working paper

07 22R

The National Banking System: The National Bank Note Puzzle

by Bruce Champ

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 07-22R May 2011*

The National Banking System: The National Bank Note Puzzle

by Bruce Champ

The era of the National Banking System (1863–1913) has been a puzzling one for monetary theorists and economic historians for well over a century. The puzzles associated with this period take various forms. Despite calculations of high profit rates on note issue for certain periods of the era, national banks never fully utilized their note-issuing powers. Relatedly, the behavior of interest rates during the period is also puzzling given the regime of bank note issuance put in place by the National Bank Acts. On the surface, it appears that an arbitrage condition is broken. The observed inelasticity in aggregate national bank note issue also is puzzling, particularly given the behavior of interest rates. This paper examines many of the puzzles of the national banking era and provides a summary of the current attempts to explain those puzzles.

This paper represents a preliminary chapter from a forthcoming monograph on the period of the National Banking System. Other chapters of the monograph appear in the Federal Reserve Bank of Cleveland's working paper series as working paper 07-19R and working paper 07-23R.

Key words: bank notes, interest rates, National Banking System, national banks.

JEL codes: E42, E43, G21, N21.

*Originally posted December 2007. First revision posted September 2010.

Bruce Champ is an economist at the Federal Reserve Bank of Cleveland. He can be reached at bruce.a.champ@clev.frb.org

Chapter 5

The National Bank Note Puzzle

This monograph would not be complete it if did not take a detailed look at what has often been referred to as the "national bank note puzzle."[1] In fact, the national bank era is marked by several odd features. The period has perplexed monetary historians and theorists for nearly a century. In fact, Spurgeon Bell was the first to clearly state the puzzling nature of national bank behavior in an important 1912 article, in which he investigated the profitability of issuing national bank notes. On the surface, the system seems simple and appears to have some obvious implications regarding the behavior of interest rates and note issue during the period. However, we will see in this chapter that these implications are not borne out by the behavior of the data during this period. Despite more than a century of research, several aspects of the puzzles still remain, although progress has been made in recent years.

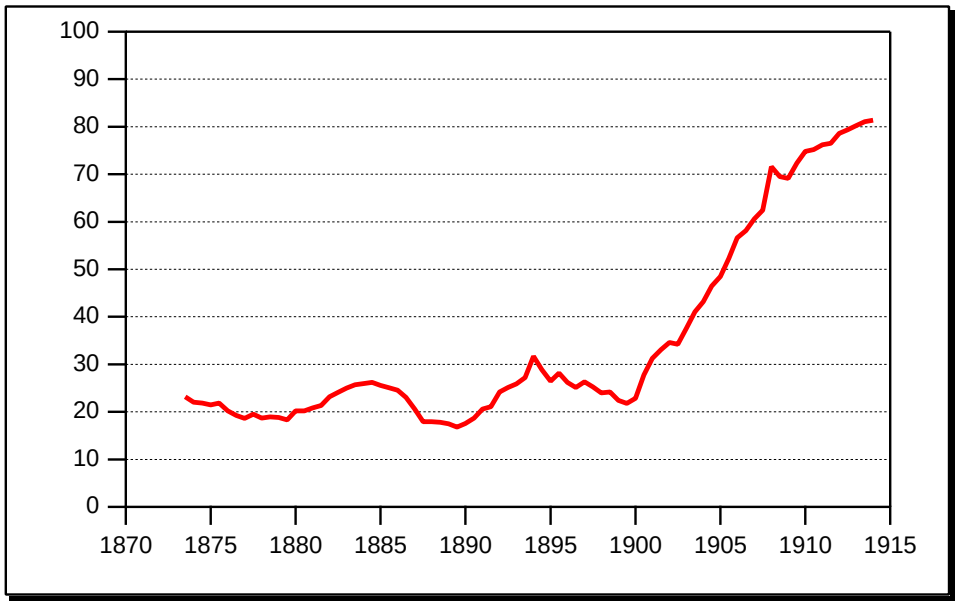
5.1 The Puzzle Defined

Recall that national banks could issue notes fully backed by U.S. government bonds. If we look at the data, we find that national banks never fully utilized their note-issuing powers. If banks had fully done so, then they would have completely intermediated the bonds that were eligible as backing. As we in see in Figure 5.1, they never did

[1]This chapter was revised December 5, 2010.

so. In fact, for the period prior to 1900, banks only held 20–30% of the bonds eligible

Figure 5.1: Percentage of U.S. Government Bonds Held as Backing for National Bank Notes (1874–1914)



Source: Commercial and Financial Chronicle (various years).

as backing for national bank notes. If banks had fully intermediated the bonds, then the percentage should have been close to 100% throughout the period.[2] I will refer to this phenomenon as the "underissuance of national bank notes."

Underissuance of national bank notes appears paradoxical since, by previous calculations, issuance of these notes was a relatively profitable enterprise for national banks.

[2]Some financial intermediaries other than national banks were legally required to hold U.S. government securities. Also, national banks were required to hold U.S. government bonds as collateral for U.S. government deposits. For these reasons, it is unreasonable to believe that national banks would have held *all* the securities eligible for backing.

5.1.1 Cagan's profit rate on national bank note issuance

In two papers written during the 1960s, Phillip Cagan presented a formula which purports to measure the profit rate on the issuance of national bank notes. This formula is presented in equation 5.1.

$$r_c = \begin{cases} \frac{r_b p - \tau\alpha \min(p,1)}{p - \alpha \min(p,1)} & \text{if } p > \alpha \min(p,1) \\ \infty & \text{if } p = \alpha \min(p,1) \end{cases} \tag{5.1}$$

where

r_c = Cagan's annual rate of return on the issuance of national bank notes
p = price of the bond held in backing, in dollars (assuming a par value of \$1)
r_b = annualized yield to maturity on the bond held as backing
α = fraction of the value of a given deposit of bonds that could be issued as notes
τ = annual expense in dollars of issuing $\alpha \min(p, 1)$ in notes

Note that $\alpha \min(p, 1)$ denotes the amount of notes that are returned to the issuing bank by the U.S. Treasury from the deposit of an eligible bond with price p. The variable τ denotes the annual cost of issuing notes. It consists of the tax on note issue (for \$1 in notes this is \$0.01 before 1900 and \$0.005 on 2% coupon rate bonds after 1900) as well as miscellaneous costs of note issue.[3] Cagan uses an estimate provided by the Comptroller of the Currency for the miscellaneous costs of note issue of 0.00625 for a one-dollar deposit in government bonds.[4] Translating these costs to one dollar's worth of note issue implies that

[3]The miscellaneous costs of note issue include the costs of redeeming national bank notes as well as the cost of the plates used to print notes.

[4]Comptrollers often used a value of \$62.50 for the costs associated with issuing the amount of notes a \$100,000 bond purchase would allow. The \$62.50 figure was broken down into the following components: cost of redemption, \$45; express charges, \$3; plates, \$7.50; and agents' fees, \$7.

$$\tau = \begin{cases} 0.01 + \frac{0.00625}{\alpha} & \text{for bonds with a 1\% tax on circulation} \\ 0.005 + \frac{0.00625}{\alpha} & \text{for bonds with a 1/2\% tax on circulation (for 2\% bonds after 1900)} \end{cases} \tag{5.2}$$

Also note that for most of the the period, eligible bonds sold above par so that $\alpha \min(p, 1) = \alpha$.

The numerator of Cagan's formula (equation 5.1) represents the dollar return from the bond held as backing less the total costs associated with note issue. The denominator represents the amount of capital tied up in the process of intermediating the bond into national bank notes. The amount of capital tied up is the difference between the price of the bond and the amount of notes received from the U.S. Treasury.[5]

As an example, suppose that a bank in 1890 (implying that $\alpha = 0.9$) purchased a government bond for \$1.10. This bond had a yield to maturity of 4 percent. In this case, the total cost of note issue is $\tau = 0.01 + \frac{0.00625}{0.09} \approx 0.01694$. This implies that the profit rate on issuing notes backed by this bond is

$$r_c \approx \frac{(0.04)(1.10) - (0.01694)(0.9)}{1.10 - 0.9} \approx 14.375\%.$$

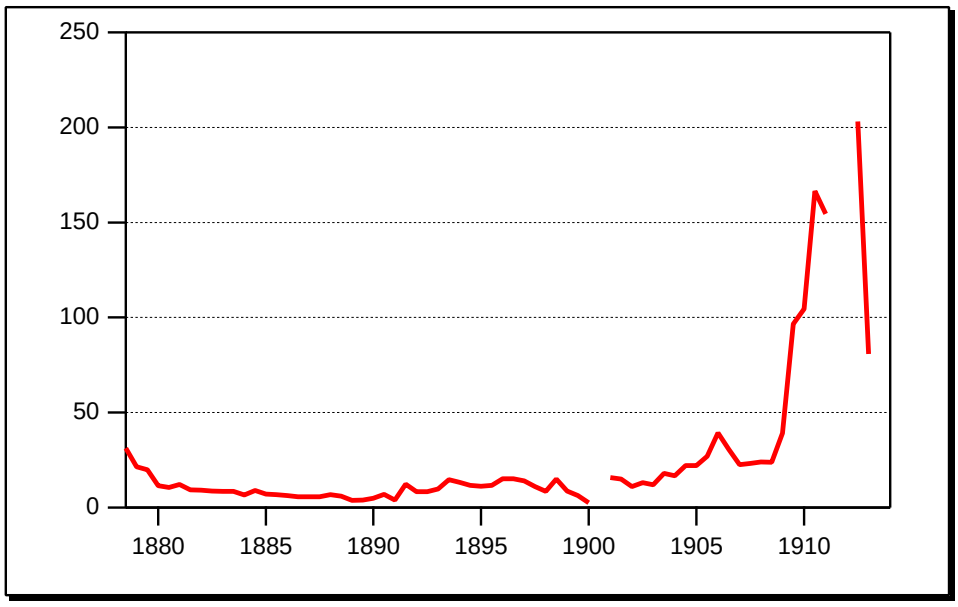
Cagan (1963) calculates annual rates of return on note issue and finds profit rates of 20%–30% for the late 1870s. Goodhart (1965), using Cagan's formula, calculates profit rates of 18% to infinity during the period 1901 to 1913. An infinite profit rate could only occur in Cagan's formula after 1900 (when $\alpha = 1$) and when the backing bond sold below par. In such a case, the amount of notes returned to the issuing bank would be exactly equal to the price paid for the backing bond, implying

[5]Bell (1912) originally suggested this was the appropriate measure of capital tied up in the intermediation of eligible bonds.

that the amount of tied-up capital would be zero. Presumably, the bank could earn infinite profits by using the acquired notes to purchase additional government bonds *ad infinitum*.[6]

Most certainly, the implied profit rates are, at times, far in excess of rates of return on alternative uses of bank capital. Figure 5.2 shows my calculations of the profit rates on issuing national bank notes using the formula suggested by Cagan. This

Figure 5.2: Cagan's Profit Rate on National Bank Note Issuance, 1878–1913 (percent per annum)



Source: Author's calculations using Cagan's profit rate formula, equation 5.1. The bond price data comes from various issues of the *Commercial and Financial Chronicle*. The costs of note issue were compiled from data presented in the *Annual Report* of the Secretary of the Treasury.

data portrayed in this figure uses more accurate representations of the costs of note issue than those estimated by the Comptroller of the Currency. Rather than being constant as in Cagan's original formulation, the cost of note issue estimates presented

[6]Kuhlwein (1992) criticizes this view by claiming that dealers in government bonds may not have always accepted national bank notes in payment.

here vary over time. Note the gaps in the graph. These gaps correspond to infinite profit rates as measured by Cagan's forumla. All of these instances occur when the the prices of 2% coupon bonds fell below par.

On the surface, profit rates far in excess of alternative uses of bank capital are puzzling. One would have expected banks to pursue the relatively more profitable enterprise of issuing notes. To do so would require the purchase of eligible bonds, which, in turn, would put upward pressure on bond prices. This would cause the return on note issue to fall. One would assume that this process would continue until the return on note issue would be brought in line with that on alternative uses of bank capital, adjusted for risk.

The fact that that national banks did not fully exploit their note-issuing powers in the face of apparently high profitability of pursuing this enterprise is the main puzzle of the national banking era. As Cagan notes, "It is nevertheless puzzling, why, in view of the large profit in issuing notes after the mid-1890s, their expansion occurred so slowly and never reached 100 percent of the amount allowed." (Cagan 1965, p. 94) Friedman and Schwartz (1963) also noted the puzzling nature of national bank note issuance when they stated, "The fraction of the maximum issued fluctuated with the profitability of issue, but the fraction was throughout lower than might have been expected. We have no explanation for this puzzle." (Friedman and Schwartz 1963, p. 23) Friedman and Schwartz go on to state, "Either bankers did not recognize a profitable course of action simply because the net return was expressed as a percentage of the wrong base, which is hard to accept, or we have overlooked some costs of bank note issue that appeared large to them, which seems much more probable." (Friedman and Schwartz 1963, p. 24) When Friedman and Schwartz mention that "the net return was expressed as a percentage of the wrong base," they are referring to calculations performed by contemporary Comptrollers of the Currency, who found low profitability

of note issue during the era. The Comptrollers' calculations are discussed in the next section.

Cagan's findings of high profit rates would have surprised Secretary of the Treasury Carlisle who summarized the prevailing views concerning bank note profitability when he stated before the House Committee on Banking and Currency in December 1894, "It is well known, of course, that the profits of the circulation of a national bank constitute a very small item of the total profits of the institution." Certainly, one component of the solution to the national bank note puzzle is to reconcile the contemporary view that note issue was not very profitable with Cagan's findings of high profitability.

5.1.2 Deriving measures of the profitability of note issue

To better understand Cagan's profit rate and to clarify some of the issues surrounding it, let's look in detail at a bank's decision to issue additional notes. To do so, it is useful to look at a simplified version of a typical national bank's balance sheet. This is depicted in Table 5.1, where only the items most relevant to a bank's decision to issue notes are presented. (A more detailed version of a typical national bank's balance sheet appears in Section 2.4.)

Table 5.1: Balance Sheet of a Note-Issuing National Bank

Assets	**Liabilities**
Reserves	Bank notes in circulation (N)
Eligible bonds (B)	Deposits
Other earning assets (A)	Net worth
	Paid-in capital)(K)
	Surplus (S)

Consider the purchase of a government bond with price p and par value 1. Let's

look at the effect of this on the bank's balance sheet. Here, $\Delta B = p$. These bonds are deposited with the U.S. Treasury and notes are returned to the bank. The amount of notes returned is $\alpha \min(p, 1)$, so that $\Delta N = \alpha \min(p, 1)$. The difference between the bonds purchased and the notes issued is $\Delta B - \Delta N = p - \alpha \min(p, 1)$. This amount is either positive or zero depending on p and α. When this difference is positive, it must be financed either either by a decline in other assets, reserves, or an increase in bank capital. For now, we will assume that it is financed by a reduction in other assets. Here $\Delta A = \Delta N - \Delta B$.

What is the effect of this on the banks profits? There are a number of effects. First, there is interest income on the bonds held as backing. If the annual yield on the bond is r_b, the total interest income will be $r_b p$. However, the decrease in other earning assets causes a decline in revenue. If r_a is the annual yield on other assets, the decreased revenue from the decline in earning assets is $r_a[p - \alpha \min(p, 1)]$. Furthermore, the bank incurs costs of issuing notes. This amount will be $\tau \Delta N = \tau\alpha \min(p, 1)$. Adding all this up implies a change in profits ($\Delta\pi$) of

$$\begin{aligned} \Delta\pi = \; & \text{Interest income on bonds held} \\ & - \text{Interest foregone due to decline in assets} \\ & - \text{Costs associated with note issue} \\ \\ = \; & r_b p - r_a[p - \alpha \min(p, 1)] - \tau\alpha \min(p, 1). \end{aligned} \tag{5.3}$$

Profit calculations by the Comptroller of the Currency

In the late 1800s, various Comptrollers of the Currency made estimates of the profitability of issuing notes. They performed these calculations to argue that national

banks were not making large profits off the issuance of notes.[7] The comptrollers presented calculations of $\Delta\pi/p$, which they labeled the "profit in circulation in excess of 6 percent on the investment." For r_a, they used 6% (hence the label). The comptrollers found values for this typically between 0.4 and 1.3 percent. Because of these low values, they concluded that note issuance was not very profitable.

Cagan (1965) and Friedman and Schwartz (1963) criticized the Comptrollers' measurements, asserting that p is not the proper amount of bank capital tied up in the issuance of notes. Their view is that instead of $\Delta B - \Delta N = p - \alpha \min(p, 1)$ being financed by reductions in assets, it was financed by an injection of capital. As we have seen, Cagan measures the profit rate as the return on eligible bonds less the cost of note issue as a percentage of the amount of capital tied up (equation 5.1).

Accounting for idle notes

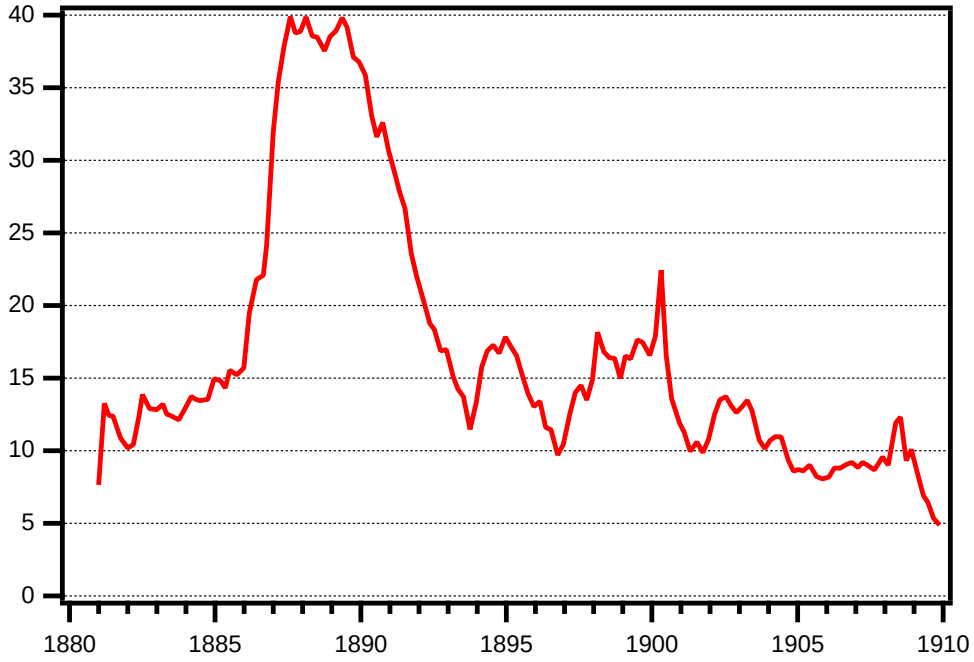
Champ, Wallace, and Weber (1992) present an important criticism of earlier profit rate calculations. The profitability calculations presented earlier in this chapter (equations 5.1 and 5.3 assume that all notes were constantly in circulation, an assumption that is counterfactual. Implicitly the assumption is that since the notes were used to by interest-bearing assets (such as making loans), they were always earning the bank the rate of return r_a. However, idle notes did not earn this rate of return. They earned nothing.

Notes were idle either when they were sitting in the vaults of the issuing bank or when they had been redeemed and were en route from the Treasury back to the issuing bank. The amounts of these idle notes were not trivial as seen in Figure 5.3.

[7] Contemporary critics of the National Banking System claimed that national banks made "double profits" from the interest earned on U.S. government securities held as backing for bank notes and from interest earned when the notes were lent out to borrowers. Comptrollers of the era frequently countered such claims by presenting profit rate calculations in their annual reports.

During the late 1880s and early 1890s, idle notes amounted to around 40% of the

Figure 5.3: Idle Bank Notes as a Percentage of Bank Notes Outstanding, 1881–1910



Source: Commercial and Financial Chronicle (various years).

national bank notes outstanding.

If we account for idle notes, we must alter our calculation of the effect of issuing notes on profits. Suppose that, on average, a bank expects that the fraction ϕ of notes outstanding will be in circulation at any point in time. Then the difference between bonds purchased and notes in circulation is no longer $\Delta B - \Delta N$, but is $\Delta B - \phi \Delta N$. Alternatively, this amount is $p - \phi\alpha \min(p, 1)$. Multiplying this by the rate of return on assets gives us the amount of interest foregone due to the decline in other earning assets. Second, the tax and miscellaneous were only levied on notes in circulation so that the total effect on profits due to the cost of note issue should now

be $\tau\phi\alpha \min(p, 1)$. So we should amend our calculation of the effect on profits to be

$$\Delta\pi^* = r_b p - r_a[p - \phi\alpha \min(p, 1)] - \tau\phi\alpha \min(p, 1). \tag{5.4}$$

One the one hand, idle notes reduce profits by not earning the rate of return on assets, r_a. However, idle notes lower the cost of note issue since idle notes are not taxed; the semiannual tax on note issue only applied to notes *in circulation*. If we compare this equation with our previous measure of the effect on profits, we can see the effect that ϕ has on profits.

$$\Delta\pi^* - \Delta\pi = (r_a - \tau)(\phi - 1)\alpha \min(p, 1). \tag{5.5}$$

Since $r_A > \tau$, then idle notes ($\phi < 1$) will imply that profits will be lower than when there are no idle notes. In other words, idle notes reduce the profitability of note issue.

Calculating a rate of return on equity

National banks were required by law to purchase a minimum amount of government bonds that was based on the amount of bank paid-in capital. A bank's total capital consisted of two elements—paid-in capital and surplus. Paid-in capital was that amount of a banks subscribed capital that had actually been paid into to the bank. Surplus was the accumulation of past retained earnings.

The requirement was that banks had to hold bonds equal to $\beta = 1/4$ of their paid-in capital, so that the minimum amount of bond holdings was βK. The law was a bit more complicated than that in reality, but this is what we will use for our computations. The maximum amount of note issue was based on the bank's capital. So, for a bank issuing with a paid-in capital of K, the total change in profits arising

from moving to the minimum amount of note issue was

$$\begin{aligned}\Delta\pi_E &= \left[\frac{K}{\min(p,1)} - \beta K\right]\Delta\pi^* \\ &= K\left[\frac{1}{\min(p,1)} - \beta\right]\Delta\pi^* \end{aligned} \tag{5.6}$$

Champ, Wallace, and Weber also felt that the Comptrollers choice of $r_a = 0.06$ is too high. For r_a, we used the bond held as backing with the highest yield to maturity. We also believed that the Comptroller's estimates of costs were too high. These costs should only include marginal costs in the calculation. Costs like plates (basically a one-time expense), which were included in the Comptroller's cost estimates should not be included. So, instead of using the Comptrollers estimate of $\tau = 0.01 + \frac{0.00625}{\alpha}$, we used:

$$\tau = \begin{cases} 0.01 + \frac{0.005}{\alpha} & \text{for bonds with a 1\% tax on circulation} \\ 0.005 + \frac{0.005}{\alpha} & \text{for bonds with a 1/2\% tax on circulation (2\% bonds after 1900)} \end{cases} \tag{5.7}$$

Calculating $\Delta\pi_E$, Champ, Wallace, and Weber obtain a median value of \$1,910 and conclude that this is not a large addition to profits.

To get a better handle on the magnitude, we calculate the value of $\Delta\pi_E$ relative to the size of bank net worth. This gives us a measure of the rate of return to equity, a more meaningful figure. Dividing $\Delta\pi_E$ by total capital to obtain the change in the rate of return on equity (Δr_E):

$$\begin{aligned}\Delta r_E &= \frac{\Delta\pi_E}{S+K} \\ &= \frac{\left[\frac{1}{\min(p,1)} - \beta\right]\Delta\pi^*}{1+\frac{S}{K}}. \end{aligned} \tag{5.8}$$

When we calculate Δr_E, we get quite small estimates. The median estimate is around 0.5%. This does not seem to be a large value since we ignore some of the costs of note issue.

5.1.3 The national bank note puzzle restated

We can look at the national bank note puzzle from either a bond price or bond yield perspective. Given the inverse relationship between bond prices and bond yields, the two perspectives are equivalent. Cagan and Schwartz (1991) note that the national bank note puzzle can be interpreted in such a manner—in particular with respect to the prices on U.S. government securities that were eligible as backing for national bank notes. They specifically ask the question: Why was the the price of eligible collateral so low? Cagan and Schwartz state, "The real puzzle, we show, is why the market prices of some of the eligible bonds did not reflect their value for securing note issues." (Cagan and Schwartz 1991, p. 293) Stated differently, why was the yield on eligible collateral so high during this period?

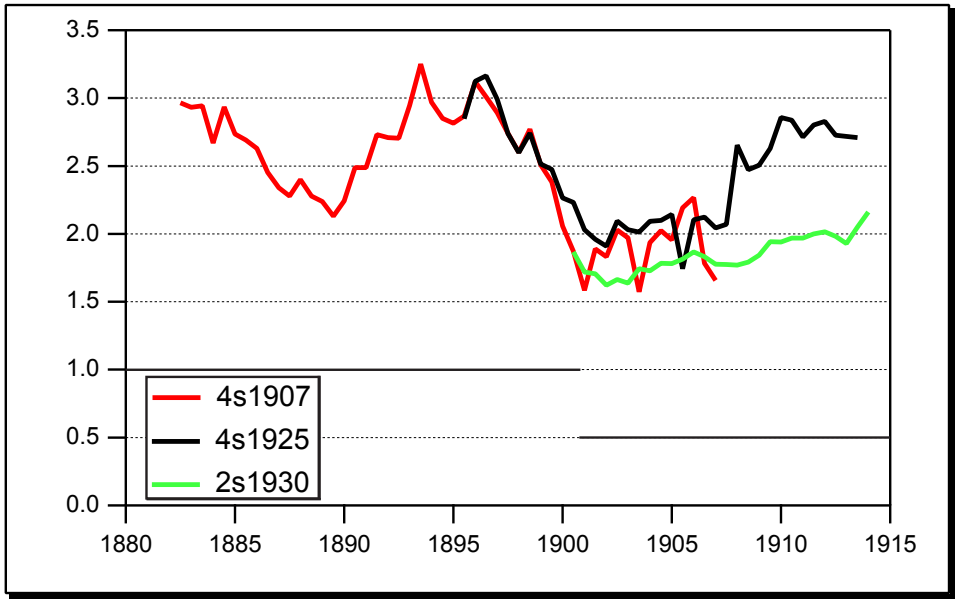
Essentially, national banks during this period could borrow from the federal government at the tax rate on circulation (1% before 1900 and, effectively, 1/2% thereafter) and receive notes from the U.S. Treasury. These notes could be used to purchase assets of various types. We would expect national banks to have exploited any interest rate spread between the tax rate and interest-bearing assets until the spreads were eliminated after accounting for risk. This implies that we should expect the yields on government bonds and safe short-term assets to be driven down to the tax rate on note issue.

There are times, typically in the fall, that short-term interest rates on call loans in New York exceeded 25%. It is puzzling that, during such times, national banks did not choose to expand their circulation in the form of loans in these markets. However,

as noted earlier, there is very little, if any, seasonal variation in note issue by national banks, at least before 1907. On the surface, it appears that an arbitrage condition is being broken.

We saw in Chapter 2 that the yields on government bonds far exceeded the tax rate on note issue. Figure 5.4 illustrates the yields on several classes of bonds that were important components of bank holdings eligible securities for backing national bank notes. The black horizontal lines in Figure 5.4 represent the tax rate.

Figure 5.4: Yields on Selected Eligible U.S. Government Securities, 1882–1913 (percent per annum)



Source: *Commercial and Financial Chronicle* (various years).

Recall that the tax rate on note issue was one percent before 1900 and, effectively, 0.5 percent after 1900. This implies that the yields on eligible securities exceeded the tax rate by 100 to 200 basis points throughout the period. It would seem that a national bank, set up purely as a means to intermediate government bonds, could

have exploited this differential to make profits. What are we missing?

5.2 Possible Explanations for the National Bank Note Puzzle

The observed calculations of high profit rates of issuing national bank notes in the face of low issuance of notes is puzzling. We have seen that we can also state the national bank note puzzle as one involving a perplexing spread between the tax rate on note issue and the yields of eligible bonds. Over the decades, several potential explanations for this puzzle have been provided.

5.2.1 Risks associated with national bank note issue

One area of research that attempts to explain the national bank note puzzle focuses on risks associated with note issue. Profit rate calculations ignore any impact of risk associated with note issue. Studies that pursue this route for an explanation have focused on risks associated with a Congressional termination of the circulation privilege and the riskiness of the eligible collateral for national bank notes.

Risk of revocation of the circulation privilege

Goodhart (1965) emphasizes that Cagan's profit formula does not take into account certain elements of risk or all of the redemption costs associated with bank note issuance. Goodhart begins by noting that national banks were uncertain regarding a continuation of the system in which they operated and that this has implications regarding Cagans formula. As Goodhart states,

> ...the profit formula of Cagan assumed that the legal arrangements controlling the issue of banknotes were, and were expected to remain, permanent. But there was always a probability, and after 1907 a probability, that the circulation privilege would be terminated. (Goodhart 1965, p. 94)

In this view, the fact that certain classes of U.S. government securities could be used as backing for national bank notes should have resulted in them having a premium over other classes of bonds. Goodhart argues that if Congress were to have terminated the right of national banks to issue notes, the prices of the bonds used as backing would have fallen. This risk must have been perceived by national banks but is not taken into account by Cagan's formula.

However, it should be noted that after 1907, when opposition to the National Banking System appears to have become strongest, the proportion of national bank notes outstanding to the legal maximum was at its highest levels. To resolve the paradox that note issue was relatively high during a period in which note issue appeared to be most risky due to the possibility of a termination of the circulation privilege, Goodhart refers to differing behavior between city banks and country banks.

He observes that while central reserve city banks reduced their circulation of notes from $82.4 million in 1908 to $78.3 million in 1913, reserve city and country banks actually increased their note issue from $531.3 million to $646.1 million over the same period. Thus, the net result was an increase in total notes outstanding by $110.7 million over the period. Goodhart suggests that relatively lower costs of bank note issuance for country banks caused this disparity in behavior.

The reasons for Goodharts differential cost argument lie in the procedure for redeeming national bank notes. If a particular bank found itself holding the notes of other banks and the public was not willing to absorb these so-called "redundant notes," the bank holding them could redeem the notes by sending them to the U.S. Treasury (before 1874 notes were redeemed through reserve agent banks).[8] The Treasury then would send lawful money (*e.g.*, greenbacks) to the redeeming bank. If neces-

[8]The reason that national banks were reluctant to hold these redundant notes is that they did not satisfy legal reserve requirements. Only "lawful money" satisfied such requirements. See Section 1.3.3 for details on legal reserve requirements during this time period.

sary, the issuing bank would then be required to replenish its redemption fund, which was to be maintained at five percent of the value of a bank's outstanding circulation. The Treasury would then return the bank notes to the issuing bank.

However, this redemption procedure was not without cost. As Spurgeon Bell, who introduced many of the arguments extended by Goodhart, wrote, "The redeeming bank must forego the use of funds in transit during the week or two weeks elapsing between the time of shipping and the time of receiving the redemption money."[9] The redeeming bank was also required to absorb the shipping charges to the Treasury.

However, an alternative to redeeming redundant notes existed for the country banks. Country banks could deposit these notes with a correspondent bank in a reserve city where the deposit would typically earn two-percent interest.[10] Reserve city banks could, in turn, send the notes to a central reserve city bank and also earn interest on the deposit. The end result was an accumulation of redundant national banknotes in the central reserve city bank vaults. As Goodhart points out, the viable alternatives facing a central reserve city bank were to redeem the redundant notes, which was costly, or reduce (or at least not further expand) their own circulation in the hope that the notes would be absorbed by the public.[11] Due to the costly nature of redeeming bank notes, Goodhart suggests that central reserve city banks chose the latter option; hence, the differing behavior of note issue referred to above.

However, central reserve city banks also chose to redeem national bank notes, as did other banks. Figure 5.5 portrays the percentage of outstanding national bank

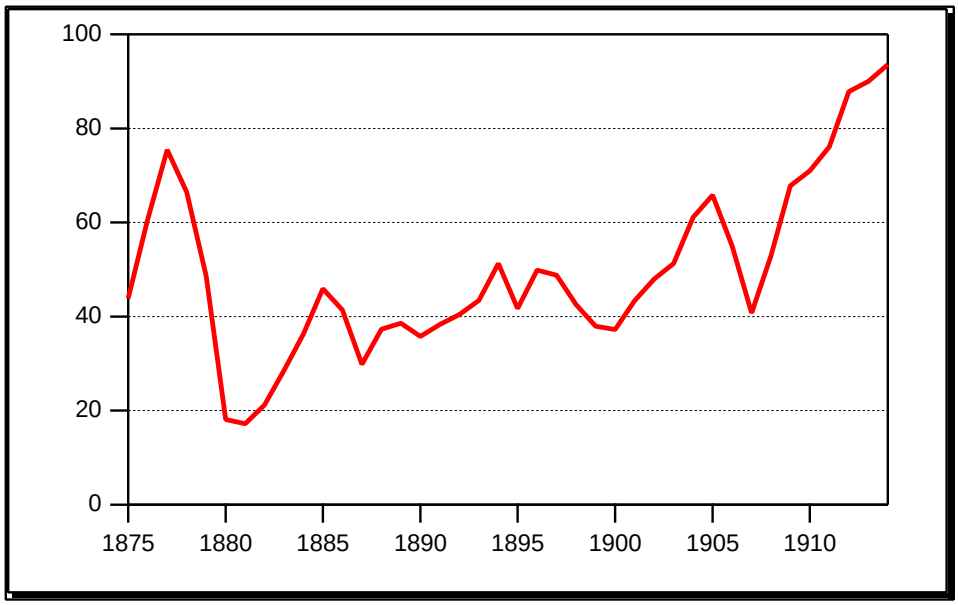
[9]See Bell (1912, p. 45). Given the current state of technology (*e.g.*, electronic funds transfers), it is easy to forget the significance role played by time spent in transit during the period of the National Banking System.

[10]The two-percent interest paid on such bankers' balances was remarkably constant over the period of the National Banking System.

[11]Goodhart actually refers to four alternatives open to banks. However, the two mentioned here appear to be the most reasonable alternatives. For a further discussion see Goodhart (1965, p. 521–22).

notes that were redeemed in each year from 1875 to 1913. Annual bank note re-

Figure 5.5: Redemptions of National Bank Notes as a Percent of Bank Notes Outstanding, 1875–1913



Source: Annual Report of the Secretary of the Treasury(various years).

demptions typically amounted to around 50% of the volume of notes outstanding, and about half of these notes were sent by large banks in New York City. Goodhart argues that the bulk of the redemption costs were borne by central reserve city banks. If this were the case, there must have been some incentive for central reserve city banks to attract notes to their vaults. Otherwise it is difficult to understand why central reserve banks paid two-percent interest on deposits from banks which used them as correspondents, an action which undoubtedly did indeed attract redundant notes. Also, if the city banks rid themselves of redundant notes in an indirect manner by reducing their own circulation, as Goodhart suggests they did from 1908 to 1913, this action required sending lawful money to Washington and waiting for the

return of the bonds held as backing. This procedure would have been just as costly as redeeming the redundant notes in the first place. Given the consequences, it is difficult to understand why city banks attracted the notes in the first place. Thus, it is unclear whether Goodhart‘s differential cost argument explains the differing behavior between city and country banks during this period.

In any case, it would appear that Goodhart‘s differential cost argument sidetracked him from the main issue. The national banking system as a whole assumed the costs of bank note redemption. However, only those costs which affected the profitability of note issue should be considered in explaining bank note underissuance. The redemption costs referred to by Goodhart had little relation to a bank's decision whether to issue additional notes. However, certain costs associated with the redemption procedure were borne by the issuing bank. These costs and their implications are detailed in the following section. A part of these costs which affected note profitability were ignored in Cagans profit calculations.

Champ (1990) extends the Bell-Goodhart analysis associated with the risk of a termination of the circulation privilege by Congress. I conclude that after examining the prevailing attitudes toward the National Banking System, it becomes apparent that national banks must have been uncertain regarding a continuation of the system throughout the period, not just after 1907. National banks would have been extremely naive to believe that the circulation privilege could not be revoked. Further, the existence of a positive probability that this privilege would be revoked has important implications for the issuance of national bank notes. In *A History of Currency in the United States*, A. Barton Hepburn summarizes the problems facing the National Banking System when he states

> The national banking system, designed, as we have seen, to give the people a permanent paper currency, lacked the necessary support even in the ranks of those who had created

> it. Only by the most strenuous efforts was the system able to survive the political attacks upon the one hand and the untoward conditions of a decreasing volume and enhancing prices of bonds, producing diminution of profits, on the other. (Hepburn 1915, p. 338)

Champ (1990) documents Congressional opposition to the National Banking System and provides estimates of the potential capital losses that national banks might incur had the circulation privilege been revoked.

Risky collateral

If banks encountered changes in the demand for national bank notes, they may desire to adjust the amount of notes outstanding and their holdings of U.S. government securities held as backing. Although banks could costlessly store idle notes in their vaults, they may have wished to routinely alter their holdings of U. S. government securities held as backing for notes. This means that banks may have been concerned with fluctuations in the prices of eligible bonds. Champ (1990) and Kuhlwein (1992) examine risk features of the bonds that national banks held as backing for national bank notes.

Since banks could only intermediate long-term U. S. bonds implies that banks had to deal with a fair amount of market risk in issuing notes. As Spurgeon Bell stated in 1912,

> The name of a government bond carries with it, for the great majority, the idea of security; but the large bankers who are familiar with the fiscal operations of the government and are accustomed to dealing in government bonds have found them an investment of unusual risk. If an officer of a large bank is asked why he does not issue more bank notes the reply will usually be that he does not want to risk such a large percent of the bank's capital in government bonds. (Bell 1912, p. 51)

Bell points to specific instances of dramatic changes in U. S. government bond prices during the period of the National Banking System. W.C. Mitchell also emphasized the instability of prices for eligible bonds in his 1911 study of yields on

railroad and government bonds: "Instead of providing the stablest of American securities from the investors' point of view, government bonds have proved to be the least stable among the bonds for which yields have been computed." (Mitchell 1911, p. 285)

5.2.2 The opportunity cost of issuing national bank notes

James (1976) critiques the works of Goodhart and Cagan and presents an explanation of the underissuance of national bank notes which hinges upon opportunity costs. The analysis of James is partially based on observations made by the Indianapolis Monetary Commission in 1898.[12] James suggests that Goodhart's argument that national banks underissued notes on account of risk associated with political opposition to the system "is limited to the period 1907–1913 and has little relevance to the question of low note issue in the 1890s; moreover, it is not an operational hypothesis." (James 1976, p. 360)

James asks, "If note issue was so risky after 1908, then why did the proportion of notes outstanding remain so high?" James refers to the differential costs argument which Goodhart proposed as a possible answer to this question. Again, Goodhart claims that country banks could avoid part of the costs of redeeming national bank notes by depositing the notes in a correspondent city bank. However, the city bank at the end of this chain of deposits had no such option. If the city bank chose to redeem the notes, it would forego the interest on these funds during the period between shipping the redundant notes and receiving the lawful money from the redemption fund (a period of about two weeks).

James then shows, through an example, that this foregone interest could not have

[12]See Taylor (1898) for a summary of the findings of the Indianapolis Monetary Commission.

been significant and "is a trifling one to explain the change in country and city bank issue of notes." James also points out that Goodharts differential cost argument is not substantiated empirically; the data does not appear to support it over certain time periods. However, even if one accepts the statement made by James that the difference in costs facing city versus country banks could not have been too significant, this is not to say that the costs associated with the redemption procedure that faced each individual bank were inconsequential. An individual bank, when deciding whether to issue additional circulation could not be certain that the public would absorb the notes. This added uncertainty as to the actual costs and, hence, profitability of issuing notes. Again, this matter and its implications will be discussed in a more satisfactory manner in the following sections.

After presenting a critique of Goodhart's analysis, James presents his own argument to explain the underissuance of national bank notes. James suggests that underissuance was caused "by the high opportunity costs of note issue in the form of [high] local loan rates in southern and western states." James argues that in areas in which local loan rates were relatively high, issuing notes would imply foregoing the high interest on the difference between the purchase price of the bond used as backing and the amount of notes that could be issued (this difference is merely $p - \alpha \min(p, 1)$ in Cagan's profit rate formula). James backs up his argument by demonstrating that the data tends to show a lower willingness for banks located in southern and western states to issue notes and that the local loan rates were indeed higher in these areas.

5.2.3 Redemption costs

Goodhart (1965) emphasized that variable redemption costs might be an important aspect ignored by previous authors attempting to explain the national bank note

puzzle. Two papers have emphasized the importance of redemption costs in explaining national bank behavior during this era.

Champ, Wallace, and Weber (1994) provide a possible reason to doubt some of the assumptions behind prior calculations of the profitability of national bank note issue. Prior descriptions of the period claim that the nonbank public viewed national bank notes and lawful money as perfect substitutes.[13] Champ, Wallace, and Weber refer to this as the "equivalence view" and discuss its implications. Because of their reliance on the equivalence view, prior studies implied that a national bank "could always get its own notes into circulation and, in effect, keep them outstanding." (Champ, Wallace, and Weber 1994, p. 344) However, Champ, Wallace, and Weber show that the equivalence view must be rejected.

Champ, Wallace, and Weber set up a model of note issuance in which the assumption of perfect substitutability is maintained. We show that if the collateral constraint on note issue is not binding (which, as noted previously, it was not during the period of the National Banking System), then yields on the collateral (government bonds) should be equal to the tax rate on note issue (1% before 1900, effectively $\frac{1}{2}$% thereafter). This, in turn, would imply that during this period the price of government bonds of a given class should have been fixed, independent of Treasury debt policies. Furthermore, we show that nonbindingness of the collateral constraint also implies safe short-term rates should have been pegged at the tax rate.

However, the data appear inconsistent with these implications. Yields on government bonds were 200 to 300 basis points above the tax rate throughout the period. Short-term interest rates were considerably above the tax rate and were highly variable. This calls into question the assumption that national bank notes and lawful money were perfect substitutes, a key assumption behind prior studies.

[13]See, for example, Friedman and Schwartz (1963, p. 21).

Champ, Wallace, and Weber (1994) suggest an alternative explanation for the national bank puzzle that hinges on redemption costs. In their view, national bank notes did not work as well for some purposes as for others. Since the costs of redeeming national bank notes were borne by the issuing bank, a bank would be concerned about how quickly its notes would be redeemed. Notes issued to buy securities in organized markets (the exact markets for which we have the high interest rate observations) probably would be quickly deposited in a bank and sent through the redemption process since national bank notes did not count as reserves for national. Hence, it is possible that, although a high rate of interest is earned, it is only for a short period of time, and the redemption costs outweigh the earned interest. So national banks would not be able to arbitrage away the differential between the interest rate on securities and the cost of issuing notes.

National banks would be more willing to issue notes in, say, a rural area where the notes might stay in circulation for a long period of time before being redeemed. Evidence will be given to support this assertion. As noted earlier, this is consistent with the significant number of large New York City banks that chose not to issue notes. The "float" on notes issued by them would be so short as to not be profitable. They turned to other operations (like being big players in the payments system) for profits.

This view is made explicitly in a model by Wallace and Zhu (2007).

5.3 Redemption Costs

Champ, Freeman, and Weber (1999) provide a model that explicitly incorporates costly bank note redemptions.

5.3.1 Aggregate versus disaggregate data

Calomiris and Mason (2006) claim that looking solely aggregate data can be misleading. It's a novel approach to the puzzle. They claim that even though banks *in the aggregate* are not fully intermediate the eligible bonds, there were limits on *single* banks that were binding. MORE HERE????

5.4 Other Puzzles from the National Banking Era

5.4.1 Inelasticity of national bank note issuance

The most common complaint voiced during the National Banking System concerned an inelasticity of the currency.[14] In fact, reference to any undergraduate textbook on money and banking which discusses the National Banking System will point to this as one of the major defects of the System. The concern with this defect was one of the main factors which provoked passage of the Federal Reserve Act, a major purpose of which was "to provide a more elastic currency." What was meant by currency inelasticity, and why did complaints regarding it exist? Furthermore, how can we account for these complaints?

The complaints regarding currency inelasticity came in two basic forms. One of these forms referred to bank notes which were backed by a fixed supply of bonds as being inherently inelastic.[15] This type of complaint was apparently based on the premise that the aggregate quantity of notes that could be issued was ultimately limited by the amount of bonds which were eligible as backing. This, undoubtedly, could have been a problem if banks were fully intermediating all the eligible bonds.

[14]Friedman and Schwartz also discuss the issue of currency inelasticity. See Friedman and Schwartz (1963, pp. 168–70).

[15]This criticism of the National Banking System remains to this day. See, for example, Meltzer (2002).

However, as noted before, this was never the case during the period under study. This is particularly true before 1900, when banks only held 20 to 30 percent of the eligible bonds. Thus, this form of the complaint was merely hypothetical; it referred to a possible problem related to note issue backed by bonds but not to a problem which actually existed during the period.

However, there is another sense in which critics viewed bank note issue as inelastic. This complaint regarded a view that bank note issuance did not expand and contract with "the needs of trade". As the National Monetary Commission stated,

> ... of our various forms of currency the bank-note is the only one which we might expect to respond to the changing need of business by automatic expansion and contraction, but this issue is largely dependent on the amount and price of U.S. bonds. (National Monetary Commission, Senate Document No. 243, 62nd Congress, 2nd Session, 1912, p. 7).[16]

This type of inelasticity is also difficult to understand given that banks were not fully intermediating the eligible bonds. Furthermore, it would, at first glance, appear that incentives existed for the issuance of notes to expand and contract as needed. As Cagan states, "When money tightened and interest rates rose, the price of U.S. bonds tended to decline.... When money eased, the opposite occurred." (Cagan 1965, p. 91) These incentives work in the right direction for providing an elastic currency. Thus, why did the complaints arise? We can only answer this question if there is some sense in which the restrictions placed on banks were binding.

I THINK THE BEST WAY TO PRESENT THIS IS TO DISCUSS WHY ISSU-

[16] As a humorous note, in another study, the National Monetary Commission attempted to measure the elasticity of the various components of the money stock. They found that national bank notes

> Do not appear to exhibit any considerable seasonal elasticity, i.e., rise and fall according to the seasonal variations in demands of trade.... There is no evidence of contraction when the crop-moving demands are over, the national bank-note elasticity being (to use a rather inelegant expression) of the chewing-gum variety.

See Kemmerer (1910, p. 153).

ING NOTES TO SATISFY A SHORT-TERM LIQUIDITY NEED. THE REDEMPTION COSTS MAY SWAMP IT???

It is perhaps best to view these complaints concerning currency inelasticity as complaints regarding the terms on which people could obtain credit. We can do this by telling the following story. By law, banks were restricted as to the type of debt that they could intermediate. Only long-term bonds of the government qualified. Commercial paper and other short-term private securities did not qualify. Suppliers of these private securities would like to have banks be able to intermediate their debt into notes, but legal restrictions prevented this from occurring. Banks could intermediate their debt in the form of demand deposits, but perhaps only at a higher cost than through the issuance of notes. Thus, these suppliers of debt borrow at a higher interest rate than would exist if their debt could be intermediated directly into notes. Complaints regarding an inelasticity of the currency arose because credit conditions were more constrained than they would have been if legal restrictions had not prevented the direct intermediation of private short-term securities into notes. This argument is much the same as that presented in Wallace (1983). As Wallace states, if there is a binding prohibition on private note issue, "then some borrowers face higher interest rates on loans than they would if they, directly or through 'banks,' were able to borrow by issuing small-denomination bearer notes. The prohibition puts a barrier between borrowers and lenders. . . ." (Wallace 1983, p. 5)

Due to the risk features of note issue detailed above and the long-term nature of eligible bonds, banks may not have perceived the eligible bonds as perfect substitutes for private short-term securities. So, even though banks were not issuing notes up to the legal maximum, the restriction on the particular class of security which could be intermediated may have been binding. Once it is granted that these two classes of securities, long-term U.S. bonds versus short-term private securities, were not perfect

substitutes, it would not be paradoxical to observe underissuance and complaints over currency inelasticity.

If the National Banking Act had allowed the use of securities such as the modern-day Treasury bill or private short-term securities as backing for notes, it would be difficult to make this argument. In fact, the Aldrich-Vreeland Act of 1908 did allow the use of commercial paper as the basis for the issuance of an "emergency currency" during periods of financial crises. The consequences of this bill's relaxation of the legal restrictions on the type of debt that could be intermediated were tested only once at the outbreak of World War I. However, it appears that the act worked well in adding an elasticity to the supply of notes.

As Friedman and Schwartz state, "To judge by that one episode, the Aldrich-Vreeland Act provided an effective device for solving a threatened interconvertibility crisis without monetary contraction or widespread bank failures." (Friedman and Schwartz 1963, p. 172) This lends support to the view that currency inelasticity was caused by a bindingness of legal restrictions on the type of debt that could be intermediated into notes. Once this restriction was relaxed, bank note issuance responded well during a period of financial crisis.

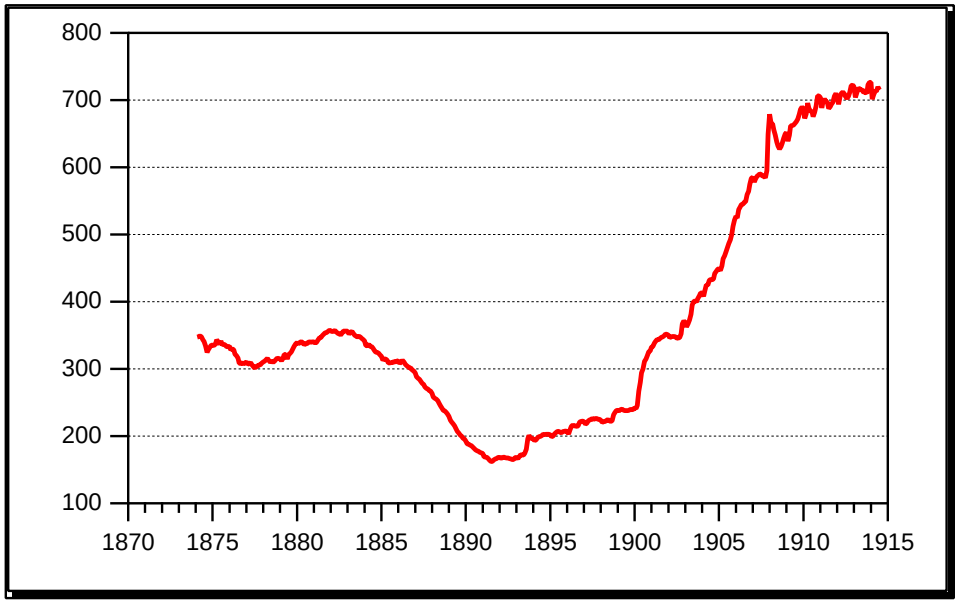
We can also point to certain legal restrictions which interfered with the ability of national banks to adjust the amount of outstanding bank notes. Banks were restricted by law as to the aggregate amount of notes that could be retired in a given month. Further, any bank which reduced its circulation could not subsequently increase its circulation for a period of six months. In addition, Friedman and Schwartz (1963, p. 169), as well as Bell (1912), refer to the existence of significant delays between a bank's decision to issue additional currency and actually receiving the notes from the Comptroller of the Currency. However, given that national banks could costlessly keep a stock of their own bank notes idle in their vaults, delays in obtaining new

notes does not seem to adequately provide and explanation for currency elasticity. A bank did not have to wait for new notes to be printed to put notes into circulation.

5.4.2 Low note issuance during the 1880s

I have noted that throughout most of the national banking era, note issue was below the maximum allowable by law.

Figure 5.6: National Bank Note Circulation, 1874–1914 (millions of dollars)



Sources: Andrew (1910) for data up through 1909. *Annual Report* of the Comptroller of the Currency for 1910–1914.

Champ and Thomson (2006) provide additional insights into the particularly low issuance of the 1880s. We claim that at least part of the decline in note issue during that period was demand-driven.

Bibliography

Andrew, A. Piatt. 1910. *Statistics for the United States: 1867–1909*. Washington: U.S. Government Printing Office.

Bell, Spurgeon. 1912. "Profit on National Banknotes." *American Economic Review*, no. 2:38–60.

Cagan, Phillip. 1963. "The First Fifty Years of the National Banking System." In *Banking and Monetary Studies*, edited by Deane Carson, 15–42. Homewood, IL: Richard D. Irwin.

———. 1965. *Determinants and Effects of Changes in the U.S. Money Stock, 1875-1960*. New York: National Bureau of Economic Research.

Cagan, Phillip, and Anna J. Schwartz. 1991. "The National Bank Note Puzzle Reinterpreted." *Journal of Money, Credit, and Banking* 23 ((August): 293–307.

Calomiris, Charles W., and Joseph R. Mason. 2006. "Resolving the Puzzle of the Underissuance of National Bank Notes."

Champ, Bruce. 1990. "The Underissuance of National Banknotes, 1875-1913." Ph.D. diss., University of Minnesota.

Champ, Bruce, Scott Freeman, and Warren E. Weber. 1999. "Redemption Costs and Interest Rates under the U.S. National Banking System." *Journal of Money, Credit, and Banking* August, part 2:568–589.

Champ, Bruce, and James B. Thomson. 2006. "National Bank Notes and Silver Certificates." Federal Reserve Bank of Cleveland Working Paper (WP06-22).

Champ, Bruce, Neil Wallace, and Warren E. Weber. 1992. "Resolving the National Bank Note Paradox." *Federal Reserve Bank of Minneapolis Quarterly Review* Spring:13–21.

———. 1994. "Interest Rates Under the U.S. National Banking System." *Journal of Monetary Economics* 34 (December): 343–58.

Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton University Press.

Goodhart, C. A. E. 1965. "Profit on National Bank Notes, 1900-1913." *Journal of Political Economy* 73 (5): 516–522 (October).

Hepburn, A. Burton. 1915. *A History of Currency in the United States*. New York: The Macmillan Company.

James, John A. 1976. "The Conundrum of the Low Issue of National Bank Notes." *Journal of Political Economy*.

Kemmerer, Edwin Walter. 1910. *Seasonal Variations in the Relative Demand for Money and Capital in the United States: A Statistical Survey*. Washington: U.S. Government Printing Office.

Kuhlwein, Michael. 1992. "The National Bank Note Controversy Reexamined." *Journal of Money, Credit, and Banking* 24:111–126.

Meltzer, Allan H. 2002. *A History of the Federal Reserve*. Volume 1. University of Chicago Press.

Mitchell, W. C. 1911. "Rates of Interest and the Prices of Investment Securities: 1890-1909." *Journal of Political Economy* 19 (4): 269–308 (April).

Taylor, F. M. 1898. "The Final Report of the Indianapolis Monetary Commission." *Journal of Political Economy* 6 (3): 293–322 (June).

Wallace, Neil. 1983. "A Legal Restrictions Theory of the Demand for 'Money' and the Role of Monetary Policy." *Federal Reserve Bank of Minneapolis Quarterly Review* 7 (Winter): 1–7.

Wallace, Neil, and Tao Zhu. 2007. "Float on a Note." *Journal of Monetary Economics* 54:229–46.