left\{ \pi(t,j) \ln \left[\frac{R^m(t,j)d(t)}{\rho(t)} \right] + [1-\pi(t,j)] \ln \left[\frac{R^n(t,j)d(t)}{\rho(t)} \right] \right\} \right]$$

subject to (3.4), (3.1), (3.2),

$$p(t)S(t) \leq d(t) \tag{3.5}$$

and

$$B(t,j) \leq \beta(t). \tag{3.6}$$

The maximization is over $\{R^m(t,j), R^n(t,j), \ell(t), S(t), B(t,j), N(t,j)\}$.

Substitute for $\ell(t)$ using (3.4) and let $\lambda(t,j), \theta(t,j), \mu(t), \nu(t,j)$ be the Lagrange multipliers for (3.1), (3.2), (3.5), and (3.6), respectively. The first order conditions for the banks' optimization problem are

$$\delta^t \frac{\alpha(t,j)}{R^m(j,t)} - \lambda(t,j)\frac{d(t)}{p(t)} = 0, \qquad \forall t,\ j \tag{3.7}$$

$$\delta^t \frac{\alpha(t,j)}{R^n(t,j)} - \theta(t,j)\frac{d(t)}{p(t)} = 0, \qquad \forall t,\ j \tag{3.8}$$

$$\lambda(t,j) - c_B\theta(t,j) - \nu(t,j) \leq 0, \qquad \forall t,\ j \qquad \text{with equality if } B(t,j) > 0 \tag{3.9}$$

$$\lambda(t,j) - c_N\theta(t,j) \leq 0, \qquad \forall t,\ j \qquad \text{with equality if } N(t,j) > 0 \tag{3.10}$$

$$\lambda(t,g) + \lambda(t,b) - R(t)[\theta(t,g) + \theta(t,b)] - \rho(t)\mu(t) \leq 0,\ \forall t \ \text{ with equality if } S(t) > 0 \tag{3.11}$$

$$\mu(t) \geq 0, \qquad \text{with equality if } p(t)S(t) < d(t) \tag{3.12}$$

$$\nu(j) \geq 0, \qquad \forall t,\ j \qquad \text{with equality if } B(j,t) < \beta(t) \tag{3.13}$$

and the constraints.

3.3. Market clearing

The market that has to clear in this model is that for loans. The quantity of loans sought by borrowers has to be equal to quantity of loans that banks are willing to make. That is

$$\rho(t)y(t)/R(t) = d(t) - p(t)S(t). \tag{3.14}$$

3.4. Equilibrium

For defining equilibrium in this economy, we face a choice of whether to take prices or the quantity of specie as exogenous. We choose prices. This means that banks can hold the quantity of specie that the want as reserves. Thus, we are implicitly assuming that there is a market where banks can trade specie for goods or goods for specie at $p(t)$. If we think of our model as strictly of the banking sector, then this market is the rest of the economy. If we think of our model as the domestic economy, then this other market is the rest of the world, and banks' adjustment of specie holdings corresponds to international specie flows. Under this latter interpretation, we admit that there is some inconsistency, because we are, in effect, assuming that the economy is a small country with respect to goods,

but a large or at least isolated economy in terms of capital. However, we chose as we did, because we wanted to have banks' decisions regarding reserves matter.
Assumption P:

$$1 - \pi(t, g) < \frac{\rho(t)y(t+1)}{c_B d(t)} < 1 \qquad \forall t$$

This assumption guarantees that banks will always want to issue notes in both states and will want to hold positive amounts of reserves.
Condition B:

$$\frac{y(t+1)}{c_B p(t+1)} - [1 - \pi(t, b)]\frac{d(t)}{p(t)} < \beta(t).$$

If this condition is satisfied, bank notes issuance in both states will be less than $\beta(t)$.
Condition N:

$$\frac{y(t+1)}{c_N p(t+1)} - [1 - \pi(t, g)]\frac{d(t)}{p(t)} > \beta(t)\left[1 - \frac{c_N - c_B}{c_N}\pi(t, g)\right].$$

If this condition is satisfied, banks will issue expensive notes in both states.

Definition 3.1. *Positive reserves and note issuance equilibrium. Under assumption P, given* $\{z(t), y(t), p(t), \beta(t), \alpha(t, j), \pi(t, j)\}$, δ, c_B, *and* c_N, *a set* $\{R^m(t, j), R^n(t, j), R(t), S(t) > 0, B(t, j) > 0, N(t, j) \geq 0\}$ *is a Nash equilibrium if it satisfies household and bank maximization and loan market clearing.*

The following lemma will be useful for the discussion which follows:

Lemma 3.2. *Under Assumption P, any equilibrium will have the properties:*

1. $[\beta(t) - B(t, j)]N(t, j) = 0$,
2. $c_B \leq R(t) \leq c_N$,
3. $c_B R^m(t, j) \leq R^n(j, t)$ *with equality if* $B(t, j) < \beta(t)$,
4. $c_N R^m(t, j) \geq R^n(t, j)$ *with equality if* $N(t, j) > 0$,
5. $R^m(t, b) \leq R^m(t, g) \leq c_N$, and

6. $c_B \leq R^n(t,g) \leq R^n(t,b)$.

   Proof: Properties (1) through (4) follow directly from the banks' first order conditions. Proofs of properties (5) and (6) are given in the Appendix.■

Property (1) states that a bank will not issue expensive notes in a state unless it has issued all the cheap notes it can in that state. Property (2) states that the nominal interest rate on loans has to at least cover the cost of issuing cheap notes, but has to be less than or equal to the cost of issuing expensive notes. Property (3) states that if the constraint on issuing cheap banknotes is not binding in a state, then the difference in the interest rates paid to movers and nonmovers in that state will just cover the cost of redeeming the notes. If the constraint is binding, however, the difference can exceed this cost. This property makes it incentive compatible for nonmovers to not want to withdraw deposits from the bank early. Property (4) states that if banks issue expensive notes in a state, then the difference between the interest rates paid to movers and nonmovers must cover the cost of redeeming the expensive notes. Property (5) states that the interest rate paid to movers in the good state will always be at least as great as that paid to movers in the bad state, but that this interest rate will never exceed the cost of issuing the expensive notes. Property (6) states that the interest rate paid to nonmovers in the bad state will always be at least as great at the interest rate paid to movers in the bad state, and this interest rate will always be greater than the cost of issuing cheap notes.

The following lemma confirms the intuition that if the restriction on issuing cheap bank notes does not bind, then arbitrage will lead to the interest rate on loans being equal to the cost of issuing bank notes.

Lemma 3.3. *Under Assumption P, if Condition B is satisfied, then there exist values of* $\{z(t), y(t), p(t), \beta(t), \alpha(t,j), \pi(t,j)\}$, δ, c_B, *and* c_N, *such that an equilibrium exists in which banks do not issue expensive notes. This equilibrium has the properties:*

$$R^m(t,j) = 1$$
$$R^n(t,j) = R(t) = c_B$$
$$S(t) = \frac{d(t)}{p(t)} - \frac{y(t+1)}{c_B p(t+1)}$$
$$B(t,j) = \frac{y(t+1)}{c_B p(t+1)} - [1 - \pi(t,j)]\frac{d(t)}{p(t)}$$
$$N(t,j) = 0$$

Proof: Follows directly from the definition of equilibrium.■

The next lemma shows that when banks issue expensive notes in both states, the rate of interest on loans will be equal to the cost of issuing the expensive notes. The interest rate paid to nonmovers is the same in both states and exceeds the rate of interest on loans. The interest rate paid to movers is also the same in both states and exceeds unity.

Lemma 3.4. *Under Assumption P, if Condition N is satisfied, then there exist values of* $\{z(t), y(t), p(t), \beta(t), \alpha(t,j), \pi(t,j)\}$, δ, c_B, *and* c_N, *such that an equilibrium exists in which banks issue expensive notes in both states. This equilibrium has the properties:*

$$R(t) = c_N$$

$$R^m(t,j) = c_B R^n(t,j)$$

$$R^n(t,j) = c_B + \beta(t)(c_N - c_B)\frac{p(t)}{d(t)}$$

$$S(t) = \frac{d(t)}{p(t)} - \frac{y(t+1)}{c_N p(t+1)}$$

$$B(t,j) = \beta(t)$$

$$N(t,j) = \frac{y(t+1)}{c_N p(t+1)} - [1 - \pi(t,j)]\frac{d(t)}{p(t)} - \beta(t)\left[1 - \frac{c_N - c_B}{c_N}\pi(t,j)\right]$$

Proof: Follows directly from the definition of equilibrium.■

In the cases covered by Lemmas 3.3 and 3.4, interest rates are unchanged for any changes in the parameters or exogenous variables. Thus, if we hope to match the facts for this period, we must consider cases where neither Condition B nor Condition N are satisfied. That is, we have to find cases in which the constraint on issuing expensive notes is binding in state b, but where it is not profitable for banks to issue expensive notes in state g.

In examining the cases not covered by Lemmas 3.3 and 3.4, we noticed that interest rates could be written as functions of $\rho(t)y(t+1)/d(t)$ and $\beta(t)p(t)/d(t)$ as well as the other parameters of the model. Figure 4, we present a representative illustration of interest rate behavior in this model as a function of $\rho(t)y(t+1)/d(t)$, holding the values of the parameters and other exogenous variables fixed. Initially we are in the case covered by lemma 3.3. Then, as $\rho(t)y(t+1)/d(t)$ increases, banks hit the constraint on issuing cheap notes if the bad state should occur. At this point, $R(t)$, $R^m(t,g)$, $R^n(t,g)$, and $R^n(t,b)$ begin to increase; $R^m(t,b)$,

however, begins to decrease, so that movers actually get back less than a dollar for each dollar of deposits that they have to withdraw early should the bad state occur. As $\rho(t)y(t+1)/d(t)$ continues to increase, interest rates reach the point where it pays banks to issue expensive notes in the bad state. At this point, the slopes of the interest rate curves change, and for $R(t)$, $R^m(t,g)$, $R^n(t,g)$, and $R^n(t,b)$ the slopes become flatter. Moving right from this point, banks continue to issue more cheap notes in the good state and expensive notes in the bad state until eventually they hit the constraint on issuing cheap notes in the good state. After this point, all interest rates begin increasing until eventually interest rates are high enough so that it pays banks to also issue expensive notes in both states and we are in the case covered by Lemma 3.4. Note that in this case the interest rate paid to nonmovers is higher than the interest rate on loans.

4. Calibration

We now want to see how well a version of the model calibrated to call report data on national banks during the period 1882 to 1913 predicts interest rate levels and changes and bank note circulation during the period 1882-1912. The strategy is the following. First, we match the national bank to the theoretical constructs in the model. Then, we use the national bank data to calibrate the probability and cost parameters in the model. In the next section, we determine how well this calibrated model performs.

4.1. The data

The model constructs that we want to match to the national bank data are note issue, specie holdings, deposits, and the endowments of borrowers. To this we have to decide at what point in the timing of the model the bank balance sheets are being observed. The difficultly is if we assume that banks' balance sheets are observed at the end of a period, then banks should not have specie on their books, since it was all paid out to young movers before the period ended. If we assume that the balance sheets are observed at the beginning of a period, then banks do not have any note labilities, since they don't have to issue notes until households actually move later in the period. In actuality, national banks' balance sheets have both notes and specie at the same time.

Our solution is to assume that bank balance sheets are observed shortly after beginning of a period, so they the current young have had the opportunity to make

their deposits and take out their loans. We also assume that banks have been able to adjust their specie reserves for the period to the desired level. However, we also assume old borrowed who have sold part of their endowment for bank notes have not yet had a chance to return to a bank to pay off their loans. Thus, notes outstanding correspond to an amount of loans made in the previous period.

With this assumption, we map the items on the balances of national banks and the variables in our analysis as:

Notes: $[B(t,j)+N(t,j)]$ "National Bank circulation." "State bank circulation" is not included since national banks could not issue more of these notes.

Specie: $[S(t)]$ "Fractional currency" + "Trade dollars" + "Specie" + "Legal-tender notes" + "5% fund with Treasury" + "Due from U.S. Treasury" + "Due from reserve agents." The "5% fund with Treasury" is included, because these were funds that national banks were required to deposit in specie with the U.S. Treasury as reserves for note issue, but which were also counted as reserves against deposits. The "Due from Treasury" is included because this is specie owed by the Treasury to national banks due to notes of other banks that they have sent in for redemption. The "Due from reserve agents" is included because these were counted as reserves against deposits, and represented funds that banks could call on to meet demands to redeem deposits.

Deposits: $[D(t) = d(t)/p(t)]$ Total liabilities and net worth less notes." Net worth is included as part of deposits, since national banks used their capital to fund loans.

Endowment of borrowers: $[Y(t+1) = y(t+1)/p(t+1)]$ Total assets less specie and notes. If we accept the above timing of the observation of bank balance sheets, then what is on the asset side of the balance sheet after specie and notes (assumed to equal to unpaid loans from the previous cohort) are subtracted is loans made to the current cohort, $L(t) = \ell(t)/p(t)$. The variable we want in the analysis, however, is the endowment of borrowers, $Y(t+1) = R(t)L(t)$, which involves the gross interest rate. Since the gross real interest rate is close to one, we ignore it and approximate $Y(t+1)$ as $L(t)$.[7]

[7] During this period, national bank held deposits for the government – "United States deposits" and "Deposits U.S. disbursing agents" – which they had to back with government bonds

In Figure 5, we plot these variables normalized by deposits, since such a normalization will be useful when we compare the predictions of the model with the data. Note that with this normalization, the fall in notes is still pronounced, but notes as a fraction of deposits (as we measure them) recovers only very slowly after hitting the minimum in 1890. Also note that the ratio of loans to deposits appears to fluctuate more than the ratio of notes to deposits or specie plus notes to deposits, but that is an illusion due to the difference axes used for the different ratios.

4.2. Parameters

The next step is to calibrate the probabilities and costs. Our choices:

The probability of being in the good state: $\alpha(t, g) = 0.9$. We choose the basic time period of the analysis as a year. Sprague identifies four years – 1884, 1890, 1893, and 1907 – in which financial disruptions occurred. XX and XX point out, however, that there were only general restrictions of payments in New York City in 1893 and 1907. We come up with our value for $\alpha(g)$ by averaging these occurrencs over the thirty-one years for which we have data.

The probability of moving in the good state: $\pi(t, g) = 0.2$. Given that the gross interest rates during this period were close to one, we can obtain this probability as the ratio of notes plus specie to deposits from (3.1). From Figure 5, this value is approximately 0.2 over the period.

The probability of moving in the bad state: $\pi(t, b) = 0.23$. We examine the sum of specie and notes normalized by deposits in the call reports near times that Sprague identifies as financial disruptions. These are marked by black diamonds in Figure 5. We see that in two of these panics, these values were about 3 percentage points higher than at previous or subsequent call reports. Thus, we added 0.03 to $\pi(t, g)$ to get the calibrated value $\pi(t, b)$

The cost of issuing cheap notes: $c_B = 1.01 + 0.000625/.9$ (before 1900), $1.005 + .000625$ (after 1900). We take the cost of issuing cheap notes to be the tax on notes (1 percent per annum before 1900, one-half percent thereafter) plus the other costs associated with note issue as computed by the Comptroller

– "Bonds for deposits" – as collateral. We have treated these like other deposits and loans in terms of our mapping from data to theory since they were not equal on bank balance sheets, so eliminating them would mean that our asset and liability measures would not add up.

of the Currency. The Comptroller estimated these costs to be \$62.50 per \$100,000 par value of the bond. Since notes could only be issued up to 90 percent of the minimum of the market or par value of bonds prior to 1900, we divide the additional costs by this number for that period.

The cost of issuing expensive notes: $c_N =?$ We have no way of calibrating this parameter. Instead, we will do here is try some several values to see if we can get good predictions with reasonable numbers.

The restriction of the quantity of cheap notes that can be issued as a percentage of deposits: $\beta(t)p(t)/d(t) = 0.06$. What we wanted was a value such that the restriction on issuing cheap notes would in general not be binding, but would bind occasionally, especially in times of panics. In Figure 5, we have plotted the line for $\beta(t)p(t)/d(t) = 0.06$. We see that this value implies that banks were issuing expensive notes at the beginning and end of our sample period. They also issued them around the time of the crisis of 1893 and in 1897. For most of the rest of the time, however, only cheap notes were being issued. Thus, our choice for this parameter seems reasonable.

5. Results

We now see how well a calibrated version of our model fits with some actual data during the period. In this calibrated version, we will keep the values of the parameters constant over the entire time period. The only exogenous variable that we will allow to change is the rate of the endowments of borrowers and lenders. We do this to limit the number of free parameters that we have. We first examine how the predictions of this calibrated model compare to the level and movements in the yields on long-term government bonds. Then we examine its predictions compared to the discounts on certified checks and the interest rates on call loans during periods of financial panics. Finally, we compare the level and movements in note circulation in the calibrated model with actual note circulation.

5.1. Government bond yields

To see how well a calibrated version of our model performed versus actual government bonds yields, we computed $R(t)$ from our model for each call report date between 1882 and 1910. The actual government bond yields we use for comparison are those on the 4s of 1907 up to 1900 and those on the 2s of 1930 thereafter. We

choose these bonds because they made up the largest component on those held by national banks during this period and they were they ones for which there was the least uncertainty about call dates that might be affecting yield computations. [check this] The call report data are for various days during a month; the government bond yield data are all end of month. For the comparisons we used the end of month observation closest to the call report date.

Comparisons of $R(t)$ from a calibrated version of the model with actual government bond yields over the period 1882 to 1910 for three different values of c_N are shown in the table below.

	Model			Actual
	$c_N = 1.10$	$c_N = 1.15$	$c_N = 1.20$	
Mean	1.56	1.66	1.67	2.23
Variance	0.121	0.191	0.206	0.206
Correlation	0.549	0.534	0.534	

The results look quite good for the model. The means of the calibrated series are too low; they are below the mean of the government bond yields by approximately 60 basis points. This is not unexpected, however, since by choosing the Comptroller's calculations for c_B seem very conservative as they do not include any other costs that a national bank might incur due its note issuing activities. What is more important from our point of view, however, is that the variances of the computed and actual series are very close. Thus, we are able to explain the fluctuations in interest rates during this period.

Further, the correlations between the computed and actual series are all greater than 0.5, which indicates that the computed series tracks the actual series quite well. This is illustrated in Figure 6, where we show the comparison using $R(t)$ computed with a value of $c_N = 1.15$ in the calibration. Specifically, the calibrated model captures the fall in government bond yields from 1882 to 1890, the rise from 1890 to 189x, and then the sharp fall in yields from 189x to 1900. The calibrated model also captures the increase in bond yields after 1900, although the figure also shows that the model does a better job of tracking interest rate movements before 1900 than it does after that.

5.2. Interest rates during financial disruptions

Next we want to examine how well a calibrated version of our model does explaining the interest rates paid to depositors, especially during periods in which there

were financial disruptions. There is no direct evidence on the interest rates paid to depositors under the National Banking System. However, we can make some inferences about what how these rates moved and how high they might have been, at least during periods of financial disruption.

Before the establishment of the Federal Reserve System in 1913, when banks suspended payments on deposits, which they did in 1893 and 1907, they did not close their doors for business. They continued to go about the normal business of banking, with the exception that they would not pay out specie for deposits. However, they would certify checks for their depositors. This basically made the check a liability of the bank and meant that the bank would pay the check when the suspension of payments ended if the bank were solvent at the time. According to Sprague, these certified checks circulated and could be discounted for specie. We interpret the fact that certified checks could be discounted against specie during periods of financial disruption as meaning that movers received less than a dollar for each dollar of their deposits during such periods.

Sprague designates four periods in which financial disruptions occurred: the panic of May 1884 (Gorton lists this as occurring in June), the financial stringency November 1890, the crisis from May through September 1893, and the crisis from October through December 1907. He does not present evidence of there being any discounts on certified checks for the panic of 1884 and the financial stringency 1890. However, for the crisis of 1893 he says that discounts on certified checks were mostly between 1 and 2 percent, with some of the discounts reaching 4 percent. For the crisis of 1907, he says that discounts were most in the range of 3/4 to 2-1/2 percent, with again some reaching the level of 4 percent.

Drawing inferences about the interest rates paid to nonmovers is more difficult. Our approach is to use the interest rate on call loans to provide some information about interest rates paid to nonmovers. Call loans were loans made to individual based upon stocks as security. They were referred to as call loans, because they were payable when demanded by the bank. It is well know that the interest rates on call loans tended to increase sharply and to be well above government bond yields during periods of financial disruption. [Should we put in a graph?] We think that the interest rates on such loans provide some information about the interest rate that nonmover depositors might have been able to earn, because presumably such households had the option of withdrawing their funds from banks and making short-term loans directly to other agents. Although they may not have been able to earn the interest rates on call loans, we think the returns on such alternatives would tend to move with call money rates.

Thus, we have the stylized facts that the certified checks were discounted and call money rates rose sharply during financial disruptions. In terms of our model, we interpret these are meaning that the interest rate paid to movers was below unity and that the interest rate paid to nonmovers rose sharply was about the rate of interest on loans during periods of financial disruption.

To see if the calibrated version of the model delivers predictions consistent with these stylized facts, we compute the interest rates paid to movers and nonmovers for our model designating the periods around Sprague's four periods of financial disruptions as periods in which the bad state actually occurred. The results are shown in Figure 7. From the figure we see that the interest rates paid to movers are usually around one percent or less. However, in the periods of financial disruption, these interest rates fall sharply and become negative, consistent with the stylized facts. Interest rates paid to nonmovers are usually very slightly below the interest rate on loans, but in periods of financial disruption, they rise sharply and are above the interest rate on loans during these periods. This is also consistent with the stylized facts.

The next question is how well the magnitudes of interest rates for movers and nonmovers during panics periods from our calibrated model match those observed in actuality. Interest rates paid to movers in panics periods are about a negative 7 or 8 percent in the calibrated model. As mentioned above, in actuality the discounts on certified checks were usually only 1 or 2 percent, although sometimes they got as high as 4 percent. Thus, it would seem that the calibrated model overpredicts the losses to movers in periods of financial disruption. However, it must be remembered that what is being reported for certified checks is the discount, the immediate loss that the holder of the certified check suffers when attempting to convert it to specie. The variable reported for the calibrated model is an interest rates, the amount that the mover receives at the end of the period as a percentage of the amount deposited. The time period in the calibration is assumed to be a year. The time period over which the discount on certified checks was actually realized by households during the period we are considering may have been much shorter. Hence, the differences between the calibrated model and the discounts reported by Sprague may not be all that different.

[Paragraph on call loan rates]

5.3. Note circulation

Finally, we see how well the calibrated version of our model predicts the level and movements of actual note issue during this period. Our method is similar to that used for the long term government bond yields, with one major exception. The interest on loans in the model economy is invariant to the state of the economy; banks' note issuance is not. Thus, we have to make an assumption about when the actual economy was in the bad state. We assume that the actual economy was in the bad state in those period identified as periods of financial disruption above.

Comparisons of $[B(t,j)+N(t,j)]/D(t)$ from a calibrated version of the model with actual values over the period 1882 to 1910 for the same three values of c_N used above are shown in the table below.

	Model			Actual
	$c_N = 1.10$	$c_N = 1.15$	$c_N = 1.20$	
Mean	0.045	0.044	0.044	0.073
Variance	0.00009	0.00007	0.00006	0.00108
Correlation	0.225	0.227	0.215	

The results do not look as good as those for interest rates. The means of the calibrated series for note issues are only about half those of the actual series. Further, the variances of note issue for the calibrated model are more than one hundred times smaller than actual note issues. Thus, the calibrated model demonstrates even more inelasticity of note issue than was actually the case. Lastly, the correlation between the model's predictions and actual is quite low.

The difficulties that the model has predicting note circulation is illustrated in Figure 8, where we show the comparison using $[B(t,j)+N(t,j)]/D(t)$ computed with a value of $c_N = 1.15$ in the calibration. The calibrated model badly misses the fall in note circulation from 1882 to 1890, although it does capture to some extent the rise in note circulation after 1900. We checked the extent to which the early miss in circulation was the reason that the model appeared to do so poorly. When the period from 1882 to 1890 is eliminated from the comparison, the difference between the means of the calibrated model and the actual series is reduced, but the actual is still larger than the model. The variance of actual note circulation falls dramatically, and the variance of actual and model become virtually the same. However, the correlation between actual and model falls to below 0.1, indicating that there seems to be no relation between actual note circulation and that predicted by the model.

6. Conclusion

The so-called National Bank note paradox is that a large and persistent arbitrage opportunity seemed to go unexploited from the time the National Banking System was established until the establishment of the Federal Reserve System. This unexploited arbitrage opportunity was that the yield on government bonds was consistently above the cost of issuing national bank notes.

Several authors have attempted to explain this paradox. In this paper, we focus on the paper by CSW. They have attempted to explain this paradox by arguing that national bank note issue was inelastic due to "the mechanics of note issue." We agree with CSW that national bank note issue did not fluctuate much over short periods of time, but we point out that note displayed large secular fluctuations. Thus, we disagree that whatever inelasticity there was in national bank issue was due to exogenous factors. Instead, we think that there were times when national banks could rationally expected that any additional notes they might issue would be quickly presented for redemption. Because there were costs associated with note redemption that were independent of the time that notes were in circulation, these notes were more expensive to issue than notes that would be in circulation for longer periods.

Thus, we built a model in which national banks could issue two types of notes – cheap notes and expensive notes – and in which there was an exogenous upper bound on the quantity of expensive notes that can be issued. We found that such model implies that for certain parameter values, interest rates can be above the cost of issuing cheap notes and that interest rates can fluctuate in response to changes in endowments. When we took a calibrated version of the model to data, we found that it did quite well. It was able to predict both the level of and changes in the yields on long term-government bonds during the period. However, it did not do as well in accounting for fluctuations in note issuance. The model displayed more note inelasticity than the actual data.

References

[to be added]

Figure 1 -- Long Term Govt Bond Yields, 1882-1914

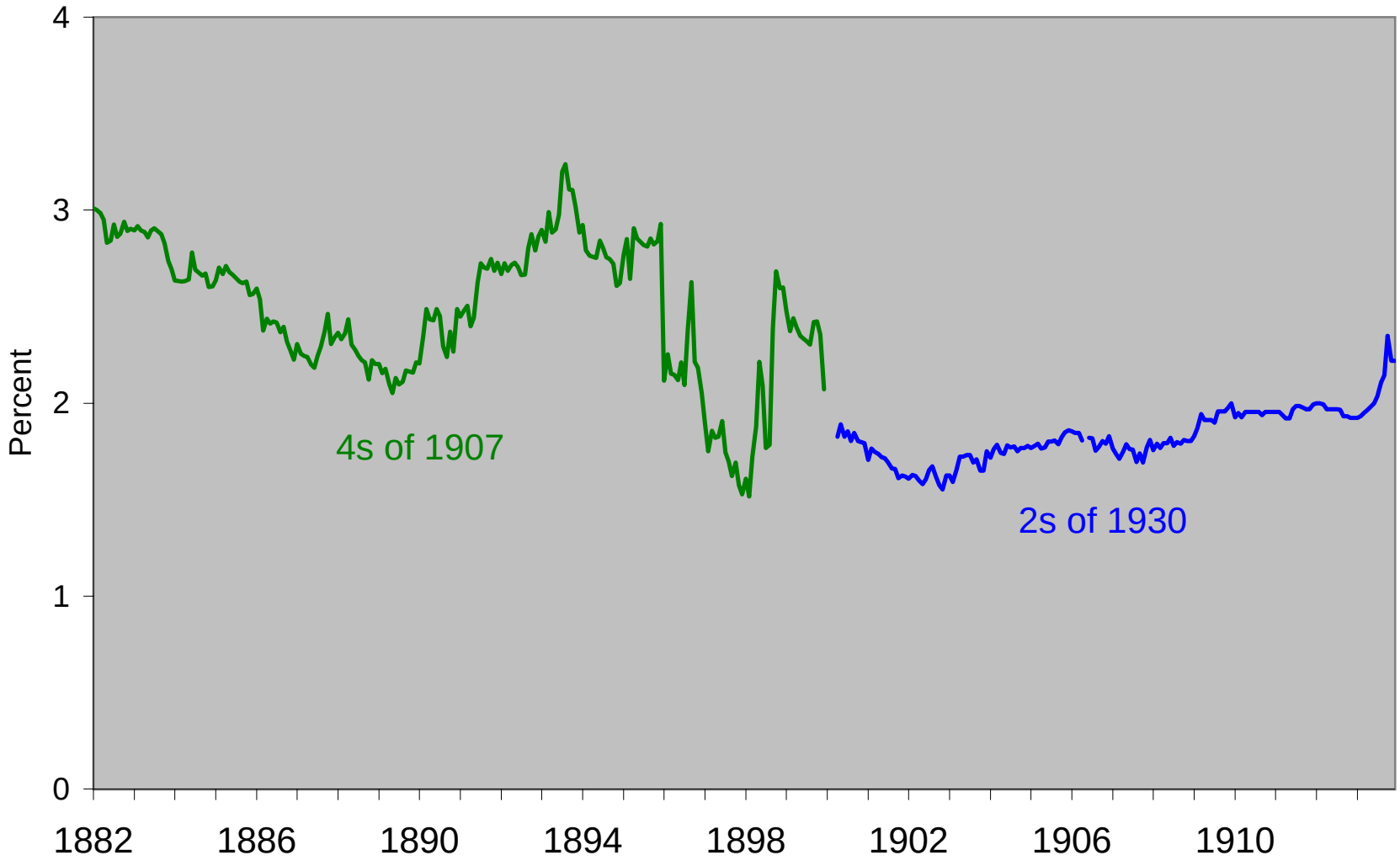

Figure 2 -- Percentage changes in National Bank Note Circulation, 1882 - 1910

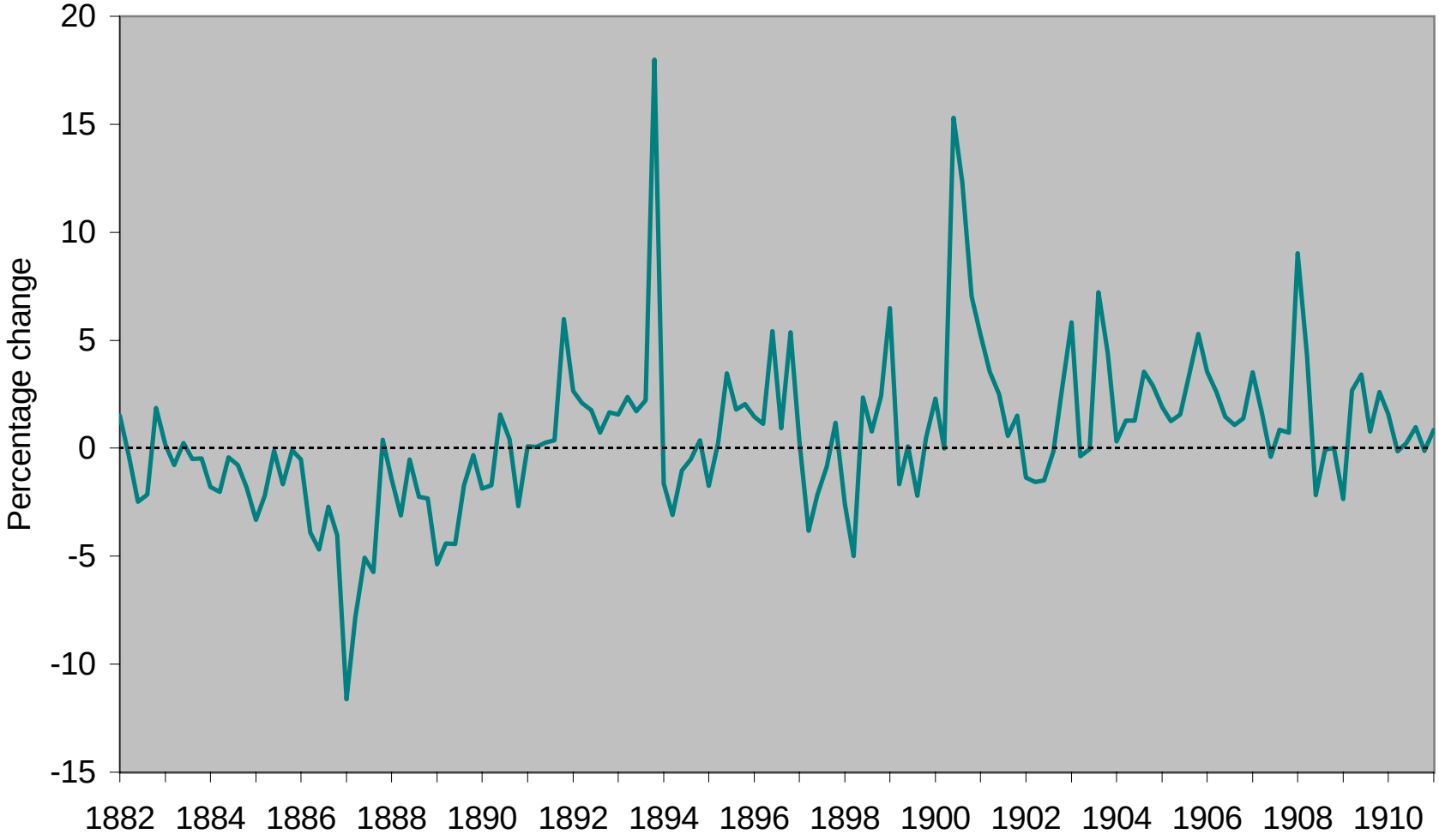

Figure 3 -- National Bank Note Circulation, 1882 - 1913

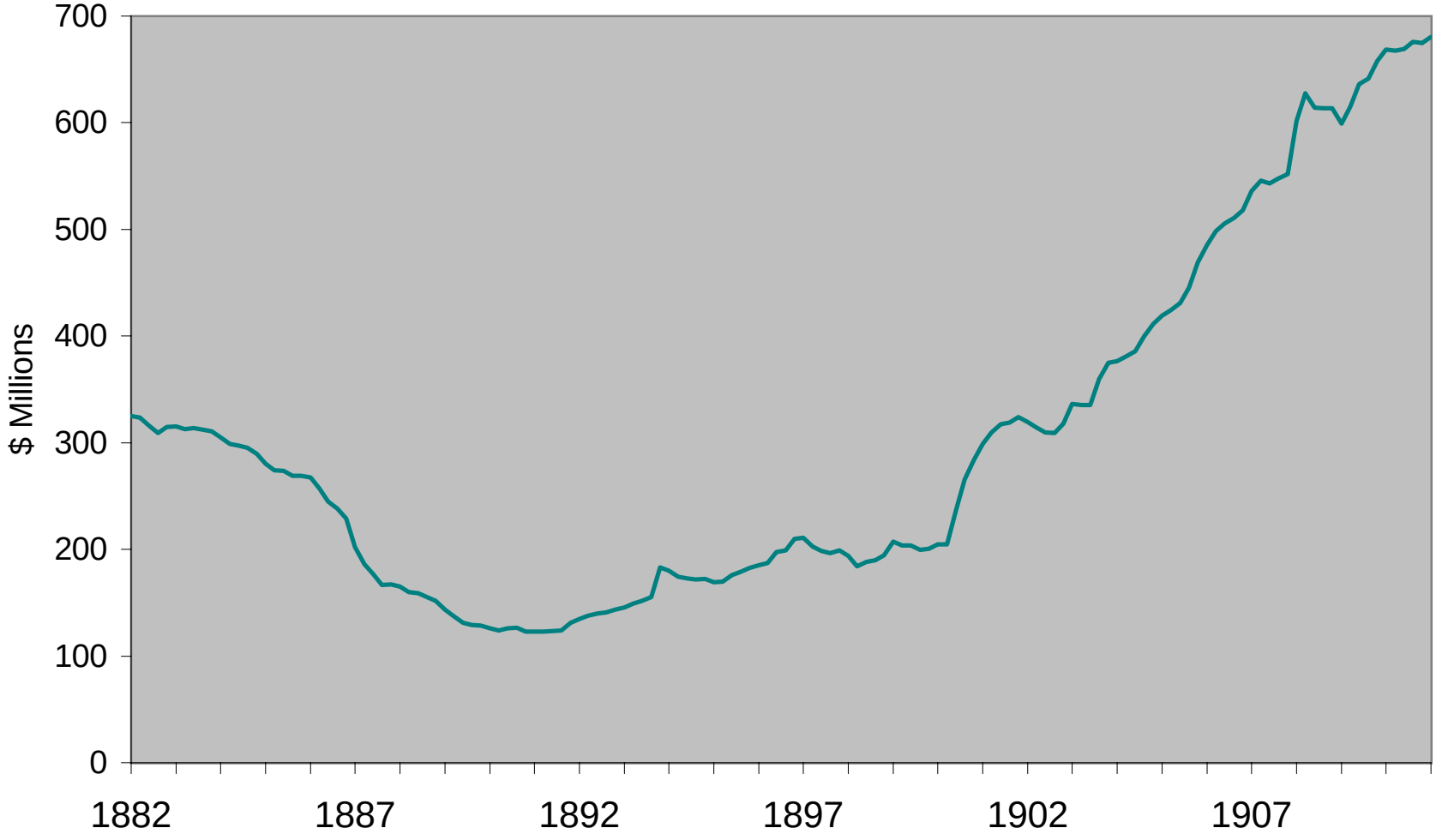

Figure 4 -- Representative Model Interest Rates

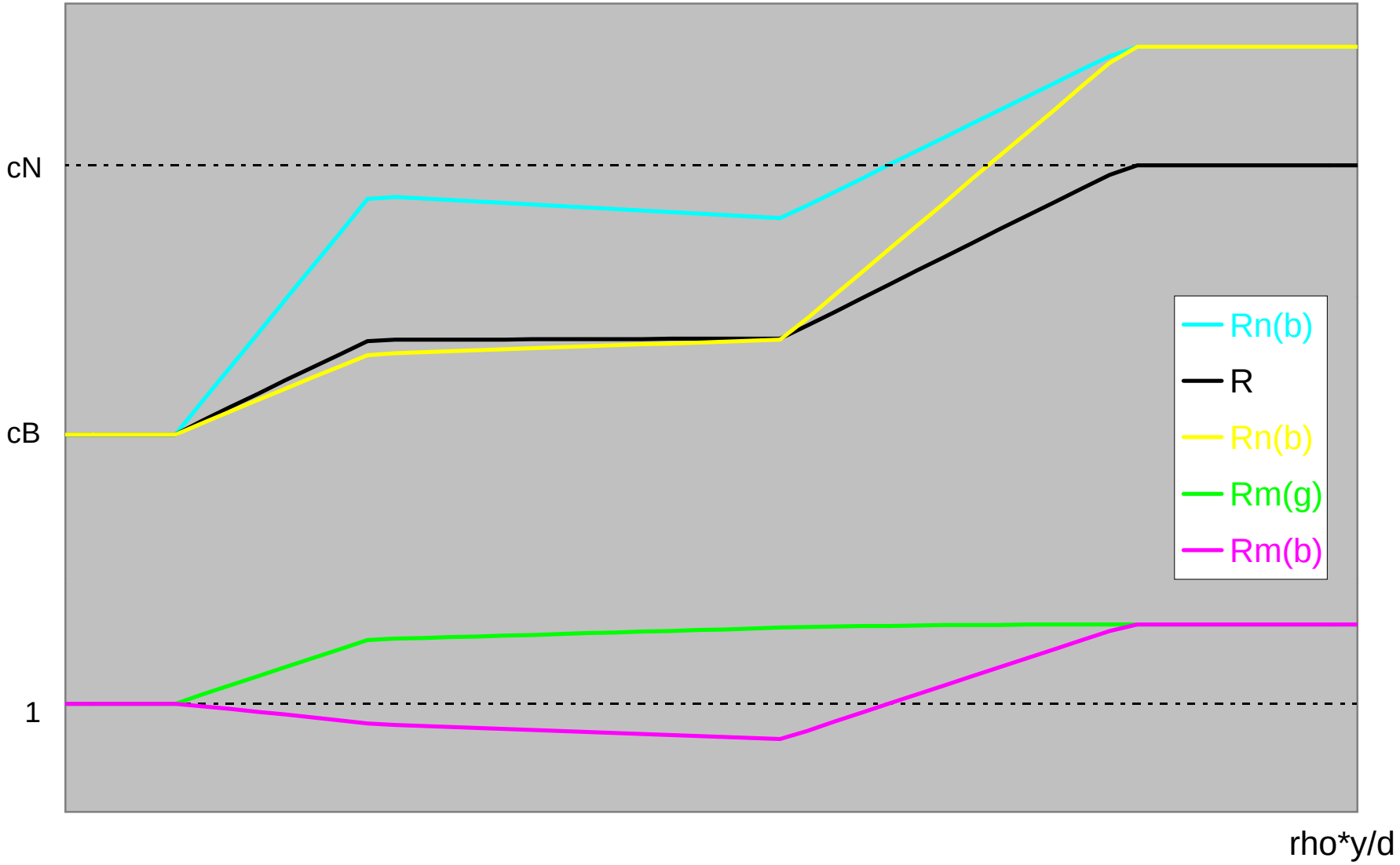

Figure 5 -- Ratios of model variables, 1882 - 1910

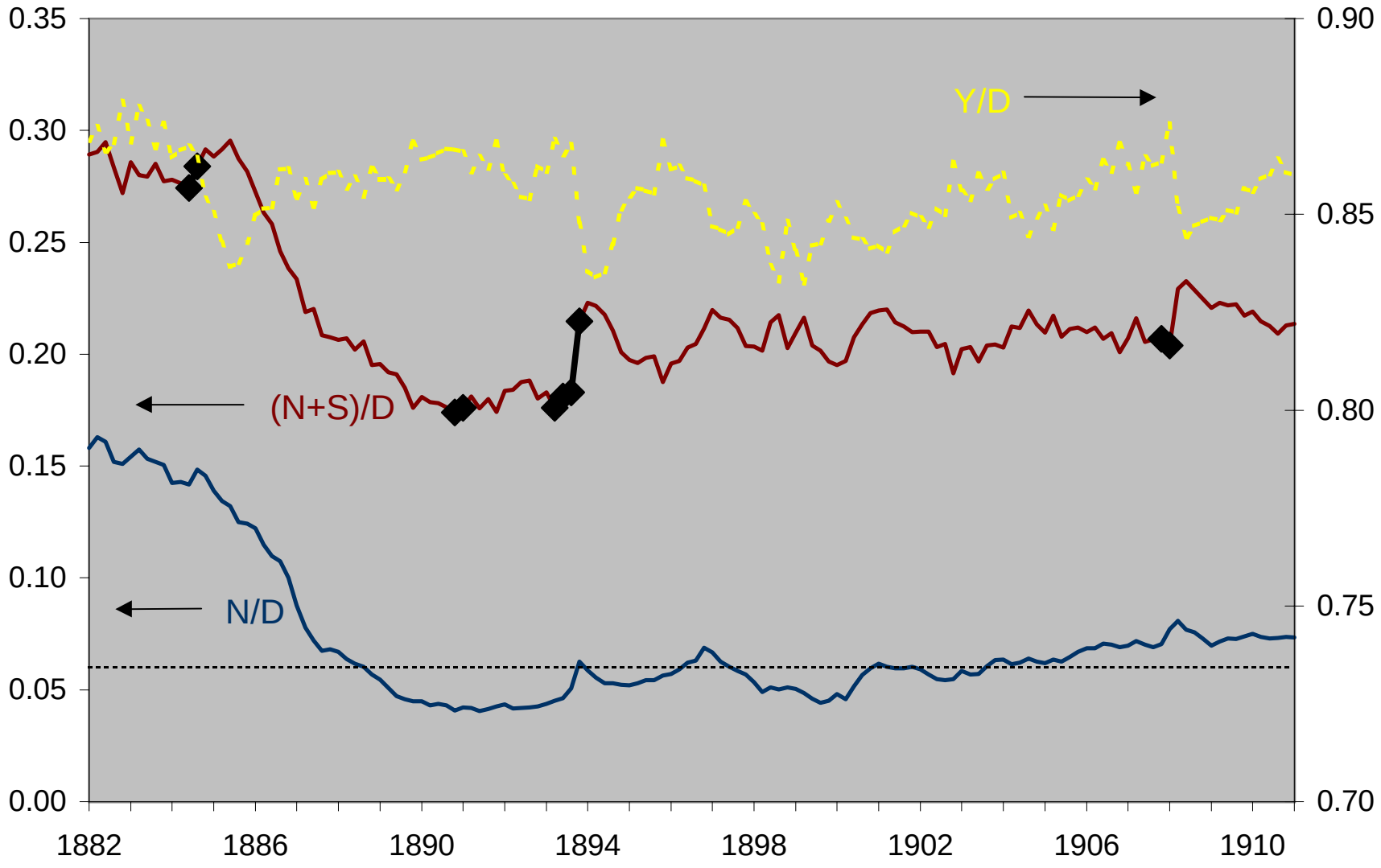

Figure 6 -- Interest rate comparisons

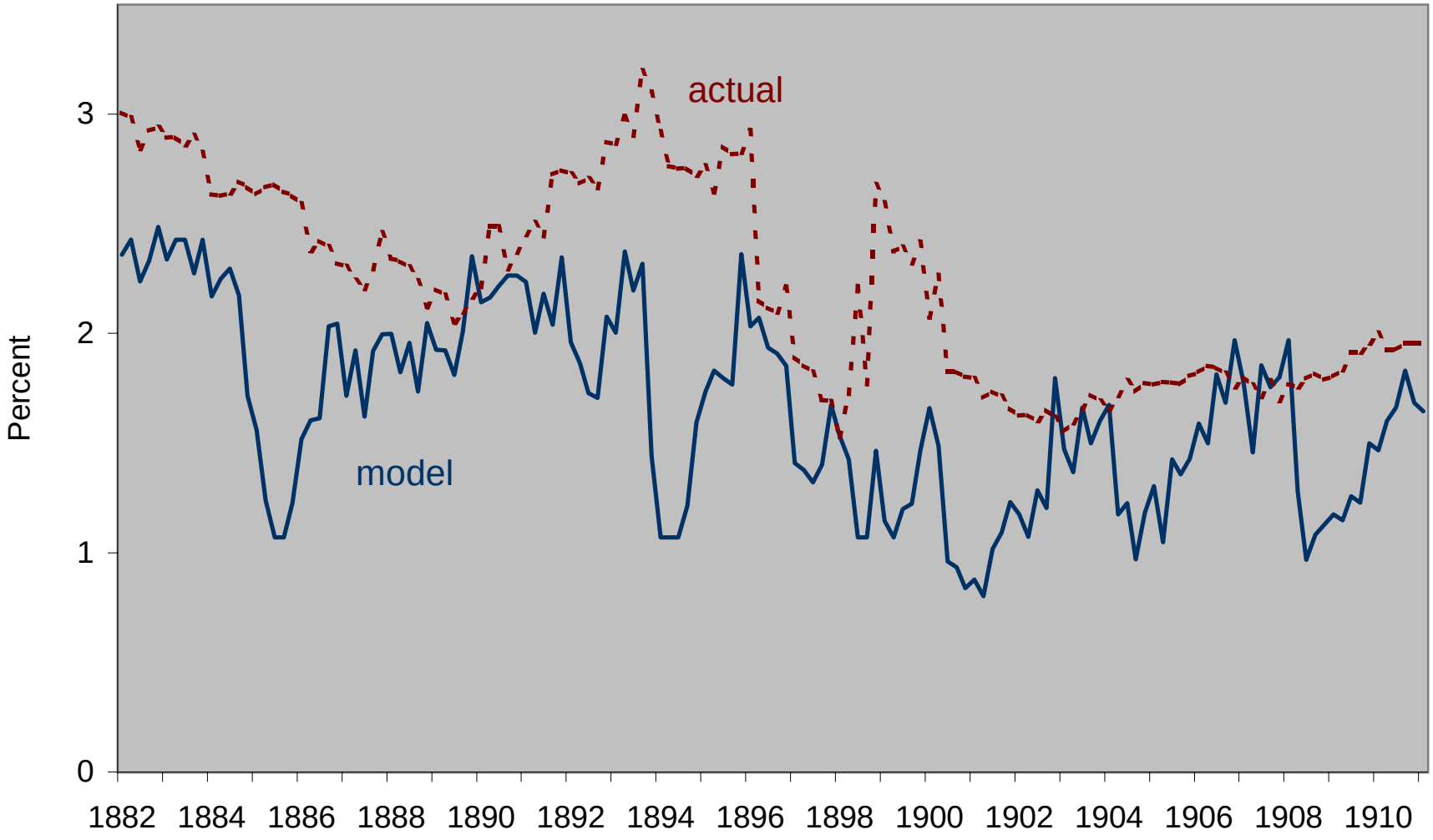

Figure 7 -- Predicted interest rates from model

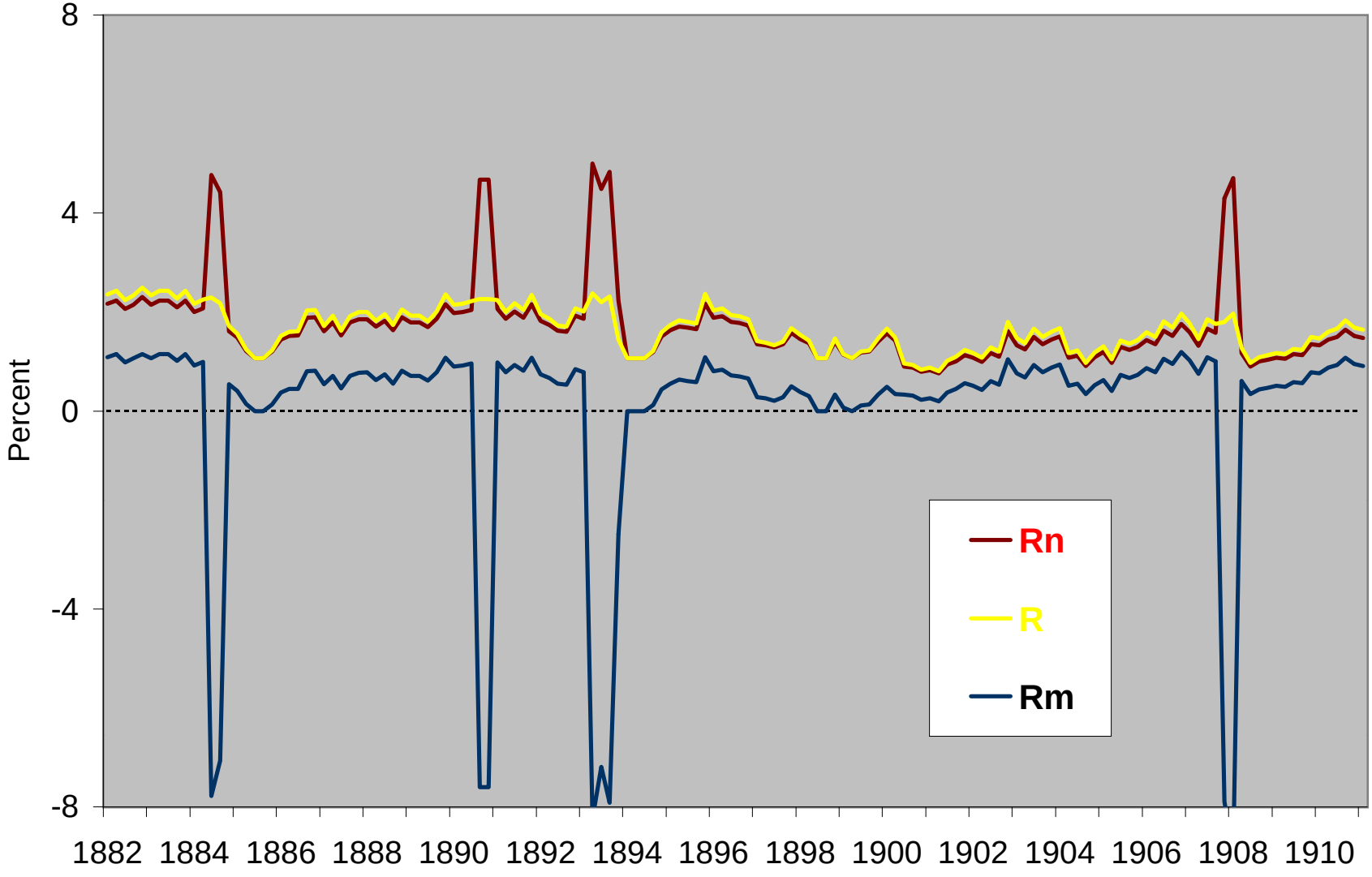
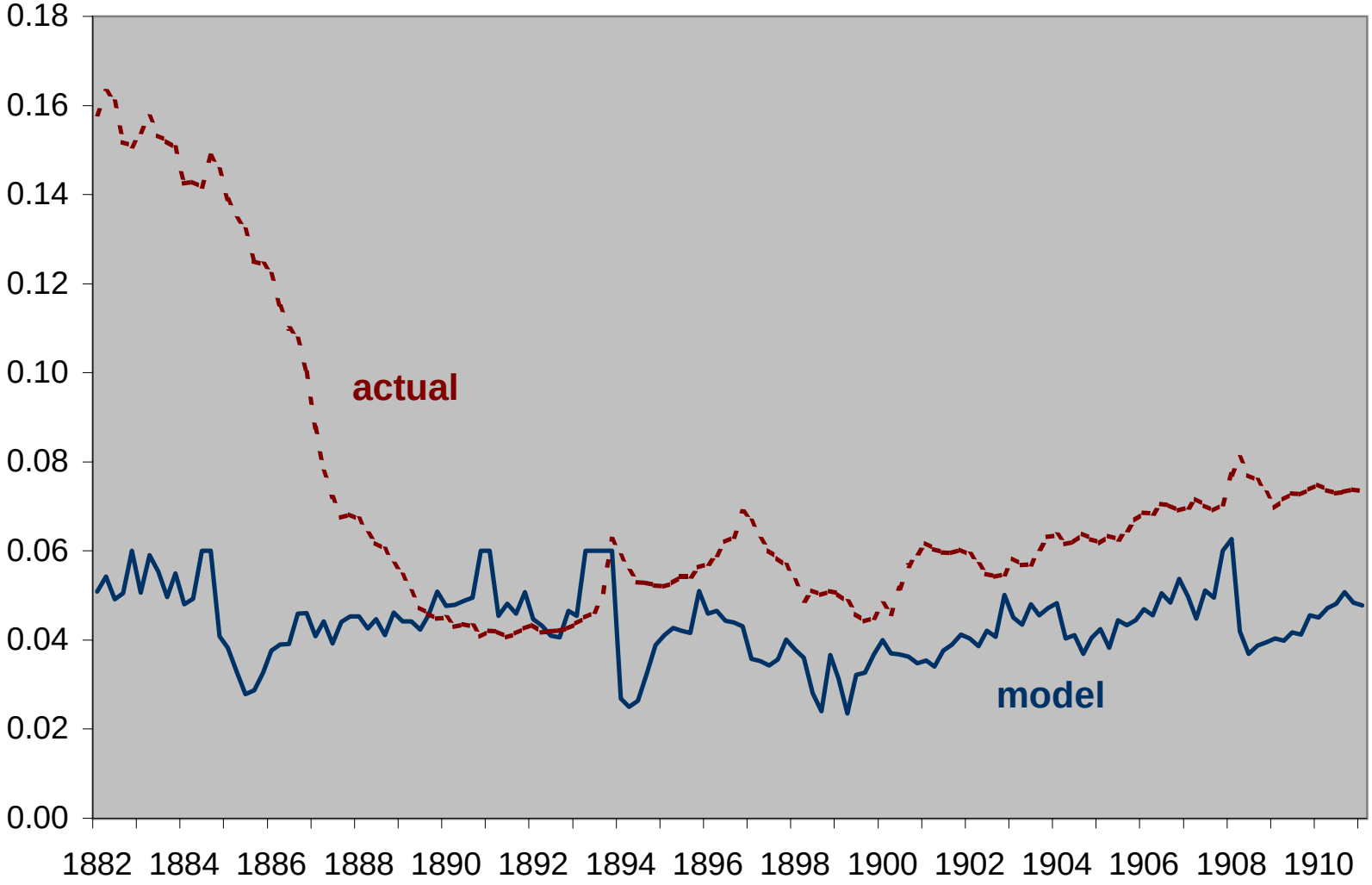
Figure 8 -- Note circulation comparisons