

I. Introduction

During the Civil War, the federal government began to charter national banks. These banks enjoyed the privilege of being licensed to issue national bank notes, which were default-free liabilities of the banks, backed 111% by U.S. Treasury bonds deposited by issuing banks at the U.S. Treasury. The creation of these new banks, combined with a 10% annual tax on state bank note issues, soon resulted in the supplanting of state banks' notes by the new national bank notes.

Scholars have long puzzled over the observation that national banks did not take greater advantage of the authority to issue notes. The supply of notes never reached its maximum permissible level, despite calculations measuring the profitability of allocating capital toward bank note supply collateralized by bonds (e.g., as derived by Cagan 1965) indicating that national bank note issuance was more profitable than the typical profit earned by allocating bank capital toward lending funded by a combination of deposits and capital. Friedman and Schwartz (1963, p. 23) write:

Before 1890 the amount outstanding ranged around 20 per cent of the possible maximum, by 1900 it had risen to about 28 per cent, and by World War I to about 80 percent. The maximum was in fact approached only in the twenties, when for the first time U.S. bonds deposited to secure circulation and government deposits (which also required such security) nearly equaled the total of eligible bonds. Before 1905, the capital stock of national banks set narrower limits to their maximum possible note issue than did the total of eligible bonds, but the actual issue did not approach this lower limit either. Thereafter, the capital stock of national banks exceeded the total of eligible bonds and hence was not the effective limit on note issue. Yet, despite the failure to use fully the possibilities of note issue, the published market prices of government bonds bearing the circulation privilege were apparently always low enough to make note issue profitable except in the years 1884 to 1891. 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) and Cagan (1965) argued that profits from note issue were large on the margin, because bond issues to back note issues remained cheap and because banks could easily leverage their capital devoted to those bond purchases.

James (1978) argued that cross-sectional variation in the regional supply of bank notes was consistent with regional variation in the opportunity cost of note issuance (that is, regional variation in the profitability of bank lending). In James' view, at least some of the puzzle of low bank note issuance was explicable by the high profitability of bank lending in the South and West, where note issuance was relatively low. But James' explanation was not a complete one. After 1874, there were no regional limits on note issuance, suggesting that banks in the East should have substantially increased their outstanding notes.

Hetherington (1990) showed that some of the time variation in the extent of note issue could be explained by changes in rules governing note issues. But that approach did not explain the puzzle posed by Friedman and Schwartz (1963) and Cagan (1965); like James' (1978) explanation of cross-state variation in note issuance, Hetherington's (1990) explanation of some of the variation in supply over time did not address the persistent underissue of bank notes: the level of bank notes remained far below its maximum despite the high profitability of note issuing. Champ, Wallace, and Weber (1992), recognizing that previous empirical analysis has not resolved the puzzle of national bank underissue, posit hidden transacting costs in a desperate attempt to explain why banks did not find it profitable to issue notes.

No such desperate explanations are necessary. In this paper, we resolve the puzzle of note underissue by disaggregating data on national banks and analyzing individual bank note issuing. We take account of the legal constraints facing individual banks that limited their maximum permissible note issuance. Friedman and Schwartz (1963) and Cagan (1965) compared average profitability of note issuing and lending. But that is not the relevant comparison. If, as James argued, some banks face high opportunity costs (leading them to limit note issuing), and if other banks earning lower profits from lending decide to issue the maximum amount of notes that they are permitted to issue by law, then apparent "underissuance" may still occur in the aggregate (in the sense that the average profits from lending may be below the marginal profit of note issuing) even though all banks are maximizing profits. We argue that this, in fact, is the explanation for the apparent underissuance of national bank notes.

We develop a data set linking individual bank note issuing behavior, the specific legal constraints on note issuing faced by each bank, and the profitability of bank lending (as measured by bank performance and other bank characteristics that should have influenced the note supply decision). Our measures of the opportunities for lending include characteristics of the economic environment in which banks operated. Those environmental characteristics are measured using state- and regional-level data.

By disaggregating, and analyzing the constraints faced by individual national banks in 1880, we are able to show that most banks (roughly two-thirds of all national banks) were issuing the maximum amount of notes they were allowed to issue. For the remaining banks that issued less than the maximum amount allowed, we find that cross-sectional variation in banks' opportunity costs go a long way toward explaining the extent of note issuance by these banks. Those same measures of opportunity costs also explain which banks tend to be maximum note issuers. We conclude that a combination of legal restrictions on maximum note issuing and banks' opportunity costs explain the extent of bank note issuing in a manner fully consistent with bank profit maximization.

II. Supply and Demand for National Bank Notes

The quantity of national bank notes in circulation was determined by supply and demand in the market for bank notes. While there were legislative limits placed on the aggregate quantity of notes outstanding and on the geographic distribution of note issuing, those limits were not effective constraints on individual bank issuance. Prior to 1874, whenever the amount of notes came close to reaching the maximum allowable supply, the law was changed to accommodate more note issues. In 1874, the law was changed to remove any limits on note issues, although the limited supply of U.S. Treasury bonds (to serve as 111% collateral for note issues) effectively placed a non-binding upper bound on the potential supply of notes.

Notes were essentially perfect substitutes for transactions purposes with U.S. notes (greenbacks) and coins, and traded at par with those alternative transacting media (except briefly during 1873, when the special value to banks of holding greenbacks, which were a legal reserve currency, led their value to temporarily exceed that of bank notes – see Friedman and Schwartz 1963, pp. 21-2). National bank notes and greenbacks were always inframarginal sources of transacting media set by suppliers and were unresponsive to shifts in the demand for transacting media; increases in demand for transacting media on the margin were met by changes in the supply of specie currency (see Calomiris 1988, 1994, Hetherington 1990).

In essence, equilibrium in international markets under the classical gold standard simultaneously determined gold-denominated interest rates and specie flows to equilibrate the markets for goods and money (Calomiris and Hubbard 1995). The supply of national bank notes adjusted endogenously to the level of interest rates set in the money market. Calomiris (1988, 1994) shows that (so long as the supply of Treasury bonds to back national bank notes was greater than the amount demanded for that purpose) the supply of national bank notes for banks that were not at a corner solution should have been determined by (a) the yield on government bonds, (b) the profitability of bank lending, and (c) the tax rate charged on national bank note issues. The supply of notes, in this model, is set by the profit-maximizing choices of national banks about whether to allocate capital toward (a) the business of deposit taking and lending or (b) the business of producing national bank notes, backed by government bond purchases. In any empirical model of cross-sectional differences in note issuance (for our sample of national banks in 1880) the profitability of lending is the only variable that should predict cross-sectional differences in the propensity to issue notes, since the taxation rate and market yield on bonds are the same for all banks.

Of course, this model does not apply to banks that are at a corner solution, either because of legal limits that require them to issue less than they would like, or legal limits on minimum issues that require them to issue more than they would like. Among banks issuing the maximum permissible amount of notes, cross-sectional variation in bank characteristics will have no explanatory power for note issuing. Thus, before applying the model to individual bank data, one must take account of the various limits on note issuing that might constrain banks to operate at a corner solution.

As of 1880, the limits on note issues relating to capital for national banks were as follows (see National Monetary Commission 1910). Banks chartered before March 1865 could not issue notes in excess of 90% of bank paid in capital. Banks chartered from March 1865 through July 1870 were governed by the following limits on note issue relative to capital: Those with capital less than \$500,000 could issue up to 90% of capital; those with capital of between \$500,000

and \$1,000,000 could issue up to 80% of capital; those with capital of between \$1,000,000 and \$3,000,000 could issue up to 75% of capital; and those with capital in excess of \$3,000,000 could issue up to 65% of capital. Banks chartered after July 12, 1870 were governed by the following limits: No bank could issue more than \$500,000 in notes; banks with capital less than \$500,000 could issue up to 90% of capital; and banks with capital between \$500,000 and \$625,000 could issue up to 80% of capital. National banks that also had outstanding state bank issues (from before they became national banks) had to include those notes in any measure of total permissible note issues. All banks also faced an effective minimum note issuing requirement, since all national banks were required to maintain at least \$30,000 in bond holdings or one third of bank capital, whichever was higher. As we shall show below, these ceilings and floors on permissible note issues were often binding on individual banks.

III. Empirical Analysis

Our dataset consists of information about 1,882 national banks and the states in which they reside, for the year 1880. The dependent variable we analyze is the extent of underissue. Specifically, underissue is defined as:

$$U = [1 - (\text{Actual Notes}/\text{Maximum Permissible Notes})].$$

The median value of U is 0.66% and the mean is 9.78%. Fully 75% of national banks have values of U less than 11%. We assume that banks with U less than 2% are effectively at their maximum amount of note issue (given potential rounding effects from the minimum denomination of bond issues and random variation in outstanding notes associated with redemptions). By that definition, 67% of national banks were issuing the maximum amount of notes permitted. In other words, two-thirds of national banks were at a corner solution in their issuing of national bank notes. Figure 1 plots a histogram of the distribution of U for the other 33% of national banks.

We turn to regression analysis of U . We report three regressions: an OLS regression for the one-third of banks that were not issuing the maximum permissible amount of notes in 1880, a Probit regression for whether banks are maximum issuers or not, and a Tobit regression for the entire sample of national banks, which takes account of left-censored banks (those issuing the maximum amount of notes permitted). In all three regressions we exclude right-censored banks (those issuing the minimum amount of notes permitted). Including these banks would not materially affect our results.

Our measures of bank opportunity costs take account of a variety of potentially relevant factors. We include the return on assets (ROA) of the bank as a measure of bank profitability. Of course, ROA is a noisy indicator of long-run lending opportunities and fails to capture dynamic growth or contraction in expected loan opportunities, which would be relevant to bank decisions about allocating capital between note production and lending. Thus, as additional bank-level proxies, we also include the ratio of loans and discounts relative to assets other than U.S. Treasury bonds in 1880 (LOANRAT), as well as the change in that variable from 1879 to 1880 (LOANGR). We include other characteristics of banks that might also influence their lending and lending growth opportunities, such as asset size (SIZE) and bank age (AGE), defined as years since its national bank charter, and the interaction between the two (SIZExAGE). It occurred to us that a bank's decisions about note issuance might also be influenced by the extent of a bank's connections with the federal government as a fiscal agent of the government (or that those connections might be indicative of the opportunity set of the bank or its overall strategy), so we include the ratio of debt due to the bank from the Treasury relative to total bank assets (DFTTA) as a regressor. We interpret DFTTA as a measure of the extent to which the bank acted as an agent for the government in the bond market.

We also include variables that capture aspects of the economic environment in which banks operate, which might be relevant for measuring opportunity costs, and which might not be fully captured by our noisy bank-level measures of lending opportunities. These include: DIVCAP (dividends paid by national banks in the state / capital of national banks in the state), URBAN (an indicator variable that takes the value 1 if the bank is located in New York, Philadelphia, Boston, Chicago, New Orleans, or San Francisco, and zero otherwise), and six regional indicator variables (MIDATL, MIDWEST, SOUTH, APPAL, WEST,

and the excluded category, EAST). To control for regional differences in links between banks and the federal government, we interact regional variables with DFTTA. Finally, we include the ratio of national banks in the state to total banks in the state (NBRAT) to control for the possibility that the presence of other national banks in a bank's state might influence its note issuing behavior.

Table 1 reports regression results for each of the three regressions (OLS for underissuers, Tobit for all national banks, and Probit for all national banks). The results of the three regressions are broadly consistent and indicate substantial variation in the extent of underissuance that is traceable to bank-specific, state-specific, and region-specific characteristics. Some of the regressors have more explanatory power in some of the regressions than in others.

Interestingly, LOANRAT and LOANGR have opposite signs in all three regressions; LOANRAT is associated with less note issuing and LOANGR with more (after controlling for LOANRAT). One interpretation of this finding is that there are adjustment lags in note issuance and redemption. Banks whose high LOANRAT is due to recent growth in loans may have more notes outstanding than they would like to have in the long run.

Banks with high DFTTA tend to issue more notes, and this varies according to the region of the country. As in James (1978), banks in the West, South, and Midwest tend to issue less notes. ROA, SIZE, and AGE are highly significant in the Probit and Tobit, but not in the OLS.

IV. Conclusion

Most of the puzzle of underissuance of national bank notes disappears when one disaggregates the data to the level of individual banks and takes account of the limits banks faced on their maximum permissible note issues. 67% of banks in our 1880 sample were maximum note issuers.

The hypothesis that bank opportunities in lending can explain cross-sectional differences in bank note issuing is confirmed by our regression analysis. Banks with poor lending opportunities, or those operating in regions with poor lending opportunities, issued more notes or were more likely to be maximum issuers.

We conclude that the puzzle of underissuance of national bank notes is an object lesson in the value of disaggregating data, and thus avoiding "representative bank" analysis relating average bank behavior and opportunities.

References

- Cagan, Phillip (1965). *Determinants and Effects of Changes in the Stock of Money, 1875-1970*, New York: Columbia University Press.
- Calomiris, Charles W. (1988). "Price and Exchange Rate Determination During the Greenback Suspension," *Oxford Economic Papers* (December), pp. 189-220.
- Calomiris, Charles W. (1994). "Greenback Resumption and Silver Risk: The Economics and Politics of Monetary Regime Change in the United States, 1862-1900. In M.D. Bordo and F. Capie, eds., *Monetary Regimes in Transition*, Cambridge: Cambridge University Press, pp. 86-132.
- Calomiris, Charles W., and R. Glenn Hubbard (1995). "International Adjustment Under the Classical Gold Standard: Evidence for the U.S. and Britain," Working paper, Columbia Business School.
- Champ, Bruce, Neil Wallace, Warren E. Weber (1992). "Resolving the National Bank Note Paradox," *Quarterly Review*, Federal Reserve Bank of Minneapolis (Spring), pp. 13-21.
- Friedman, Milton, and Anna J. Schwartz (1963). *A Monetary History of the United States, 1867-1960*, Princeton: Princeton University Press.
- Hetherington, Bruce W. (1990). "Bank Entry and the Low Issue of National Bank Notes: A Reexamination," *Journal of Economic History* (September), pp. 669-76.
- James, John A. (1978). "The Conundrum of the Low Issue of National Bank Notes," *Journal of Political Economy* (April), pp. 359-67.
- National Monetary Commission (1910). *Laws of the United States Concerning Money, Banking, and Loans, 1778-1909*, Washington D.C.: U.S. Government Printing Office.

Table 1: Regressions on National Bank Note Issue, 1880

Model Type	OLS	Tobit	Probit
Dep. Var.	Underissue	Underissue	Maximum Issue=1
N	604	1785	1785
Uncensored Values	604	604	604
R ² / Log-Likelihood	0.3411	-671.16	-1030.27
Adj R ² / Unrestricted Log-Likelihood	0.3185	-876.95	-1180.91
	Coef	Coef	Coef
	<u>Std. Error</u>	<u>Std. Error</u>	<u>Std. Error</u>
Intercept	-0.208	0.657	-3.552
	<i>0.446</i>	<i>0.570</i>	<i>2.133</i>
LOANRAT	0.253	0.217	-0.316
	<i>0.053</i>	<i>0.067</i>	<i>0.248</i>
LOANGR	-0.065	-0.169	0.503
	<i>0.029</i>	<i>0.040</i>	<i>0.150</i>
SIZE	0.026	-0.054	0.293
	<i>0.034</i>	<i>0.044</i>	<i>0.163</i>
AGE	0.027	-0.093	0.481
	<i>0.029</i>	<i>0.037</i>	<i>0.141</i>
SIZExAGE	-0.002	0.007	-0.037
	<i>0.002</i>	<i>0.003</i>	<i>0.011</i>
NBRAT	-0.009	-0.004	-0.025
	<i>0.076</i>	<i>0.098</i>	<i>0.366</i>
DFTTA	-6.531	-10.724	23.500
	<i>1.056</i>	<i>1.672</i>	<i>6.166</i>
MA_DFTTA	-7.015	-6.875	17.813
	<i>2.061</i>	<i>2.839</i>	<i>10.318</i>
MW_DFTTA	-12.305	-18.084	44.026
	<i>1.997</i>	<i>2.959</i>	<i>10.884</i>
SO_DFTTA	-12.204	-32.095	100.045
	<i>9.747</i>	<i>11.736</i>	<i>44.802</i>
AP_DFTTA	-9.073	-19.905	71.429
	<i>5.320</i>	<i>7.081</i>	<i>26.381</i>
WE_DFTTA	-7.012	-26.783	79.941
	<i>17.001</i>	<i>19.478</i>	<i>79.419</i>
ROA	-0.529	6.409	-24.213
	<i>2.168</i>	<i>2.733</i>	<i>10.275</i>
DIVCAP	2.842	-1.171	8.088
	<i>1.667</i>	<i>1.909</i>	<i>7.183</i>
URBAN	-0.048	-0.021	-0.083
	<i>0.028</i>	<i>0.039</i>	<i>0.151</i>
MIDATL	0.033	0.085	-0.340
	<i>0.040</i>	<i>0.058</i>	<i>0.220</i>
MIDWEST	0.161	0.302	-0.855
	<i>0.042</i>	<i>0.062</i>	<i>0.236</i>
SOUTH	0.205	0.448	-1.350
	<i>0.121</i>	<i>0.159</i>	<i>0.627</i>

Fig 1: Histogram of National Bank

