If financial development matters, then how? National banks in the United States 1870–1900

Scott Fulford*

October 2011

Abstract

Despite long-standing interest in the effects of financial development, it has been difficult to determine how banks affect growth since they typically grow with the economy and engage in multiple activities. I consider a time when banks were limited to commercial loans to understand how banks mattered for growth. I construct a novel dataset tracking the size and location of every national bank in the United States from 1870-1900. A large minimum capital requirement meant that otherwise similar counties had very different amounts of banking and I use this discontinuity to estimate the effect of banking on economic development. Even though national banks could not take land as collateral, proximity to a national bank increased agricultural production per capita and tilted the composition of production away from manufacturing. Agricultural gains came from increasing the land under cultivation, not by increasing yields per acre. Additional banking in 1900 still increased incomes 70 years later, suggesting that these results are highly persistent. Although the literature on financial development often focuses on investment as the conduit from finance to growth, this paper points to an alternative: relieving the short-term liquidity needs of commerce.

JEL classification: O16, D91
Keywords: National banks, commercial banking, financial development, growth

*Boston College Department of Economics, 140 Commonwealth Ave, Chestnut Hill, MA 02467; email: scott.fulford@bc.edu. I thank Shahed Kahn for his help entering and checking the national bank accounts, Mashfiqur Kahn for his help cross-checking, Perry Mehrling for his suggestions on the early banking literature, and Don Cox for his rhetorical good taste. This paper previously circulated under the title "Gilded or Gold? National Banks and Development in the United States 1870-1900."

1 Introduction

Do banks matter for growth? If so, how? These questions are difficult to answer since banks typically accompany growth and are involved in multiple activities. That makes it difficult to determine not just whether banks help create growth or simply respond to it, but which functions of banks matter for growth. There is a growing consensus that financial development contributes to growth both internationally and within nations.[1] Yet since banks do many things and may alleviate many different types of constraints, it still unclear how banks help growth. Such questions are increasingly important as some of what used to be the main activities of banks are taken over by new entities in both developed and developing countries.[2]

To answer these questions, I examine a period in United States history when there were strict limits on the activities of banks. From 1870-1900 the US expanded economically and geographically, settling its vast interior. National banks—banks chartered and regulated by the federal government—expanded with the rest of the country and were by far the most important financial institutions in the period. National banks could issue money directly in the form of bank notes. They could not, however, take real estate as collateral, and were limited by the banking practice of the day to make short self-liquidating loans.[3] They were thus commercial banks, facilitating trade

[1]See Levine (2005) for a summary of recent cross country literature, and the question of whether bank based or financial market based funding matters more across countries. Burgess and Pande (2005) and Fulford (2011) examine the experience of India during its large expansion of branch banking in the 1970s and 1980s. In Italy Guiso, Sapienza, and Zingales (2004) find that local financial institutions aid growth, Benfratello, Schiantarelli, and Sembenelli (2008) show they matter for process innovation, but have little impact on product innovation, and Pascali (2011) examines the long term importance of banks. Dehejia and Lleras-Muney (2007) examine state banking regulations and their effects in the US from 1900-1940. Examining a period when some of the strict limitations on banks used in this paper were relaxed, Jayaratne and Strahan (1996) find positive effects from allowing interstate branching. Driscoll (2004), however, finds that in post-war US data, changes in loans do not affect output at the state level.

[2]For example, mobile phone payment systems are creating new ways to move funds from place to place and payment clearing, which used to be a key function of banks using their correspondence network. For an example of one rapidly expanding network see Jack and Suri (2010) on the expansion of M-Pesa in Kenya.

[3]This banking theory was known as the "real-bills" doctrine (James, 1978, p. 59-64). Banks may have occasionally skirted its rules, in particular the self-liquidating requirement, by renewing loans when they became due. Nonetheless, loan maturities were short: James (1978, p. 61) suggests that the average maturity was about 60 days. Sound banking theory, and the value of commercial banking, were clear at the time. See, for example, the eleventh edition of *Practical Banking* (Bolles, 1903, p. 88): "the first and most important function of a bank is, by the use of the capital which it controls, to bridge over the periods of credit which necessarily intervene between production and consumption, in such a manner as to give back to each producer, or middleman, as quickly as possible, the capital invested by him in

through short-term loans and direct money creation, not investment banks. If these banks mattered for growth during the period, particularly in the agricultural sector, it was because of their commercial, not investment, activities.

To analyze the effects of these banks, I create a rich new data set which gives the exact geographic coordinates and size of every national bank in 1870, 1880, 1890, and 1900. Charged with regulating the national banks and the money they issued, the Comptroller of the Currency collected and published the balance sheet of every national bank each year. Since national banks were not allowed to branch, the place of business listed in the accounts allows me to locate each bank precisely—and to examine local financial development with greater nuance than studies that are limited to regional aggregates. Every decade the census collected detailed data on manufacturing and agricultural output in each county, as well as the amount of land under cultivation and the size of farms. Combining the census data with the location of the banks gives me a detailed panel of banking and output in every county of the US over three decades.

Simply comparing areas that had banks with those that did not does not identify the effect of banking, since counties where banks want to enter are likely to also be areas of high economic activity. To identify the effects of the national banks, I use a combination of regression discontinuity (Lee and Lemieux, 2010) to uncover causal effects and multiple imputation to incorporate underlying uncertainty (Rubin, 1987). Concerned with the stability of the money supply, Congress required national banks to have a large minimum capital. The large minimum size meant that banks were limited in where they could enter profitably: not every county could support a bank of the minimum size, and many banks opened at exactly the minimum size. Some counties had significantly more banking than they would have received if banks were allowed to open at their optimum size; others had much less. How much more banking? How much less? Where and when banks choose to enter from decade to decade and banks' behavior after the minimum capital requirements

such products, in order that he may use it again in new production or new purchases." Since the national banks funded mostly short term loans, it was accepted that they could not meet the investments needs of agriculture, even if they could meet its commercial needs (Wright, 1922, pp. 46, 70).

were reduced in 1900 provide information about the distribution of which counties were close to the line between a bank opening and not. I draw from this distribution and use multiple imputation so that the estimates reflect the underlying uncertainty of how close banks were to entering. The combination of multiple imputation and regression discontinuity in a panel setup, while a straightforward use of each method individually, provides a flexible way of using a discontinuity when the assignment is observable but the assignment variable is not.

Despite the lending limitations, proximity to a bank increased output per capita, largely by increasing agricultural production. For the marginal county, gaining a bank of the minimum size increased total production per capita by 13 percent, and agricultural production by 14 percent. Banks were not just following growth, but helping to create it. The mix of production in counties with banks shifted towards agriculture, despite the rapid rise in manufacturing over the period and the limitations on the kind of loans banks could make. Access to a bank increased agricultural production because farmers brought new fields into production, not by increasing output on current fields. The pattern suggests that banks helped agriculture expand on the extensive margin by providing working capital and the liquidity necessary to bring products to market, not by providing investment capital to improve yields. Whether it was from money creation, working capital to farmers, or credit to merchants, it is clear that the commercial activities of banks contributed to growth, even though these banks did not fund capital investments. Moreover, the counties with more banking capital had higher incomes even 70-100 years later which suggests that the effects of the national banks are extremely persistent.

While the focus of much of financial development theory has been on how financial institutions fund new investments, the commercial activities of banks or other financial institutions, particularly in developing countries, may be equally or even more important.[4] In 1870, the GDP per capita

[4] Aghion and Bolton (1997) present one version of the constrained entrepreneur. The entrepreneur might be making a human capital investment as in Banerjee and Newman (1993) or in Galor and Zeira (1993) the entrepreneur is someone deciding on an occupational choice. Galor and Moav (2004) provide an even more nuanced growth story, with similar underlying choices. In Greenwood and Jovanovic (1990) and Townsend and Ueda (2006) the entrepreneur faces a risky high return, or a safe low return investment, and financial markets bring diversification. Banerjee (2001) presents a model that nests several versions of credit constraints for an entrepreneur.

of the United States would have put it someplace between India and China today (Maddison, 1995, p. 196), and as Updike (1985) argues, the United States shares characteristics with many developing countries today. In particular, the poor transportation infrastructure made getting goods to market costly and time consuming, particularly from rural areas, which remains a problem for many developing countries today (Bank, 2009). Recent empirical work has also suggested the importance of working capital and the availability of liquidity for small enterprises in the urban areas of developing countries (Banerjee et al., 2009; de Mel, McKenzie, and Woodruff, 2008). While banks may engage in many activities, this paper suggests that facilitating commerce by relieving short-term liquidity needs, whether to the producer or merchant, is a key avenue for financial development to affect growth.

2 Banking and financial markets 1864–1914

> A national bank affords a safe place for the deposit of all the little hoards and savings which otherwise would be unemployed. It aggregates these into a fund which becomes useful and powerful in stimulating trade and enterprise.
> —Hiland R. Hulbard, Comptroller of the Currency, 1871[5]

This section briefly discusses the financial markets and the economy of the United States from 1864 to 1914, and the literature examining them. While the national banking system was an important factor in early discussions of financial development, more recent literature has largely focused on the system's role in integrating financial markets. Despite the importance of the national banking era in the development of the American economy, this paper is the first to estimate the economic effects of national banks.

The national banking system largely replaced the state-chartered banking system that preceded it. Before the Civil War (1861-1865), there was no national system of banking. States chartered and regulated, or chose not to regulate, their own banks. These state banks issued their own

[5] Report of the Comptroller of the Currency to the Second Session of the Forty-Second Congress of the United States, 4 December 1871, p. XIII.

banknotes—bank-issued currency, backed only by the issuing bank's willingness and ability to pay—which circulated widely, and there was no central clearing system, although regional associations of banks created various clearing arrangements (Bodenhorn, 2000). The Civil War (1861–1865) allowed the Republicans in the US Congress, who no longer faced opposition from southern legislators, to move forward in creating a new banking and currency system. The National Currency Act of 1863 and the National Banking Act of 1864 allowed the newly created Comptroller of the Currency to charter national banks, which could issue national bank notes backed by US treasury bills—in effect allowing these banks to issue and back US currency. State banks were slow to convert to national banks, and so in 1865, Congress passed a new act which established a 10% tax on state banknotes. Not surprisingly, over the next year almost all state banks converted to national banks (White, 1983).

The goal of the National Banking Acts was to create a uniform bank note currency that would trade at par and to help raise funds for the Federal (Northern) war effort. To help ensure the stability of the new monetary and banking system, the acts imposed several restrictions on the new banks. The acts placed minimum capital requirements to form a national bank: a national bank needed at least $50,000 in capital to form in a town with no more than 6,000 inhabitants, at least $100,000 in cities between 6,000 and 50,000, and at least $200,000 in larger cities (U.S. Congress, 1864, sec. 7).[6] Moreover, the acts prohibited direct mortgage lending by national banks, and banks could not hold any mortgages obtained indirectly for more than five years (U.S. Congress, 1864, sec. 28).[7] National banks, and most state banks at the time, were not allowed to set up branches, and so all

[6]The evidence does not suggest that the limits above $50,000 were strictly enforced. Between 1870 and 1880, counties with cities with populations between four and six thousand gain somewhat more banks per capita, but the difference is statistically insignificant, despite the doubling of required capital as population increased over six thousand. The capital size appears to have been reasonably easy to circumvent by opening in a nearby town. The city populations are from the census in 1880 which also reports 1870 populations (Census Office, 1880, pp. 416-425). I matched these cities to locations using the same process as with the banks as in appendix D, giving the geographic location of each city, and the 1890 county it falls in. While it seems possible to use the change in capital requirements around 6,000 as a discontinuity to estimate the effect of banking, it does not have good explanatory power.

[7]Keehn and Smiley (1977) suggests that the ban of mortgage lending was not perfect, but was nonetheless extremely restrictive. Loans secured by mortgages or real estate were less than two percent of total value before restrictions were relaxed in 1914, and rose afterwards.

banks were unit banks, facing the same constraints.[8]

While the larger limits do not appear to have binding, the minimum capital of $50,000 limited banks' entry into many areas. The laws allowed allowed banks to open before they were fully capitalized, but required them to quickly become fully capitalized. In 1870, 1880, and 1890, every bank that reported less than $50,000 had attained at least the minimum capital by the next year or shut down. As suggested by Sylla (1969), the best evidence that the minimum capital requirement was binding even at the end of the period is what happened when it was loosened. The Gold Standard Act of 1900 reduced the minimum capital requirement to $25,000 for towns under 3,000, and over the next decade thousands of new national banks were formed with capital below $50,000.

Immediately after 1865, the number of national banks grew quickly as state banks converted into national banks. Growth in the total number of national banks then slowed, before accelerating in the 1880s during a boom in banking. Figure 1 shows the growth of national and state banks. Until the late 1880s, national banks had few and ineffective competitors. In the late 1880s as deposit banking became more important and states allowed banks to form without a special charter, the number of state banks increased rapidly, filling an apparent void left by the national banking system.[9] Yet there still seems to have been a strong desire for national banks, as the surge of smaller banks after 1900 suggests. Some of these new national banks may have been former state banks, which did not find it profitable to open with the full $50,000 in capital as national banks, but did with a smaller required capital.

The total number of banks hides a more complex process of entry and exit and spread of geographic extent. Even as the number of banks was growing, there was substantial exit from decade to decade. While most of the converted state banks were in the Northeast, new banks

[8]The second comptroller of the currency interpreted the National Banking Act as prohibiting forming branches. Although the exact language of the act does not necessarily prohibit forming branches, the rules stayed in effect until the 1920s (White, 1983, pp. 14-15). State banks were similarly constrained by state laws. For the positive effects of relaxing these laws a century later, see Jayaratne and Strahan (1996).

[9]See James (1978, pp. 29–39) and Barnett (1911, pp. 11–12, 32-33) for a discussion of the spread of state banks. After around 1890, many states had less stringent capital requirements than the National Banking Act (James, 1976c), which allowed state banks to open more easily.

spread west as the Midwest and Western states and territories became increasingly populated and productive. Figures 2 through 5 show this spread over time and space. Despite the substantial growth in national banks, many counties did not have a national bank even by 1900. In particular, the South, whose banking system had been largely destroyed during the Civil War, and the sparsely populated West, lacked banks.

Deflation and panics were important aspects of the financial markets during the Gilded Age. The United States faced a prolonged period of deflation following the war as it first resumed the gold standard at pre-war prices and then maintained it, not withstanding periods of bi-metallism (Friedman and Schwartz, 1963). Notes issued by different national banks, backed as they were by treasury deposits and uniform regulations, traded at par with each other and with currency issued directly by the government. Although aggregate notes issues were initially capped, the cap was removed in 1875 as part of the Resumption act. National banks generally chose to limit their note issues below the maximum, however, which contributed to the relative scarcity of money (James, 1976b). There were major banking crises in 1873 and 1893, and smaller disturbances in 1884 and 1890 (Wicker, 2000). At the end of the period, the crisis of 1907 prompted a reform of the system and the creation of the Federal Reserve. National banks held reserves and interbank deposits in regional reserve city banks, which in turn held reserves in New York (and to a lesser extent Chicago), which made the entire system sensitive to disturbances in New York (Cagan, 1963, pp. 36–37). The 1873 and 1893 crises seem to have accompanied a cyclical downturn (Wicker, 2000, pp. 8–11).

The national banking system played an important role in the early discussion of the importance of money, banking, and credit.[10] One area of particular concern was how poorly the national banks

[10]For example, the *Journal of Political Economy* published a four part series in 1918 , and a comment and reply in 1919, on commercial banking and capital formation (Moulton, 1918a,b,c,d; Watkins and Moulton, 1919) whose primary source of information about what banks actually do comes from national banks and their regulator the Office of the Comptroller of the Currency. The new *Review of Economics and Statistics* published a four part series from 1924 to 1927 solely on national bank statistics (Young, 1924, 1925a,b, 1927). One of the most successful textbooks on banking (Dunbar, 1892) devotes as much attention in its first edition to the national banks of the US as to the banking systems of France, England, Germany, and the Bank of Amsterdam, despite the novelty of the national banking system. The 1917 editions drops the Bank of Amsterdam in favor of discussing the new Federal Reserve.

seemed to serve agriculture. For example, Wright (1922) argues that the large minimum size meant that national banks could not profitably enter many rural areas; the prohibition from taking real estate as collateral limited the ability of farmers to borrow; and the requirement to lend only on a short-term basis meant national banks could not fund long-term agricultural investments.

More recent work has focused on the supposed instabilities of the national banking system, which led to the creation of the Federal Reserve in 1913, and its role in the integration of capital markets. This paper is the first to examine empirically whether and how the national banks mattered in the economic growth of the United States. In one of the few papers that considers non-financial effects, Campen and Mayhew (1988) describe the importance of the national banks in Knoxville, Tennessee. Much of the literature on national banks and monetary matters after the Civil War focuses on explaining regional variations in interest rates. Davis (1965) documents that national banks in the Mid-Atlantic region charged a lower average discount rate, as well as had lower returns, than banks in other regions. These gaps seem to have narrowed sometime before 1900, which Davis (1965) attributes to the development of a national commercial paper market which allowed capital to move more easily across regions. While capital flows may have increased from the more developed East, Sylla (1969) suggests that where national banks did exist in rural areas, they could act as monopolists since the minimum capital requirements and branching restrictions made it difficult to acquire sufficient funds to enter. Moreover, in many rural areas the available capital for deposit in a bank was not sufficient to make it profitable for one bank to enter and put up the minimum capital, much less a second one to offer competition. Suggesting that monopoly power may have been important, James (1976a) finds that the number of banks (including state, national, and private banks) per capita at the state level is negatively related to the interest rate between 1893 and 1901, taking into account the risk as measured by the variance of the loss rate. Binder and Brown (1991) provide somewhat more formal tests of the convergence of returns based on the timing of institutional changes, and suggesting that the timing of the 1900 relaxation of national bank capital requirements does not seem to have been important. Sullivan (2009) suggests that profits are a better measure of possible monopoly power than interest rates or returns, and finds that differences in regional profits fell after 1884.

3 Identification strategy

There are two difficult problems to identifying the effects of banking: endogeneity and dynamic effects. Suppose banks can enter freely and we observe more banking activity in wealthier areas. Should we conclude that banking causes wealth, or wealth attracts banking? Most likely the answer is some of both, but we might still like to know the effect of encouraging or discouraging banking, particularly for marginal areas likely to be affected by the policy. Does forcing, or subsidizing, banks to enter areas they might not otherwise want to enter increase productive activity, and by how much? India, for example, for years maintained a "social banking" policy which forced banks to open branches in rural areas with the express intent of fostering additional access to credit in rural areas and so promote growth (Panagariya, 2008, pp. 224-8). Such thinking is also behind the recent push for subsidized microcredit: the profits are not sufficient to bring in profit maximizing firms, but the benefits to credit are assumed to be large.

I use the observation that the minimum capital requirement forced some areas to have too much banking while restricting others to separate the endogenous effects of banking from the direct effects. To see how, suppose x_{ct} determine how profitable of a county c at time t is for banking. So x_{ct} might include population, how good the farm land is, transportation, weather, the cultural, religious, or institutional composition of the county. While some parts of x_{ct} may be observable, it is not possible to observe all of the important elements; one county may be better to open a bank because the brother of a major stockholder lives there. The amount of banking capital in a county is thus some function of this unobservable variable: $C_{ct} = C(x_{ct})$. Other important outcomes, such as manufacturing or crop production, are also determined by x_{ct} or its components: $Y_{ct} = Y(x_{ct})$.

Then if we wanted to know how much an increase in banking capital in a county affected economic output, a regression of the form:

$$Y_{ct} = \gamma_c + \gamma_t + \gamma C_{ct} + \epsilon_{ct}^Y$$

where γ_c and γ_t are county and time fixed effects, would not be informative. Since both Y_{ct}, and C_{ct} are positively related to the unobserved x_{ct}, the estimate of γ will tend to be positive, even if banking capital has no effect, or even a negative effect on output. Banks may allow farmers to drink away next year's crop, and so not be able to work hard the next day. In that case, banks actually reduce output, but there will still be more banking capital in areas where farmers have larger crops to drink away, so the estimate of γ will still be positive. The problem may become even more acute if current banking responds to predictions of future economic activity. The central empirical difficulty of how finance affects development is how to estimate the effect of banking on economic activity γ without bias.[11]

Suppose we could assign some places more banking (or less banking) than those counties would get based on their level of the unobserved x_{ct}. The extra banking can then identify γ, as long as the extra banking is not related to the unexplained economic activity ϵ_{ct}^Y. The minimum capital requirement that banks could only start with a minimum of $50,000 meant that some counties received much more capital than they would have gotten without the requirements since banks, which would have entered with a profit maximizing capital less than $50,000, instead entered with the minimum capital. Some counties, on the other hand, were denied banking that they would have had since banks did not find it profitable to enter with such a large capital stock. The minimum capital requirement thus causes the capital stock to jump discontinuously from $0 to 50,000, even though we might expect that underlying economic activity which attracts banks behaves continuously. Small changes in the underlying economic activity x_{ct} thus cause big changes in the amount of banking.

[11] See Levine (2005) and King and Levine (1993) for some attempts to deal with this problem comparing countries. Rajan and Zingales (1998) examine industries which are more likely to benefit from finance. Burgess and Pande (2005) use social banking rules in India as an instrument for which districts received more banks.

Discontinuous functions of endogenous variables can be used to estimate causal effects as long as the other factors which might explain Y_{ct} behave continuously around the transition between entry and not (Lee and Lemieux, 2010). Imagine a single county in the Midwest in 1870, with fertile land, but still sparsely populated. In 1870, despite its high potential, there were not enough people to justify a bank with $50,000 in capital. Optimal capital for a bank might have been $C^*_{c1870} = \$22,000$. Over the next decade, more people move in, and now the profit maximizing capital is $C^*_{c1880} = \$26,000$ and a bank enters. Even though optimal, endogenously determined, capital has increased by only $4,000, actual capital increased by $50,000! The county has been treated with a lot of extra banking, even though the endogenous conditions which would call forth more banking have changed very little. As long as nothing else behaves discontinuously at the transition point, then any changes in Y_{ct} come from crossing the threshold, and so can be attributed to the increase in banking. Similarly, a county nearby which had an optimal capital of $27,000 in 1870 had a bank and a lot more capital, even though the two counties were similar.

A bank makes profits on the spread between its cost of capital and the return on its loans, so entering with too much capital harms profits by requiring the bank to pay for capital it can employ less profitably. Take a simple version of the profit function of a bank considering opening in some county which treats all forms of financing as equivalent. The bank must borrow its capital at some rate $r^B_{ct}(C)$ which increases as it needs more capital, and then turns around and loans its capital at some rate $r^L_{ct}(C)$, which decreases at it attempts to loan more. Both the cost of capital, and demand for loans depend on the county. For example, it may be expensive to get capital in the Nevada desert, and there may be little demand for funds there. Profits are then:

$$\pi_{ct}(C) = r^L_{ct}(C)C - r^B_{ct}(C)C,$$

and a bank chooses its optimal capital C^*_{ct} to maximize profits. But the bank must open with at least $50,000 in capital, and so faces a constrained maximization problem. Consider how what a bank actually does (C_{ct}) varies depending on what it would like to do if it were unconstrained

(C^*_{ct}). Both are functions of x_{ct}, but C_{ct} is a discontinuous function. Figure 6 show four different profit functions which might hold in four different counties or for a county over time. In panel (A), profits for that county are less than 0 if a bank enters with \$50,000 in capital, and so it does not enter and $C_{ct} = 0$ even though C^*_{ct} is positive. In panel (B), the bank is just indifferent between entering and not, since profits are zero either way. In panel (C), the bank makes positive profits if it opens at \$50,000 capital, but would prefer to enter with less, and so opens with exactly \$50,000 in capital. In panel (D) the bank is unconstrained and opens with its optimal capital.

When forced to open with the minimum capital, the closer the minimum capital is to its optimal capital, the higher the banks profits. As the optimum capital gets further and further from \$50,000 profits decline, until the bank chooses not to enter. Let C^T be the optimal capital of the marginal county, just on the border line between getting a bank and not, which is shown in panel (B). For $C^*_{ct} > C^T$, entry occurs, and not otherwise, and so C^T is the transition capital between entry and not. In appendix A, I develop a simple profit maximization model with the bank acting as a local monopolist facing both a downward sloping demand for loans and an upward sloping cost curve for capital, which shows how the difference between desired capital stock and the required minimum capital stock affects profits. The model, based on some linearity assumptions, suggests that $C^T = \$25,000$ is a reasonable threshold, but I let the data tell me the best threshold.

Having to open with a minimum capital introduces a discontinuity. Counties with an optimal capital just above C^T have a lot more capital than counties with an optimal capital just below C^T. So the actual and observed capital in thousands in a county is a discontinuous function of C^*_{ct}:

$$C_{ct} = 50D(C^*_{ct} > C^T) + (C^*_{ct} - 50)D(C^*_{ct} > 50),$$

where $D(\cdot)$ is an indicator which is 1 if the condition is true and 0 otherwise. Figure 7 shows this discontinuous function. For very low optimal capital, profits are negative and so no entry occurs. As optimal capital passes C^T, the actual capital jumps to 50, and stays there until the bank is no longer constrained. Alternatively, define $EC_{ct} = C_{ct} - C^*_{ct}$ as the extra capital (in thousands) that a county has, or the loss of capital from not having bank. It too is a discontinuous function of C^*_{ct} which jumps by 50 at C^T.

I use the minimum capital requirement to divide the observed banking into the endogenous banking activity and the extra banking that comes from meeting the minimum capital requirement or the loss of capital and banking activity from not having a bank. Then the endogenous economic outcome equation is:

$$Y_{ct} = \gamma_c + \gamma_t + \gamma EC_{ct} + \gamma_e C^*_{ct} + \epsilon^Y_{ct}. \tag{1}$$

If $\gamma = \gamma_e$ then this equation is just the endogenous economic activity equation 3, since by definition $EC_{ct} + C^*_{ct} = C_{ct}$. In a window around C^T, however, it is possible to estimate γ using the discontinuity in EC_{ct}, even though C^*_{ct} is potentially endogenous, and EC_{ct} is a function of C^*_{ct}.[12]

Of course, I can only observe the outcome C_{ct} not the underlying C^*_{ct}, but I have a great deal of information about the optimal capital. Counties with unconstrained banks, counties with constrained banks, and counties without any banks, all give useful information on how much capital a county would have in the absence of minimum capital requirements, and I observe each county over time. So a county without a bank in one period, but with one in the next must have passed over the threshold of profitability. Suppose that the optimum capital in a county is driven by the county population P_{ct} and local unobservable business conditions η_c and local temporal shocks ϵ_{ct}. The optimal capital is:

$$C^*(x_{ct}) = C^*_{ct} = \alpha\alpha_t P_{ct}\eta_c\epsilon_{ct} \tag{2}$$

where $\alpha_{1870} \equiv 1$. Appendix A shows how such a reduced form equation would come from profit maximization with the demand for loans and supply of capital varying by county. Given the thresh-

[12]To put the discussion in somewhat more precise regression discontinuity terms, consider both C_{ct} and C^*_{ct} as functions of the assignment variable x_{ct}. Then as x_{ct} varies, C^*_{ct} behaves continuously by assumption and C_{ct} is discontinuous, producing a sharp regression discontinuity. C^*_{ct} is a continuous function and so replaces the polynomial function of x_{ct} often used as an approximation of the correct function in sharp RD (for examples see Angrist and Pischke (2009)). If there is a random component to x_{ct} (such as rainfall), then for $C^*(x_{ct})$ close to C^T, entry occurs at random, and the approach is a fuzzy regression discontinuity design, or locally an instrumental variable (Lee and Lemieux, 2010). Note that while C^*_{ct} is technically a choice variable, the (unobserved) assignment variable x^{ct} is not, so that there is not precise manipulation of assignment (even if bankers have some influence over x^{ct} they have not yet been able to control the weather).

old for positive profits, and the assumption that $\ln \epsilon_{ct} \sim N(0, \sigma^2)$ and $\ln \eta_c \sim N(0, \sigma_\eta^2)$, I estimate $\hat{\alpha}$, $\hat{\alpha_t}$, $\hat{\sigma}$, and $\hat{\sigma}_\eta$ using maximum likelihood. Counties fall in one of three categories, depending on whether there are no banks in the county, any banks with the minimum capital, or only unconstrained banks. Each county may fall into all three categories over the full panel. Maximizing the likelihood is made more difficult by the county level dependence of each time observation on η_c which requires using numerical integration to get the log-likelihood for each county conditional only on the data. Appendix B constructs the likelihood function and discusses estimation. I choose the threshold value $\hat{C}^T$ by a grid search which maximizes the likelihood.

For a given county, the underlying unobservables must be consistent with the bank entering with a given capital. The estimates $\hat{\alpha}$, $\hat{\alpha_t}$, $\hat{\sigma}$, and $\hat{\sigma_\eta}$ give the distribution of the ϵ_{ct} and η_c, conditional on observing the actual capital. Any single draw from that distribution ignores the uncertainty that that particular value is correct. For example, it is possible that a county without a bank in all four periods simply had very low ϵ_{ct} every period, but was actually an excellent place to do business (had a high η_c and population). Instead, I draw many possible realizations of optimal capital and use multiple imputation when I estimate the effects of banking on economic activity to combine them. Multiple imputation corrects the coefficients and their standard errors for the uncertainty in any given draw from the optimal capital by estimating as if each draw is the truth, then combining the estimates to account for the added uncertainty from the change in draws (Rubin, 1987; Schafer, 1997). The distribution of $\ln \eta_c$ and $\ln \epsilon_{ct}$ for each county, conditional on the observed banking, is a multivariate truncated normal, and I draw from it using a Gibbs sampler switching between drawing $\ln \eta_c$ conditional on each of the $\ln \epsilon_{ct}$ and all of the $\ln \epsilon_{ct}$ conditional on $\ln \eta_c$. Full details are in appendix C.

The second major difficulty in estimating the effect of finance or banking is that the effects may vary over time. Credit, by definition, allows some people or firms to bring forward investment or consumption, while others to delay it. Relieving credit constraints, or introducing a new savings option, is likely to have effects that vary over time, possibly dramatically, with what holds in the long term the opposite of what holds in the short term. Usefully, the problem is reasonably easy to deal with: Fulford (2011) shows that including past values of the banking variable deals with the problem by allowing the effects to vary over time. Whether or not lags matter depends on what banks do, and how people use banks, and so the structure of the lags reveals a great deal about the effect of banking.

To summarize, the full estimation strategy takes the following steps:

1. Estimate $C^*_{ct} = \alpha\alpha_t P_{ct}\eta_c\epsilon_{ct}$ using panel MLE, assuming that banks will not enter if $C^*_{ct} \leq C^T$, are constrained if $C_{ct} = 50$ and so $C^T < C^*_{ct} \leq 50$ and are unconstrained if $C_{ct} > 50$, and that the unobserved factors that affect the optimal capital are distributed by $\ln\epsilon_{ct} \sim N(0, \sigma^2)$ and $\ln\eta_c \sim N(0, \sigma^2_\eta)$. Find the C^T that maximizes the log likelihood.

2. Using the estimates $\hat{C^T}$, $\hat{\alpha}$, $\hat{\alpha_t}$, $\hat{\sigma}$, and $\hat{\sigma_\eta}$, draw M values for each constrained county of C^*_{ct} by making M draws from the joint distribution of $\ln\epsilon_{ct}$ and $\ln\eta_c$ conditional on observing the actual capital and population. Each draw is made after N iterations in the Gibbs sampler, where N is a large number.

3. For each of M imputations, estimate variations of equation 1, including lags of the extra capital stock, as if each draw of EC^m_{ct} and C^{*m}_{ct} was the actual value, then combine the estimates using standard multiple imputation. For each imputation, select only the counties with at least one observation of C^{*m}_{ct} within a window h around $\hat{C^T}$, which allows the number of included counties to vary over imputations.

3.1 Using 1902 capital to gain precision

An additional policy change allows me to gain precision in the multiple imputation. In 1900 capital requirements for national banks were halved to $25,000. Many new banks decided to enter in the next several years at the lower capital requirements, and some existing banks reduced their capital.[13] Where banks decided to enter, and which banks chose to reduce their capital stock is very informative about the optimal capital in a county. For example, a county which gets a bank after 1900 of between $25,000 and $50,000 reveals exactly what the optimal capital for that county is. While the reduced capital stock cannot be used to estimate the effects of banking directly—it was not in place and so could not have had an effect on growth before 1900—it does help tighten estimates of the optimal capital. It may have taken several years for new banks to enter at the reduced capital requirement, and so I use the capital of banks in 1902. Under the assumption that the observed capital in 1902 is what banks would have wanted to do in 1900 if the lower capital requirements had been in place, I can use the 1902 capital to tighten the estimates. To do so, I estimate the optimal capital equation allowing for two different capital thresholds: C^T in place for 1870, 1880, and 1890 is the transition capital for entry at $50,000, and C^T_{25} is the transition capital for entry in 1902 at $25,000. When I draw multiple imputation estimates the unobserved business conditions have to be compatible with the observed behavior in 1870, 1880, 1890 and 1902. But when calculating excess capital, I use the actual capital in place in 1900. Using 1902 thus allows tighter estimates of excess capital. Using 1902 changes which counties are more likely to be close to entry, and so selects a different sample of counties, and so I show estimates both using and not using the capital in 1902.

4 Data

4.1 Sources and construction

This section gives a brief description of the sources and construction of the data, and some descriptive statistics. Appendix D gives additional details. I have created a new data set at the bank level of all national banks in 1870, 1880, 1890, and 1900. The bank level data comes from the

[13] Due to the application process to the Comptroller, and the requirement to raise capital, relatively few had opened between the passing of the Gold Standard Act and the Comptroller's report. Whenever I use 1900 information, the few banks which opened below $50,000 are excluded since they could not have been open more than a few months and so their effect on growth between 1890 and 1900 would be small.

national bank accounts collected by the Comptroller of the Currency (who is still the official regulator of national banks), and reported to Congress each year. The accounts of each bank report the town or city in which it was located—which since branching was not allowed was its only place of business. I match these towns and cities with the place names from the Graphical Names Information System maintained by the US Board of Geographic Names, which gives gives a latitude and longitude. Figures 2 through 5 show the results of this placement as banks spread over time and across space. With their location, I can match the banks to the counties (defined by 1890 county boundaries).

The census collects a wide range of demographic and economic information every ten years and reports the aggregates for counties, which are sub-unit of states, almost always with their own local governments. In keeping with the importance of agriculture during the period, the census collected detailed information on farm production, yields, and farm size, as well as some manufacturing production. Haines and ICPSR (2010) collected and entered this information, as well as the aggregate at the state level, and I use the National Historical Graphical Information System (Minnesota Population Center, 2004) to match the the census data with counties and the location of banks.

4.2 Sample selection

Throughout the analysis, I exclude all counties that had an urban population of 50,000 or more in 1880, which excludes counties with major cities. Counties with with large urban population were not generally constrained by the minimum capital requirement, so the restriction mainly affects the estimates of optimal capital. Since the activities of banks in counties with large urban populations were different than banks in rural counties, it makes sense to exclude them from the analysis which focuses on the identifiable effects of additional banking in marginal counties.

For most of the analysis, I include all rural counties. The Southern banking system was largely destroyed after the war, however, and there were far fewer banks in the South over the entire period.

National banks could have had a very different effect in the South, and given the relative poverty of the Southern states, including them might affect the results. Similarly, although states were carved out of the Western territories after the Civil War, they were sparsely populated, and not consistently divided into counties, and so the Census does not provide useful data on some of them until later in the period. For comparison, I also examine results using a restricted sample of of the Union (Northern) and border states during the Civil War, which are the Northeastern, Atlantic, and Midwestern states.[14]

4.3 Descriptive statistics

Table 1 shows descriptive statistics for banking and the economic variables over counties and for each decade. The table shows statistics for both all rural counties, and the rural counties in the Union states. Combining farm and manufacturing production from the census, I create a measure of total production per capita. While it does not include services, and so does not directly measure aggregate output, services would have been a small portion of the economy at the time, and the analysis includes fixed effects or first differences, and so the constant absorbs services to the extent to which services are proportional to the rest of economic activity. I do not adjust for deflation or inflation in the table, but instead allow for time effects throughout the analysis. Deflation occurred over much of the period, particularly around the Resumption Act of 1875, and so values in nominal dollars tend to understate growth. The average county population grows substantially over the time period. Following a dip during the 1870s, so does manufacturing and farming production per capita, and in real prices both are likely growing very quickly. Manufacturing production tends to be fairly concentrated, even when I exclude counties with an urban population of more than 50,000 in 1880. So although total manufacturing production is more than half of aggregate production in

[14] I also exclude California, Oregon, and Nevada, which were states before 1870, but might not be comparable since they have incomplete coverage by the census in early decades and are far from the rest of the states. The full list of states in the restricted sample is: Connecticut, Delaware, Illinois, Indiana, Iowa, Kansas, Kentucky, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, New Hampshire, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, Vermont, West Virginia, and Wisconsin.

some years, in an unweighted average over counties it represents less than half of total production since many counties have little manufacturing. The average number of banks increases over time as well, particularly during the 1880s, when the banks per capita more than doubles. Moreover, banks tended to fill in gaps in geographic coverage—the average distance to a bank declines substantially from 1870 to 1890.

5 Results

5.1 Optimal capital

The first step in estimating the effects of banking is to estimate the optimal capital equation 2. The full estimation details are in appendix B. Column 1 in table 2 shows the estimates for the all counties, while column 2 shows the estimates using the capital stock for 1902 after the minimum capital requirement was halved. Columns 3 and 4 restrict the sample to rural counties in border or Union states. There is a great deal of fixed heterogeneity across counties (σ_η), and a much smaller individual county decade heterogeneity (σ_ϵ). The importance of the fixed component suggests that how good a place is for banking is largely fixed, and so entry depends particularly on whether a county has a sufficient population to sustain a bank of at least the minimum size, rather than on whether a county has an idiosyncratic shock. These results reflect the experience of the period of population growth accompanied by entry, as shown in figures 2 - 5, with relatively few counties losing banking (although individual banks might exit). Banks were not willing to enter with \$50,000 unless the population and business climate were enough to make their optimal capital \$19,454 ($\hat{C^T}$), which suggests that entry with the minimum capital was profitable even when the optimal capital was low. Since banks could place excess capital in reserve city banks, which paid interest on it, the cost of carrying the extra capital may not have been the full cost of capital, but instead the presumably lower difference between the cost of capital and the interest paid on reserves.

With the estimates using the capital in 1902, the entry increases to \$24,359, and banks were willing to enter at the new minimum of $25,000 if their optimal capital was at least $14,154. The optimal capital per person also increases. Both of these changes likely come from the many banks still at $50,000 in 1902. By assumption, banks at $50,000 in 1902 wanted to be at $50,000 in 1902, even if they had been forced to be there before, which tends to shift up the average optimal capital. By comparison, a bank with $50,000 in capital in 1900 could have been anywhere between C^T and $50,000. Shifting up optimal capital means the optimal capital when forced to enter at $50,000 must shift up as well. While decades seem long enough that inertia or costs of changing capital should not dominate, the same may not be true of the two and a half years between the reduction of the minimum capital and the Comptroller's report in 1902.

5.2 The causal effects of banking

Table 3 shows the effects of increased banking on the natural log of total production per capita, while table 4 breaks total production into its constituent parts, and examines whether there were changes in the mix of production. All of the estimates include time effects, or are in first differences, which should sweep out overall price changes over time. Since four time periods is a short panel, first differencing allows me to relax the relatively strong assumption that the errors are orthogonal to the amount of banking in a county at all time periods (strict exogeneity) required to make the within group transformation, for the weaker assumption that banking differences are orthogonal to error differences. First differencing comes at at an efficiency cost, however, and since the estimates are generally close, I show the first difference results only for the total production.

For both the fixed effects and first differences, I show both the possibly biased estimate from estimating equation 3 using least squares, as well as the estimates from the multiple imputation around the discontinuity of capital stock at the dividing capital between entry and not. The coefficient on capital stock and its lags thus gives a similar estimate to the optimal capital and its lags, since both represent an effect of banking that may be biased by endogeneity. The coefficients on the excess capital, on the other hand, are a causal effect of banking for those counties near the dividing line. The estimated effect is the local average effect for the marginal counties of getting a bank. By definition, these counties are close to getting or not getting a bank, and so an extra unit of capital in these counties may have a very different effect than in counties that are far from the dividing line for profitable entry. This local average is the most interesting for policy, however, which generally considers whether to encourage (or discourage) banking in areas where banks have not chosen to go themselves, such as rural areas, or serving marginal populations.

To make the estimates with different lags comparable, and put them in units which are meaningful, I calculate the effect of adding a bank with the minimum size capital of $50,000 to a county with the average county population in 1880. Using the average county population may tend to understate the marginal effect of getting a bank on the marginal county—the best county for banking that does not already have a bank—since the marginal county likely has a lower population than the average county, but it is useful as a standard comparison. For the estimates with lags the marginal value is the total effect from both lags of either the capital stock per capita in the OLS regressions, or the excess capital per capita in the multiple imputation regression discontinuity results. I also report the p-value of the hypothesis that the total effect is zero, taking into account the multiple imputations.

An additional $50,000 in capital has a big effect on the production in the marginal county, increasing the total production value per capita by between 9 and 13 percent, in the regressions in levels with a high level of confidence. Including lags seems to increase the total effect, which suggests that banking may have a continuous growth effect, not just a level effect. Of course, with only four time periods, it is not easy to distinguish between them. While later lags are smaller, particularly in the regressions not using 1902, there is no strong evidence for the effect varying over time. The coefficient on the excess capital per capita is substantially larger than either the coefficients on capital stock per capita or optimal capital per capita, which suggests that the effect of adding banking to marginal counties is larger than adding banking to counties which already have some. The coefficients on the capital stock per capita using OLS and on the optimal capital are very close. Although the two estimates are calculated in very different ways, they both capture the endogenous relationship between output and banking, and so suggest that the multiple imputation is capturing the variation in the data well.

To examine how banking affects the components of total productivity, table 4 shows the multiple imputation regression discontinuity estimates for manufacturing value per capita, farm production value per capita, and for the share of manufacturing in total output. For a county with the average population, gaining $50,000 in capital increases manufacturing production, but the estimates are not statistically significant. The increase in total production seems to be particularly driven by increases farm production value per capita, which increases by 7-14 percent depending on the specification.

The fraction of manufacturing in the total production, shown in the last two columns of table 4, suggests that additional capital slightly decreased the share of manufacturing, particularly in the second decade. A zero or negative result is particularly striking because national banks could not lend on mortgages, and so the only direct way to promote agriculture was through financing trade. These results, of course, hold only for the marginal rural county—national banks may have promoted industrialization in the cities, while facilitating trade in rural areas.

Banks seem to have promoted agriculture largely on the extensive margin, rather than on the intensive margin. Table 5 examines how banking affected the fraction of improved farmland in a county, the yield (in dollars of production per acre), and the Gini coefficient of farm size. Improved farmland is land that has been cleared and is being tilled or is lying fallow, and includes orchards and permanent pastures, and so represents land that is actively used for agriculture. Having a bank seems to have increased the fraction of total land in a county that is being actively used for agriculture by around two percent, but had no effect on the yield. Counties with more banking brought more land into production, compared to counties with less, but did not produce more value per acre. Banks seem to have increased inequality of farm size, but only after a decade. The lag suggests that it was not that banks promoted larger farms when a county was first settled, but that they encourage consolidation after settlement.

5.3 Distance to banking

Using geography to identify the effects of banking assumes that distance somehow matters. If someone in a county far from the centers of finance can get a loan just as easily as someone close by, then the presence of a local bank should make little difference. The spread of banks with population in figures 2 through 5 shows that local banking is important—otherwise all banking could be done at lower cost in one location. But in identifying a single effect of extra capital, the estimate ignores that some counties are much closer to other sources of banking than other counties. For example, a county may have banks in towns all along its border, and so have very good access to banking, even though the county itself may not have a bank. The effect of additional banking capital in such a county may be much smaller than in a county far from any bank. To test how distance matters, I construct a measure of distance to banking for each county in each year of the analysis. Since I know the location of each bank, I take the mean distance by area within a county to the nearest bank, which may be within the county itself.[15] Since the effect of the distance to banking is likely to be highly non-linear, I interact the inverse of the distance to banking with the excess capital, and so examine whether having additional capital makes more of a difference in counties which are far away from other banks.

Table 6 shows the results of including an interaction with the inverse of distance to the nearest bank for all rural counties and the restricted sample of counties in the Northern and Midwestern states. While the interaction is not generally significant, its sign is negative, suggesting that the better a county is covered by banks, the smaller difference extra capital makes. Since the sparsely

[15]More formally, the mean distance by area to the nearest bank in a county is $(1/A)\int\int_{(x,y)\in C}\min\{D((x,y),(x_b,y_b))\}dydx$ where A is county area, $D(\cdot,cdot)$ is the appropriate distance function for the projection, and the $\min$ is taken over all banks with locations (x_b,y_b). In practice, I calculate an approximation to the integral using a two dimensional Reimann sum. First I create a raster with kilometer square regions (smaller sizes over the entire US sometimes failed due to the size of the resulting data set), each of which contains the shortest distance to a bank for each year from the center of the square. Using the 1890 county shapefile, I average over distances from the raster regions which fall within each 1890 county, which gives a county level mean distance to the nearest bank.

populated and poorly banked West may be driving much of the relationship, I also estimate the effect of restricting the sample to the Northern and Midwestern states. In these states the effect of distance is large and significant. The estimate for both samples suggest that banks were indeed very local; the benefits of opening a bank start to occur only when a new bank reduces the average distance to a bank by at least eight or nine kilometers and get larger the further away other banks are.

The distance to the nearest bank, just like capital within a county, is endogenous. Banks deciding to enter might very well consider whether there are other banks in nearby counties. The estimation can deal with this endogeneity for capital since it is part of the unobservable business conditions that make a county good for banking, but since all of the variation comes from the optimal capital, it does not provide a way to identify the exogenous effects of distance. So the interaction is capturing some combination of endogenous and exogenous effects, and it is not clear the estimate is capturing a causal effect. The sign and magnitude of the effect of distance do suggest that distance does matter, and adds support to the large effects of extra capital in the main estimates. Since the marginal counties are typically also further away from other sources of banking, it makes sense for the effects of extra banking to be larger.

5.4 Robustness

To check the robustness of the results, I restrict the sample in several ways. Table 7 shows the marginal effect and its p-value for different window sizes around the the transition capital point. As the window shrinks, counties (both in cross-section and over time) become more and more similar across the line, which makes the identification stronger. But as the window shrinks the sample size also gets smaller, so significance tends to fall. As the table shows, however, the point estimates are close even as the window shrinks. The results are not driven by counties far from the transition point, but instead by the discontinuity of capital at the transition point.

The analysis has focused on using the largest sample with the most information about how much capital a county would have had with the minimum capital requirement, so table 8 examines the sensitivity of the results. Using the capital stock in 1902 as the capital counties would have had if the minimum capital had been lower in 1900 requires assuming that the banking situation in 1902 is identical to 1900 except for the variation in capital requirements, and that banks have had a chance to fully adjust by 1902. If those assumptions are met, the extra information from 1902 allows a much more precise estimate of which counties were close to the line, but that precision may come at the cost of bias. The first two columns of table 8 use only internal information from 1870, 1880, 1890, and 1900. The results are largely unchanged, although seem to increase slightly.

While I include all rural counties, including those in the sparsely populated West and poorly banked South in most of the analysis, it is possible that these areas had a different relationship with banking than the North and Midwest. Columns four through six in table 8 restrict restricted the sample to rural counties in Northern and Midwestern states, since these counties are the most comparable. Note that is is not necessarily a problem that these areas had few banks: although counties in the West did not get many banks, they also had very low populations, and so should have had few banks. Similarly, the regressions interpret the few banks in the South as meaning that the South was not a good place for banking. Excluding the South and West does not change the results much, although there are important differences when including the extra information from 1902. The estimates of the effect on total production are substantially lower when including 1902 than when not in the Union states. There were many banks at $50,000 capital in the Union states, which did not reduce their capital by 1902, while in the South and West there were fewer banks at $25,000 between 1900 and 1902. For the Union states that suggests that using 1902 may bias the results slightly downward.

Although most of the exposition has focused on the transition between no banks and a bank with $50,000 in capital, the analysis deals with the intermediate situation where there are other banks in a county, possibly with more than the minimum capital. The presence of a bank with the minimum capital, while other banks are not at the minimum, may suggest that at least one bank is constrained. Alternatively, the total banking in the county may by entirely endogenous, and the existence of a bank with the minimum capital comes from competition or agreement among banks, rather than capital constraints. The last two columns in table 8 therefore remove the relatively few counties that have banks with $50,000 and banks with more capital. The results are not driven by these counties.

5.5 The very long-term effects of national banking

Finally, I briefly examine the long-term effects of national banks during the Gilded Age. While over the short term—where "short term" here is several decades—I show that differences in banking cause differences in economic development, over the long term one might expect any initial advantage to disappear, or even turn into a disadvantage as impatient consumers adjust to the availability of credit (Fulford, 2011). For example, over the long term mobile capital should seek the best marginal return, and so initial differences in financial development may have little long-term effect. Yet a growing literature points to the longevity of institutions and initial advantages. For example, Pascali (2011) finds that the presence of Jewish communities in Italy, who could lend in the middle ages, is a significant predictor of financial development in modern Italy.

Does banking increase income not just within decades but a century later? Table 9 shows results of estimating the effect of banking in 1900 and 1870 on income per capita in 1970. I divide up the actual banking capital in a county into optimal capital and excess capital as in the previous analysis, and estimate only for counties close to the dividing line to separate between endogenously determined and exogenously determined capital. Since the data available in 1970 are very different from 1870-1900, I no longer use fixed effects, but instead include the total production value per capita calculated from the 1900 and 1870 census directly. In addition, I include state indicators, to make sure any estimated effect is not being driven by regional differences. To maintain comparability with the previous results, I calculate the effect of additional capital using the same value: an increase in capital per person equivalent to adding $50,000 in capital to a county with the average population in 1880.

The results suggest that adding a bank of the minimum size in 1900 is associate with an increase in income per capita of 2% in 1970. The value is stable including either production per capita in 1900 or state effects, or both. The effect doubles when neither state effects nor production per capita in 1900 are included (the second column), which confirms the perhaps obvious point that regional disparities which existed in 1900 have not disappeared. The poorly banked and poor South in 1900 is still relatively poor in 1970. Including state effects, however, allows the comparison to occur at the state level: counties with more banking capital in 1900 have 2% higher incomes than other counties within their states, whether or not they were more productive in 1900.

It does not seem to matter where the capital in 1900 came from. The marginal effect of the endogenously determined capital is statistically the same as the marginal effect for the capital caused by the minimum capital requirement. One might expect that additional banking coming from an institutional rule which was relaxed in 1902 would have far less persistence than banking that responded to conditions within the county, but it turns out that additional banking in 1900 has a similar effect no matter what the source. That suggests that historical accidents can have long term effects.

The last column of table 9 compares conditions in 1870 to 1970. The effects are smaller, and no longer statistically significant. That may be because of the additional 30 years. Alternatively, as figures 2 through 5 show, between 1870 and 1900 banking spread through much of the country. An advantage in banking in 1870 with much of the country unbanked may soon have been erased as banking spread, while an advantage in 1900 may have been much more persistent. The difference points to the importance of understanding the mechanisms of how financial development affects growth in the long-term: is it through the institutional persistence of banks, the persistence of the growth they cause, or simply very slow reversion to a steady state?

6 Conclusion

Although much has been written about the financial institutions during the period, this paper is the first to estimate the causal effects of banking during the Gilded Age. Banks, while certainly following growth, also contributed substantially to it: for the marginal county close to the line between getting a bank and not, the presence of a bank was worth an extra 9-13 percent in total production per capita. Banks seem to have promoted both farming, and to a less well estimated extent, manufacturing, but in marginal counties tilted the production mix to agriculture. Agriculture increased by increasing acreage rather than improving yields. Moreover, it seems that the distance to a bank is very important: banks coming into areas with few banks have a larger effect.

One way to read these results is that bad banking regulation can be very costly. An initially somewhat arbitrary decision by an administrator not to allow branching became entrenched with costly long-term consequences. The estimates are possible because of this rule, but also represent its high local cost. The no-branching rule promoted smaller unit banks, while denying many places that could have supported a branch of larger, possibly better diversified, bank the benefits of banking. In the long term the inability to branch seems to have reduced competition and harmed growth, as Jayaratne and Strahan (1996) find when examining the consequences of relaxing limits on inter-state branching a century later. Such continuing restrictions on capital mobility may suggest why counties with more capital in 1900 still had higher incomes in 1970.

National banks both issued loans, typically of short duration and often to fund goods in transit, and national bank notes, which as currency facilitated the exchange of goods. It is not clear whether the effects national banks had on production came from the increase in the local money supply or the local credit supply. Indeed, the two may not be separable: Schumpeter argues "that all forms of credit, from the bank-note to book-credits, are essentially the same thing" (Schumpeter, 1934, p. 99), and he might be right. National bank notes traveled widely, however, and were redeemed only infrequently (Selgin and White, 1994), and so their effects on local conditions may have been small after the initial offering. In either case, since banks were generally not making loans to expand businesses or farms directly, and could not take mortgages as collateral, banks helped grease the wheels of commerce, rather than provided the capital to create new enterprises.

What lessons does this episode have for modern development? First, it seems that in the face of expensive and time consuming transport, commercial banking can be very important. Since developing countries often have limited infrastructure, particularly in rural areas, the major constraint may be less the ability to expand, which can often be done incrementally, but a liquidity constraint from the timing of payments. Farmers may have income only once a year, but have expenses all year long, merchants may have to keep a stock of goods which can be acquired only infrequently, or at lower prices in bulk. Providing working capital and funding goods in transit are not secondary functions of banks during development, historically they are the key functions of banks during development. Second, distance to banking matters, at least when transportation or communication is an issue. While national capital markets may be important as a way to acquire capital, local financial institutions matter for local growth. Third, banking can encourage production, but it is also likely to encourage production at the most efficient scale, and by the most efficient producers, which may tend to increase inequality. The recent resurgence of interest in how financial institutions, whether banks or microfinance institutions, can help development has made it even more important to understand the mechanisms of how financial institutions actually helped development.

References

Aghion, Philippe and Patrick Bolton. 1997. "A Theory of Trickle-Down Growth and Development." *The Review of Economic Studies* 64 (2):151–172.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2009. "The miracle of microfinance? Evidence from a randomized evaluation." Available http://econ-www.mit.edu/files/4162 (accessed 14 August 2009).

Banerjee, Abhijit V. 2001. "Contracting Constraints, Credit Markets and Economic Development." MIT Department of Economics Working Paper 02-17 (Available at SSRN: http://ssrn.com/abstract=306990).

Banerjee, Abhijit V. and Andrew F. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101 (2):274–298.

Bank, World. 2009. *World Development Report 2009: Reshaping Economic Geography*. Washington, D.C.

Barnett, George E. 1911. *State Banks and Trust Companies since the Passage of the National Bank Act*. Washington: Goverment Printing Office.

Benfratello, Luigi, Fabio Schiantarelli, and Alessandro Sembenelli. 2008. "Banks and innovation: Microeconometric evidence on Italian firms." *Journal of Financial Economics* 90 (2):197 – 217.

Binder, John J. and David T. Brown. 1991. "Bank Rates of Return and Entry Restrictions, 1869-1914." *The Journal of Economic History* 51 (1):47–66.

Bodenhorn, Howard. 2000. *A History of Banking in Antebellum America: Financial Markets and Economic Development in an Era of Nation Building*. Cambridge, United Kingdom: Cambridge University Press.

Bolles, Albert S. 1903. *Practical Banking*. Indianapolis: Levey Bro's & CO., eleventh edition ed.

Burgess, Robin and Rohini Pande. 2005. "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment." *The American Economic Review* 95 (3):780–795.

Cagan, Philip. 1963. *Banking and Monetary Studies*, chap. The first fifty years of the national banking system—An historical appraisal. Richard D. Irwin, Inc., 15–42.

Campen, James T. and Anne Mayhew. 1988. "The National Banking System and Southern Economic Growth: Evidence from One Southern City, 1870-1900." *The Journal of Economic History* 48 (1):pp. 127–137.

Census Office. 1880. *Statistics of the Population of the United States at the Tenth Census*, vol. 1. Washington: Goverment Printing Office. Available: http://www.census.gov/prod/www/abs/decennial/1880.html, accessed 8 July 2010.

Davis, Lance E. 1965. "The Investment Market, 1870-1914: The Evolution of a National Market." *The Journal of Economic History* 25 (3):355–399.

de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *The Quarterly Journal of Economics* 123 (4):1329–1372.

Dehejia, Rajeev and Adriana Lleras-Muney. 2007. "Financial Development and Pathways of Growth: State Branching and Deposit Insurance Laws in the United States, 1900–1940." *Journal of Law and Economics* 50 (2):pp. 239–272.

Driscoll, John C. 2004. "Does bank lending affect output? Evidence from the U.S. states." *Journal of Monetary Economics* 51 (3):451–471.

Dunbar, Charles F. 1892. *The theory and history of banking*. New York: G.P. Putnam's Sons.

Friedman, Milton and Anna Jacobson Schwartz. 1963. *A monetary history of the United States, 1867-1960*. Princeton: Princeton University Press, ninth paperback printing, 1993. ed.

Fulford, Scott. 2011. "The effects of financial development in the short and long run." Boston College Working Paper 741.

Galor, Oded and Omer Moav. 2004. "From Physical to Human Capital Accumulation: Inequality and the Process of Development." *The Review of Economic Studies* 71 (4):1001–1026.

Galor, Oded and Joseph Zeira. 1993. "Income Distribution and Macroeconomics." *Review of Economic Studies* 60:35–52.

Greenwood, Jeremy and Boyan Jovanovic. 1990. "Financial Development, Growth, and the Distribution of Income." *The Journal of Political Economy* 98 (5):1076–1107.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "Does Local Financial Development Matter?" *Quarterly Journal of Economics* 119 (3):929–969.

Haines, Michael R. and ICPSR. 2010. "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002." ICPSR study ICPSR02896-v3, available: http://dx.doi.org/10.3886/ICPSR02896, accessed 7 July 2010.

Jack, William and Tavneet Suri. 2010. "The Economics of M-PESA." Available: http://www.mit.edu/~tavneet/M-PESA.pdf, accessed 30 January 2010.

James, John A. 1976a. "Banking Market Structure, Risk, and the Pattern of Local Interest Rates in the United States, 1893-1911." *The Review of Economics and Statistics* 58 (4):453–462.

———. 1976b. “The Conundrum of the Low Issue of National Bank Notes.” *The Journal of Political Economy* 84 (2):359–368.

———. 1976c. “The Development of the National Money Market, 1893-1911.” *The Journal of Economic History* 36 (4):878–897.

———. 1978. *Money and Capital Markets in Postbellum America*. Princeton: Princeton University Press.

Jayaratne, Jith and Philip E. Strahan. 1996. “The Finance-Growth Nexus: Evidence from Bank Branch Deregulation.” *The Quarterly Journal of Economics* 111 (3):pp. 639–670.

Keehn, Richard H. and Gene Smiley. 1977. “Mortgage Lending by National Banks.” *The Business History Review* 51 (4):474–491.

King, Robert G. and Ross Levine. 1993. “Finance and Growth: Schumpeter might be right.” *Quarterly Journal of Economics* 108 (3):717–737.

Lancaster, Tony. 2004. *An Introduction to Modern Bayesian Econometrics*. Malden, MA: Blackwell.

Lee, David S. and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48 (2):281–355.

Levine, Ross. 2005. *Handbook of Economic Growth*, vol. 1, chap. Finance and Growth: Theory and Evidence. Amsterdam: Elsevier, 865–934.

Maddison, Angus. 1995. *Monitoring the World Economy 1820-1992*. Paris: OECD Development Centre Studies.

Minnesota Population Center. 2004. “National Historical Geographic Information System.” Minneapolis, MN, University of Minnesota. Available: http://www.nhgis.org, accessed 7 July 2010.

Moulton, H. G. 1918a. “Commercial Banking and Capital Formation: I.” *The Journal of Political Economy* 26 (5):484–508.

———. 1918b. “Commercial Banking and Capital Formation: II.” *The Journal of Political Economy* 26 (6):638–663.

———. 1918c. “Commercial Banking and Capital Formation: III.” *The Journal of Political Economy* 26 (7):705–731.

———. 1918d. “Commercial Banking and Capital Formation: IV.” *The Journal of Political Economy* 26 (9):849–881.

Panagariya, Arvind. 2008. *India: The Emergining Giant*. Oxford: Oxford University Press.

Pascali, Luigi. 2011. "Banks and Development: Jewish Communities in the Italian Renaissance and Current Economic Performance." Barcelona GSE Working Paper 562, available: http://research.barcelonagse.eu/tmp/working_papers/562.pdf.

Rajan, Raghuram G. and Luigi Zingales. 1998. "Financial Dependence and Growth." *The American Economic Review* 88 (3):559–586.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Schafer, Joseph L. 1997. *Analysis of Incomplete Multivariate Data*. Boca Raton, FL: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

Schumpeter, Joseph A. 1934. *The Theory of Economic Development*. Cambridge, MA: Harvard University Press, sixth printing, 1959 ed. Translated by Redvers Opie.

Selgin, George A. and Lawrence H. White. 1994. "Monetary Reform and the Redemption of National Bank Notes, 1863-1913." *The Business History Review* 68 (2):pp. 205–243.

StataCorp. 2009. *Stata Logitudinal-Data/Panel-Data Reference Manual: Release 11*. College Station, TX: StataCorp LP.

Sullivan, Richard J. 2009. "Regulatory Changes and the Development of the U.S. Banking Market, 1870-1914: A Study of the Profit Rates and Risk in National Banks." In *The Origins and Development of Financial Markets and Institutions*, edited by Jeremy Atack and Larry Neal. Cambridge University Press, 262–294.

Sylla, Richard. 1969. "Federal Policy, Banking Market Structure, and Capital Mobilization in the United States, 1863-1913." *The Journal of Economic History* 29 (4):657–686.

Townsend, Robert M. and Kenichi Ueda. 2006. "Financial Deepening, Inequality and Growth: A Model-Based Quantitative Evaluation." *Review of Economic Studies* 73:251–293.

Updike, Helen Hill. 1985. *The National Banks and American Economic Development, 1870-1900*. New York: Garland Publishing.

U.S. Board of Geographic Names. 2010. "Graphical Names Information System." Available: http://geonames.usgs.gov/, accessed 7 July 2010.

U.S. Congress. 1864. "An Act to provide a National Currency, secured by a Pledge of United States Bonds, and to provide for the Circulation and Redemption thereof." Session I, Chapter 106.

Watkins, Myron W. and H. G. Moulton. 1919. "Commercial Banking and Capital Formation: A Criticism." *The Journal of Political Economy* 27 (7):578–605.

White, Eugene Nelson. 1983. *The Regulation and Reform of the American Banking System, 1900-1929*. Princeton, NJ: Princeton University Press.

Wicker, Elmus. 2000. *Banking panics of the Gilded Age*. Cambridge: Cambridge University Press.

Wright, Ivan. 1922. *Bank Credit and Agriculture under the National and Federal Reserve Banking Systems*. New York: McGraw Hill Book Company.

Young, Allyn A. 1924. "An Analysis of Bank Statistics For the United States." *The Review of Economics and Statistics* 6 (4):284–296.

———. 1925a. "An Analysis of Bank Statistics for the United States." *The Review of Economics and Statistics* 7 (1):19–37.

———. 1925b. "An Analysis of Bank Statistics for the United States." *The Review of Economics and Statistics* 7 (2):86–104.

———. 1927. "An Analysis of Bank Statistics for the United States." *The Review of Economics and Statistics* 9 (3):121–141.

A Minimum capital requirements and banking profit maximization

Minimum capital requirements mean that some banks open in areas with more than optimal capital, while some areas do not have any banks. Then the key insight into any model of profit maximization is that the further away the minimum capital requirement forces a bank to be from its profit maximizing capital, the lower are its profits, until at some point profits are negative, and the bank will not open. While this section describes a particular model which gives a break even point, other models should give similar results, and the empirical strategy only depends on the break even point, not on the particular underlying functions.

Bank profits come from the difference between the return on the loans it makes and the cost of raising the capital to make those loans. Then $\pi = r_L(L)L - r_B(K)K$ where K is total capital and L is the loans. The capital stock is in fixed proportion to the total capital $\omega K = C$, which holds approximately true in the data. Any reserve requirements and operating costs are included in the cost of raising capital.

The borrowing rate r_B and lending rate r_L depend on the amount of capital. Local demand for loans is downward sloping, but depends on the business conditions and population in a given county. Counties with a large population engaged in activities which need banking services demand more loans. Then in county c at time t with population P_{ct} the demand for loans is $r_L(L) = \alpha_0 \alpha_t P_{ct} \eta_c \epsilon_{ct} - \alpha_1 L$. In some counties η_c is high, and the need for banking services is higher, in others η_c is low, and even large populations may demand few loans. The overall demand for loans may change over time, and counties receive idiosyncratic shocks ϵ_{ct} which change their demand for loans in a given time. The cost of capital increases with the amount of capital so that $r_B(K) = \rho K$.

Ignoring minimum capital requirements, banks can choose their optimal size and so make money anywhere the demand for loans is positive. Maximizing profits the optimal capital is:

$$C^*_{ct} = \frac{\alpha_0 \alpha_t P_{ct} \eta_c \epsilon_{ct}}{2(\alpha_1 + \rho)/\omega} = \alpha \alpha_t P_{ct} \eta_c \epsilon_{ct}$$

where α absorbs all of the other model parameters. Although the model has been structured as one of demand shocks, since it seems reasonable that demand for banking services varies while the cost of capital may be similar across counties, counties may face different and time varying costs of capital as well. The end result that C^*_{ct} depends on local conditions does not change, only the underlying parameters.

But banks cannot open with any capital, they must always have at least \$50,000. If $C^*_{ct} > 50$, then the observed capital is $C_{ct} = C^*_{ct}$, since the bank could choose its capital optimally. If $C^*_{ct} \leq 50$, however, then the observed capital $C_{ct} = 50$, while the optimal capital must be somewhat less, and the bank will not enter if profits are less than zero. Using the linear assumptions for supply and demand which create a symmetric profit function, the point where profits are zero is given when the optimal capital is 25, which is found by solving $\pi(50) = 0$ and substituting C^*_{ct}.

Profits need not be symmetric (or quadratic). More generally, I assume that the optimal capital $C^*_{ct} = \text{argmax}_C \pi(C, P_{ct}\eta_c\epsilon_{ct}) = \alpha\alpha_t P_{ct}\eta_c\epsilon_{ct}$ can be expressed in a log linear form. When constrained to enter at \$50,000, define C^T as the capital that a bank would find optimal if it could enter freely when it is just willing to enter with a capital at \$50,000. More formally, if x_0 is the solution to $\pi(50, x) = 0$, then $C^T = \text{argmax}_C \pi(C, x_0)$.

An individual bank in a county acts as a monopolist. Multiple banks divide up the county among themselves and act as monopolists within their portion. Then bank i with capital C^i_{ct} gets $p^i = C^i_{ct}/C_{ct}$ of the population. A bank with $C^i_{ct} = 50$ must have $C^T < \alpha\alpha_t p^i P_{ct}\eta_c\epsilon_{ct} \leq 50$ and an unconstrained bank has $C^i_{ct} = \alpha\alpha_t p^i P_{ct}\eta_c\epsilon_{ct}$. Adding up all banks gives $C_{ct} - C^T B^{50}_{ct} < \alpha\alpha_t P_{ct}\eta_c\epsilon_{ct} \leq C_{ct}$ where B^{50}_{ct} is the number of banks with \$50,000 in capital stock.

Putting it all together, if a county has capital stock C_{ct} in thousands, which may be zero, B_{ct} banks, and B^{50}_{ct} banks with 50 capital, and the dividing line between entering not is C^T, then: (1) Optimal capital is observed capital if the bank is unconstrained: $C_{ct} = \alpha\alpha_t P_{ct}\eta_c\epsilon_{ct}$ if $B_{ct} > 0$ and $B^{50}_{ct} = 0$; (2) Optimal capital must be between C^T and 50 if the bank is constrained but enters: $C_{ct} - (50 - C^T)B^{50}_{ct} < \alpha\alpha_t P_{ct}\eta_c\epsilon_{ct} \leq C_{ct}$ if $B^{50}_{ct} > 0$; and (3) Optimal capital must be less than

C^T if the no bank enters: $\alpha \alpha_t P_{ct} \eta_c \epsilon_{ct} \leq C^T$ if $B_{ct} = 0$ and so $C_{ct} = 0$.

B Estimating optimal capital using maximum likelihood

Because of minimum capital requirements, some counties have more capital than they would get if banks could freely choose their capital. The observed amount of capital is censored at $50,000 and zero, however. Conditional on observing a bank with $50,000 in capital, the optimal capital cannot be too much less than $50,000, otherwise a bank would not want to enter. Similarly, if there is no bank in a county, the optimal capital cannot be very close to $50,000, since otherwise it would be profitable to enter.

For any given profitability cutoff C^T, suppose that the county businesses condition shifters are distributed lognormally so that: $\ln \epsilon_{ct} \sim N(0, \sigma^2)$ and $\ln \eta_c \sim N(0, \sigma_\eta^2)$. Define D_{ct} as 1 if there are any banks in the county and 0 otherwise, D_{ct}^{50} as 1 if there is a bank with $50,000 capital and 0 otherwise, and B_{ct}^{50} as the number of banks with $50,000 in capital. Then the density of observing capital C_{ct} in county c in year t, given η_c, α, α_t, and P_{ct} is:

$$f(C_{ct}|P_{ct}, \alpha, \alpha_t, \eta_c) = \left(\Phi \left[\ln(C^T/P_{ct}) - \alpha - \alpha_t - \eta_c)/\sigma \right] \right)^{(1-D_{ct})} + \\ \left(\left(\phi \left[\left(\ln \frac{C_{ct}}{P_{ct}} - \alpha - \alpha_t - \eta_c \right) / \sigma \right] / \sigma \right)^{(1-D_{ct}^{50})} + \\ \left(\Phi \left[\left(\ln \frac{C_{ct}}{P_{ct}} - \alpha - \alpha_t - \eta_c \right) / \sigma \right] - \Phi \left[\left(\ln \frac{C_{ct} - (50 - C^T) B_{ct}^{50}}{P_{ct}} - \alpha - \alpha_t - \eta_c \right) / \sigma \right] \right)^{D_{ct}^{50}} \right)^{D_{ct}}. \tag{3}$$

The conditional density of observing $\mathbf{C_c}$, the vector for capital in county c for all years, conditional on η_c is the product of all of the densities for each year:

$$f(\mathbf{C}_c|\mathbf{P}_c, \alpha, \alpha_{1880}, \alpha_{1890}, \alpha_{1890}, \eta_c) = \prod_{t=1870}^{1900} f(C_{ct}|P_{ct}, \alpha, \alpha_t, \eta_c).$$

Finally, the density unconditional on η_c comes from integrating out η_c:

$$f(\mathbf{C}_c|\mathbf{P}_c, \alpha, \alpha_{1880}, \alpha_{1890}, \alpha_{1890}, \sigma, \sigma_\eta) = \int f(\mathbf{C}_c|\mathbf{P}_c, \alpha, \alpha, \eta_c)\phi[\eta_c/\sigma_\eta]/\sigma_\eta d\eta_c. \tag{4}$$

Maximum likelihood then finds the set of parameters which maximize the sum over all counties of the log of equation 4. The difficulty is that for each county the observations in each time period are dependent on each other through η_c, and so standard maximization routines which rely on each piece of the log likelihood being separate do not apply. The solution is to actually perform the integration over η_c to get the unconditional density, or at least a numerical approximation of it, and then maximize over the county level likelihood. This approach is implemented in Stata using a Gauss-Hermite quadrature to approximate the integration in equation 4 by xtintreg (StataCorp, 2009, pp. 180–188).

C Drawing estimates of excess capital

Given estimates of the parameters which determine the optimal capital, and realizations of ϵ_{ct} and η_{ct}, a county with a 50 bank has $50 - \alpha\alpha_t P_{ct}\eta_{ct}\epsilon_{ct}$ too much capital. Counties without banks have $\alpha\alpha_t P_{ct}\eta_{ct}\epsilon_{ct}$ too little capital. I cannot observe ϵ_{ct} and η_{ct}, but I do know how they are distributed, conditional on observing a particular value of C_{ct} and the number of banks in a county. Reversing the constraints for the maximum likelihood:

1. $\ln(C_{ct}/P_{ct}) - \hat{\alpha} - \hat{\alpha}_t = \eta_c + \epsilon_{ct}$ if $B_{ct} > 0$ and $B^{50}_{ct} = 0$,

2. $\ln(C_{ct} - (50 - \hat{C}^T B^{50}_{ct})/P_{ct}) - \hat{\alpha} - \hat{\alpha}_t < \eta_c + \epsilon_{ct} \leq \ln(C_{ct}/P_{ct}) - \hat{\alpha} - \hat{\alpha}_t$ if $B^{50}_{ct} > 0$, and

3. $\eta_c + \epsilon_{ct} \leq \ln(\hat{C}^T/P_{ct}) - \hat{\alpha} - \hat{\alpha}_t$ if $B_{ct} = 0$ and so $C_{ct} = 0$.

Call lb_{ct} the lower bound for $\eta_c + \epsilon_{ct}$ for county c at time t which can be negative infinity, and similarly for ub_{ct}. Then the joint density is:

$$f(\epsilon_{c1870}, \ldots, \epsilon_{c1870}, \eta_c | lb_{ct} < \eta_c + \epsilon_{ct} \leq ub_{ct})$$

is a multivariate truncated normal. To draw from such distribution I use a Gibbs Sampler (Lancaster, 2004, pp. 207–221) which converges to the multivariate truncated normal. Since the distribution of each ϵ_{ct} is univariate truncated normal conditional on η_c, and similarly η_c is univariate truncated normal conditional on all ϵ_{ct}, the Gibbs sampler has the steps:

1. (a) Choose η_c^0 for all counties. (b) Draw ϵ_{ct}^0 from a truncated normal with bounds $lb_{ct} - \eta_c^0$ and $ub_{ct} - \eta_c^0$ and variance σ^2.[16]

2. (a) Given $\{\epsilon_{ct}^{i-1}\}$, draw η_c^i from a univariate truncated normal with lower bound $\max_t\{lb_{ct}\} - \epsilon_{ct}^{i-1}$ and upper bound $\min_t\{ub_{ct} - \epsilon_{ct}^{i-1}\}$, and variance σ_η^2. (b) Given η_c^i draw each element of $\{\epsilon_{ct}^i\}$ from its univariate truncated normal.

3. Repeat step two N times, where N is large, and take the Nth draw of $\{\epsilon_{ct}^N\}$ and η_c^N.

I choose η_c^0 by letting $(\eta_c^0)_t = lb_{ct} + (ub_{ct} - lb_{ct})/2$ if county c in time t is a type two county and $(\eta_c^0)_t = (\ln(C^T/P_{ct}) - \hat{\alpha} - \hat{\alpha}_t$ if the county is a type three county, and then averaging the $(\eta_c^0)_t$'s. For multiple draws in the imputations, I continue the sampler and use every Nth draw as an approximately independent draw from the multivariate truncated normal, thus letting the sampler run MN times.

[16] Drawing from a truncated normal using standard psuedo-random number generators takes a bit of work. If u_i is a draw from a uniform [0,1] distribution, then $x_i = F^{-1}(u_i) = (1/\sigma)\Phi^{-1}[\Phi[a/\sigma] + u_i(\Phi[b/\sigma] - \Phi[a/\sigma])]$ is a draw from a truncated normal with lower bound a, upper bound b, and variance σ^2.

D Data

D.1 County level data for 1860-1920

Counties or their equivalents are geographic subdivisions of states. For most states the county has a governmental role as a middle level of government between the state and the local governments of cities, towns, or boroughs. The US Census collects data at the county level, and for consistency I assign independent cities (such as St. Louis, Missouri) to their own counties, or to the county which contains them. The creation of several new states from territories, and their division into counties between 1870 and 1900 makes consistent geographical designations very important. In established states, counties shift boundaries and split occasionally. To create a data set which consistently refers to the same geographic area over time, I use graphically information from the National Historical Graphical Information System (Minnesota Population Center, 2004), which provides county boundaries for each decade. From the county shape file for each decade, I calculate the union with the 1890 counties, which gives the 1890 county that all counties or parts of counties belong to in a given year. I also calculate the area of each county fragment. I discard all fragments with less than 1 mile square since these fragments represent small shifts in the county polygons rather than changes in county definition. The NHGIS also provides county level census information linked to the county shape files based on the census data from Haines and ICPSR (2010). Since some 1890 counties are composed of pieces from multiple counties, I allocate census economic and demographic information by county for each non-1890 census year using the area of the county fragment. So a county whose boundary shifts to include some of another county gets additional population from the other county in proportion to the area absorbed. I create 1890 county level means or distribution variables by taking the area weighted average of the county fragments that compose the 1890 county. This procedure is exact if population or other aggregates are uniformly distributed over the surface area of changing counties. While such changes are important for the sparsely settled Western states, for the established states on which I conduct much of the analysis, county boundaries are very stable.

D.2 National Bank Accounts

Each year the Comptroller of the Currency reported the accounts of the all National Banks.[17] For each bank in 1870, 1880, 1890, and 1900, I have its number (matched to 1871 for 1870), its place of business, and its loans and discounts, capital stock, and total liabilities.[18] The Comptroller's report includes other interesting information for banks which we have not entered, including notes outstanding, deposits, and reserves. I match the place of business to the Graphical Names Information System maintained by the US Board of Geographic Names U.S. Board of Geographic Names (2010). While most places with a National Bank still exist, some have merged with other towns or cities, in which case I match with the modern city. The match gives the latitude and longitude of the bank using the North American Datum of 1983. I match the bank location with the 1890 counties above, which gives the number of banks and national bank aggregates in the geographic area of an 1890 county for each decade from 1870 to 1900.

[17] Available in pdf from the St Louis Federal Reserve, http://fraser.stlouisfed.org/publications/comp/, accessed 7 July 2010.

[18] My excellent RA Shahed Kahn entered these series by hand, with the help of some optical character recognition to speed the process of entry. I checked the 1870 and 1880 accounts, and another RA, Mashfiqur Khan, checked the later years.

Table 1: County descriptive statistics

	All Counties				Rural Union counties			
Year	1870	1880	1890	1900	1870	1880	1890	1900
County population	14254	18009	22365	27173	14881	18224	21583	24359
	(32640)	(41128)	(54905)	(70633)	(14012)	(15321)	(18598)	(21762)
Total production	106.8	83.2	100.8	157.7	112.6	96.0	112.5	153.7
value per capita	(86.3)	(70.0)	(84.6)	(311.1)	(68.8)	(63.3)	(82.0)	(93.9)
Manufacturing	41.9	35.9	49.3	64.4	44.7	43.0	58.9	73.3
value per capita	(73.0)	(60.8)	(81.9)	(103.0)	(58.3)	(57.4)	(86.0)	(100.5)
Farm production	64.8	47.3	50.3	93.3	67.9	53.0	53.0	80.4
value per capita	(48.3)	(37.7)	(40.5)	(302.3)	(39.6)	(33.0)	(28.8)	(47.4)
Fraction manuf.	0.33	0.32	0.36	0.37	0.33	0.35	0.40	0.39
in total value	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Gini farm size	0.432	0.394	0.398	0.457	0.429	0.375	0.366	0.424
	(0.170)	(0.169)	(0.154)	(0.105)	(0.141)	(0.126)	(0.123)	(0.073)
Fraction improved	0.37	0.36	0.53	0.51	0.45	0.45	0.64	0.65
farmland	(0.24)	(0.21)	(0.23)	(0.25)	(0.24)	(0.19)	(0.20)	(0.21)
Farm yield	41.60	47.58	42.38	77.14	48.44	45.01	39.70	41.91
	(65.86)	(119.78)	(82.04)	(535.20)	(25.17)	(28.44)	(32.26)	(23.29)
Number of banks	0.57	0.74	1.26	1.29	0.61	0.85	1.43	1.51
	(2.45)	(2.68)	(3.11)	(3.05)	(1.43)	(1.68)	(2.14)	(2.28)
Distance to closest bank (km)	152.0	88.7	42.1	39.6	74.6	56.2	27.6	26.1
	(176.14)	(89.65)	(37.27)	(35.44)	(82.63)	(62.85)	(20.84)	(21.04)
Banks per 1000 capita	0.016	0.021	0.051	0.042	0.023	0.030	0.058	0.052
	(0.043)	(0.050)	(0.094)	(0.073)	(0.045)	(0.054)	(0.079)	(0.061)
Capital stock	2.34	2.59	4.70	3.63	2.87	3.32	5.19	4.34
per capita	(8.11)	(7.23)	(8.71)	(6.43)	(7.24)	(7.12)	(7.79)	(5.99)
Loans and discounts	3.36	4.57	11.16	11.34	4.06	5.50	12.08	13.16
per capita	(11.67)	(13.67)	(22.25)	(22.97)	(9.63)	(11.40)	(18.90)	(18.25)
Counties	2705	2785	2800	2795	1222	1262	1262	1262

Notes: Standard deviations on in parentheses. The average is taken over counties and is unweighted. Values are in dollars from the census or national bank accounts and are not corrected for inflation/deflation (there was significant deflation between 1870-1880 as the US went back to a full gold backing of its currency). Rural Union counties are counties from Union or border states with an urban population of fewer than 50,000. Yield is the farm production value divided by the area improved farmland in a county, excluding extreme values driven low areas of improved farmland).

Table 2: Log-likelihood estimates of optimal capital

	All Rural	All Rural with 1902	Rural Union	Rural Union with 1902
α	-7.513***	-7.252***	-7.024***	-6.817***
	(0.0514)	(0.0481)	(0.0596)	(0.0556)
α_{1880}	0.119***	0.0961***	0.126***	0.106***
	(0.0341)	(0.0343)	(0.0404)	(0.0405)
α_{1890}	0.783***	0.700***	0.670***	0.601***
	(0.0329)	(0.0329)	(0.0393)	(0.0393)
α_{1900}	0.604***	0.764***	0.552***	0.720***
	(0.0329)	(0.0328)	(0.0394)	(0.0390)
σ_η	1.735***	1.642***	1.540***	1.427***
	(0.0393)	(0.0352)	(0.0452)	(0.0405)
σ_ϵ	0.680***	0.688***	0.664***	0.672***
	(0.0101)	(0.0100)	(0.0123)	(0.0122)
C^T	19.454	24.359	17.977	22.926
C^T_{25}		14.154		13.221
Observations	10804	10804	4998	4998
Counties	2745	2745	1262	1262

Notes: Estimates of the optimal capital equation $C^*_{ct} = \alpha\alpha_t P_{ct}\eta_c\epsilon_{ct}$ based on the panel using C^T as the dividing line for entering with $50,0000 in capital, and C^T_{25} for entry with $25,000 in the columns using 1902. The full estimation details are in appendix B. The first column uses all rural counties as described in section 4.2, the second column uses all rural counties and the capital in 1902 for 1900, the third and fourth columns restrict the sample to Union and border states.

Table 3: Banks and total production per capita

	log Total production value per capita									
	Level using 1902				First difference using 1902				Level not using 1902	
Capital Stock p.c.	15.52	11.76			9.764	10.68				
	(3.268)	(2.283)			(2.407)	(2.303)				
L—		10.28				8.759				
		(2.774)				(2.082)				
Excess capital p.c.			27.43	17.52			14.29	17.82	33.22	29.96
			(7.847)	(5.185)			(4.794)	(4.868)	(15.10)	(13.23)
L—				21.04				17.32		16.65
				(7.515)				(5.628)		(11.06)
Optimal capital p.c.			13.95	10.09			9.380	9.523	13.58	7.975
			(3.926)	(3.796)			(3.134)	(3.427)	(5.693)	(5.038)
L—				8.407				7.726		10.61
				(3.810)				(3.334)		(4.701)
Observations	10462	7822	6406	4802	7519	4959	4661	3079	6068	4553
R-squared	0.190	0.341			0.154	0.065				
County FE	YES	YES	YES	YES	NO	NO	NO	NO	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Counties	2743	2729	1655	1654	2666	2529	1614	1555	1563	1560
Marginal effect	0.0528	0.0749	0.0933	0.131	0.0332	0.0661	0.0486	0.120	0.113	0.159
p-val marginal $\neq 0$	1.95e-05	3.45e-06	0.00161	0.000842	0.000186	2.51e-07	0.00696	0.000327	0.0392	0.0168
Imputations			100	100			100	100	100	100

Notes: Capital stock is measured in 1000's of dollars. L— is the first lag of the variable. First differences are the first differences for both dependent and independent variables. In each imputation, only counties with optimal capital within 15 of the dividing line which maximizes the log-likelihood in table 2 are included. See appendix for estimation details. Errors are allowed to clustered at the state level. The marginal effect is the the sum of all lagged effects for a county with the average 1880 population gaining $50,000 in capital, and so is the long term effect of gaining a bank of the minimum size on the average county. Uses all rural counties. Forms the imputations of optimal capital using the capital in 1902 for all except the last columns which use only the internal panel.

Table 4: Banks and the mix of production

	log Manufacturing value per capita		log Farm production value per capita		Frac. Manufacturing in total production	
Excess capital p.c.	24.94	17.63	20.44	18.06	-0.418	-0.615
	(18.74)	(18.86)	(7.469)	(6.652)	(1.599)	(1.109)
L—		-10.88		24.00		-3.616
		(13.91)		(7.690)		(1.210)
Optimal capital p.c.	8.996	3.893	5.939	4.823	1.081	-0.0758
	(7.954)	(7.849)	(4.336)	(4.379)	(1.166)	(1.004)
L—		-2.264		7.754		-1.953
		(5.712)		(4.271)		(1.021)
Observations	5824	4436	6427	4867	6406	4802
R-squared						
County FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Counties	1555	1548	1655	1654	1655	1654
Marginal effect	0.0848	0.0230	0.0695	0.143	-0.00142	-0.0144
p-val marg. $\neq 0$	0.194	0.799	0.0110	0.00129	0.796	0.0299
Imputations	100	100	100	100	100	100

Notes: Capital stock is measured in 1000s of dollars. L— is the first lag of the variable. In each imputation, only counties with optimal capital within 15 of the dividing line which maximizes the log-likelihood in table 2 are included. See appendix for estimation details. Errors are allowed to clustered at the state level. The marginal effect is the the sum of the all lagged effects for a county with the average 1880 population gaining $50,000 in capital, and so is the long term effect of gaining a bank of the minimum size on the average county. Uses all rural counties, and forms the imputations of optimal capital using the capital in 1902.

Table 5: Banks and farm production

	Fraction improved farm land		Farm yield		Gini of farm size	
Excess capital p.c.	5.939	4.170	-31.26	-18.18	0.372	0.448
	(1.723)	(1.280)	(61.37)	(63.49)	(0.623)	(0.708)
L—		0.560		-54.74		6.523
		(1.607)		(58.90)		(1.489)
Opt. capital p.c.	3.726	3.041	4.009	-3.019	-0.545	-0.311
	(1.418)	(1.076)	(46.60)	(48.04)	(0.510)	(0.593)
L—		-1.564		-7.159		3.608
		(1.280)		(45.68)		(1.232)
Observations	6549	4882	6401	4858	6549	4882
R-squared						
County FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Counties	1655	1654	1654	1652	1655	1654
Marginal effect	0.0202	0.0161	-0.106	-0.248	0.00127	0.0237
p-val marg. $\neq 0$	0.00144	0.0697	0.614	0.433	0.555	0.000400
Imputations	100	100	100	100	100	100

Notes: Capital stock is measured in 1000's of dollars. L— is the first lag of the variable. In each imputation, only counties with optimal capital within 15 of the dividing line which maximizes the log-likelihood in table 2 are included. See appendix for estimation details. The marginal effect is the the sum of the all lagged effects for a county with the average 1880 population gaining $50,000 in capital, and so is the long term effect of gaining a bank of the minimum size on the average county. Uses all rural counties, and forms the imputations of optimal capital using the capital in 1902. Errors are allowed to clustered at the state level. The window size is 15.

Table 6: Distance to banking

	All rural counties			Union rural counties
	log Total prod. per cap.	log Manuf prod. per cap.	log Farm prod. per cap.	log Total prod. per cap.
Excess capital p.c.	40.07	31.89	32.56	60.42
	(13.08)	(13.65)	(12.92)	(15.98)
Optimal capital p.c.	14.62	16.09	5.937	34.23
	(8.410)	(12.43)	(9.458)	(9.526)
Excess capital p.c. ×	-334.3	-211.5	-293.7	-570.4
1/(Dist. to bank)	(217.2)	(204.3)	(199.5)	(201.3)
Optimal capital p.c. ×	-29.15	-78.23	-22.37	-260.1
1/(Dist. to bank)	(106.4)	(184.5)	(107.8)	(95.77)
Observations	6406	6131	6427	3446
County FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Counties	1655	1648	1655	888
Marginal effect	0.124	0.100	0.0996	0.138
p-val marginal $\neq 0$	0.00284	0.0173	0.0140	0.00730
Imputations	100	100	100	100
Min. dist. for pos. effect	8.34	6.63	9.02	9.44

Notes: The distance is the average over the county area distance in kilometers to the nearest bank. The marginal effect for the distance is the sum of the main effect of $50,000 in excess capital at the mean county population in 1880, and the interaction effect of $50,000 in excess capital times the mean (over counties) distance to the nearest bank in 1880. The last columns includes only the rural Union counties used in the rest of the analysis. The first three columns include all counties for which sufficient data exists, the last restricts the sample to include only rural Union counties. The imputations of optimal capital uses the capital in 1902. The window size is 15. Errors are allowed to clustered at the state level.

Table 7: Sample variations

	All rural counties using 1902 capital reduction					
	Window size 15		Window size 10		Window size 5	
Dependent Variable	1 lag	2 lags	1 lag	2 lags	1 lag	2 lags
log Total production	0.093	0.131	0.099	0.133	0.109	0.154
per capita	[0.002]	[0.001]	[0.003]	[0.003]	[0.013]	[0.013]
log Manufacturing prod.	0.073	0.025	0.072	0.027	0.083	0.051
per capita	[0.010]	[0.539]	[0.029]	[0.570]	[0.081]	[0.438]
log Farm production	0.070	0.143	0.071	0.136	0.080	0.139
per capita	[0.011]	[0.001]	[0.020]	[0.006]	[0.063]	[0.032]
Fraction manufacturing	-0.001	-0.014	-0.002	-0.013	-0.001	-0.010
in total production	[0.796]	[0.030]	[0.701]	[0.102]	[0.918]	[0.344]
Gini of farm size	0.001	0.024	0.002	0.024	0.004	0.025
	[0.555]	[0.000]	[0.551]	[0.001]	[0.287]	[0.008]
Fraction improved	0.020	0.016	0.021	0.015	0.021	0.016
farm land	[0.001]	[0.070]	[0.004]	[0.130]	[0.024]	[0.218]
Farm yield	-0.106	-0.248	-0.130	-0.274	-0.206	-0.242
	[0.614]	[0.433]	[0.623]	[0.485]	[0.622]	[0.658]
Observations	6406	4802	4353	3262	2101	1580
Counties	1655	1654	1125	1123	541	541

Notes: p-values in brackets, testing the combined effect of both lags of excess capital stock per capita. All regressions include county fixed effects, time effects, and cluster at the state level. Observations and counties are both the minimum number from the imputations in the estimates of log total production.

Table 8: Robustness checks

Dependent Variable	All rural counties not using 1902		Rural union using 1902		Rural union not using 1902		Excuding counties with more than one bank	
	1 lag	2 lags	1 lag	2 lags	1 lag	2 lags	1 lag	2 lags
log Total production	0.113	0.159	0.063	0.082	0.148	0.106	0.105	0.142
per capita	[0.039]	[0.017]	[0.013]	[0.014]	[0.017]	[0.165]	[0.002]	[0.001]
log Manufacturing prod.	0.085	0.023	0.047	0.036	0.152	0.081	0.083	0.024
per capita	[0.194]	[0.799]	[0.115]	[0.464]	[0.026]	[0.361]	[0.009]	[0.584]
log Farm production	0.111	0.183	0.061	0.091	0.125	0.111	0.078	0.154
per capita	[0.040]	[0.017]	[0.010]	[0.026]	[0.058]	[0.254]	[0.014]	[0.002]
Fraction manufacturing	0.002	-0.020	0.003	-0.005	0.008	-0.003	-0.002	-0.015
in total production	[0.848]	[0.192]	[0.532]	[0.290]	[0.466]	[0.833]	[0.759]	[0.036]
Gini of farm size	0.001	0.037	0.001	0.018	0.001	0.031	0.001	0.024
	[0.772]	[0.003]	[0.626]	[0.020]	[0.894]	[0.052]	[0.648]	[0.001]
Fraction improved	0.022	0.017	0.028	0.026	0.054	0.047	0.021	0.017
farm land	[0.106]	[0.357]	[0.001]	[0.001]	[0.008]	[0.023]	[0.002]	[0.077]
Farm yield	-0.166	-0.404	0.077	0.266	0.068	-0.052	-0.113	-0.254
	[0.742]	[0.611]	[0.836]	[0.420]	[0.926]	[0.954]	[0.637]	[0.476]
Observations	6068	4553	3446	2609	3110	2360	6328	4744
Counties	1563	1560	888	888	805	805	1636	1636

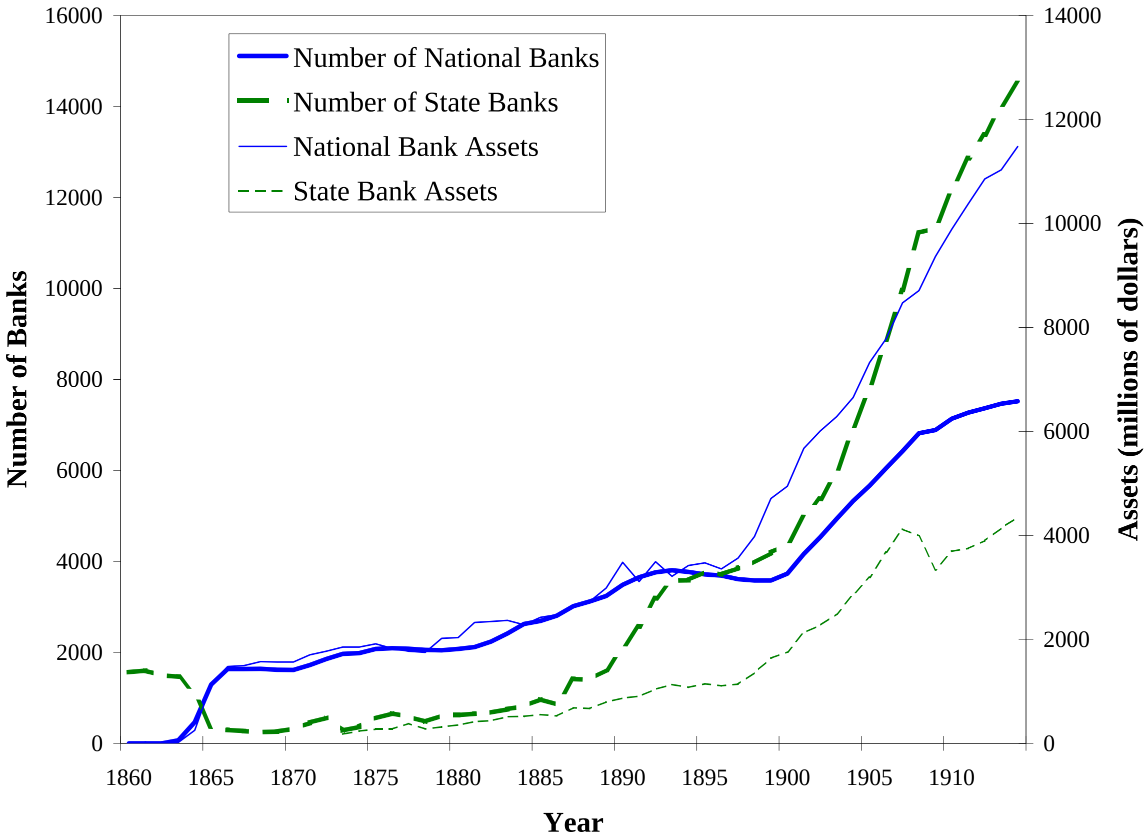
Notes: p-values in brackets, testing the combined effect of both lags of excess capital stock per capita. All regressions include county fixed effects, time effects, and cluster at the state level. Observations and counties are both the minimum number from the imputations in the estimates of log total production. The last two columns includes only counties which have banks with 50 capital or 0 capital. The optimal capital estimates are in table 2. The window size is 15. Errors are allowed to clustered at the state level.

Table 9: Long term differences in income

	log income per capita in 1970				
	Counties in 1900				Counties in 1870
log Total prod. per capita			0.126	0.201	0.0414
			(0.0225)	(0.0238)	(0.0159)
Excess capital p.c.	6.777	12.04	5.357	6.892	2.757
	(1.902)	(3.418)	(1.462)	(1.818)	(2.298)
Optimal capital p.c.	7.752	13.83	5.770	6.975	3.638
	(1.597)	(3.189)	(1.295)	(1.672)	(1.973)
State FE	YES	NO	YES	NO	YES
Counties	1629	1629	1628	1628	1536
Excess marg. eff.	0.0230	0.0410	0.0182	0.0234	0.00938
p-val excess	0.00129	0.00127	0.00114	0.000748	0.240
Opt. marg. eff.	0.0264	0.0470	0.0196	0.0237	0.0124
p-val opt.	3.81e-05	0.000120	0.000118	0.000204	0.0755
p-val opt. = excess	0.411	0.268	0.675	0.941	0.420
Imputations	100	100	100	100	100

Notes: The marginal effect for both optimum and excess is the effect of increasing capital per capita by $50,000 divided by the the average 1880 county population, which is used for comparison to earlier estimates. The marginal effect of adding capital to a county is either the optimal or the excess marginal effect, not the sum. The "p-val opt. = excess" tests whether the two effects are the same (taking into account imputations). Income per person in 1970 and county vector files comes from census data and from Minnesota Population Center (2004). Uses all rural counties constrained at some point from 1870-1900, except those in Connecticut, Delaware, Massachusetts, Rhode Island, and Vermont. These states are well banked and small, and so have few counties which are constrained and so can only be included in some imputations. The imputations of optimal capital use the additional information from 1902. Errors are allowed to clustered at the state level. The window size is 15. Capital stock is measured in 1000's of dollars. See appendix for estimation details on the division between optimal and excess capital.

Figure 1: Growth of National and State Banks 1860-1914



Notes:Series are from White (1983, pp. 12–13), but originally come from Barnett (1911) and the Annual Report of the Comptroller of the Currency from various years. The number of state banks is from the Comptroller's series, and likely understates the number of state banks after 1887 by around 100 compared to the Barnett count which includes additional state banks. The number of state banks in the Comptroller series is nearly 400 below the Barnett series in 1886, but increases sharply to 1887. The difference appears to be a large effort by the Comptroller to obtain better information on state banks (see Annual Report of the Comptroller of the Currency 1887, p. 38).

Figure 2: Population and national banks in 1870

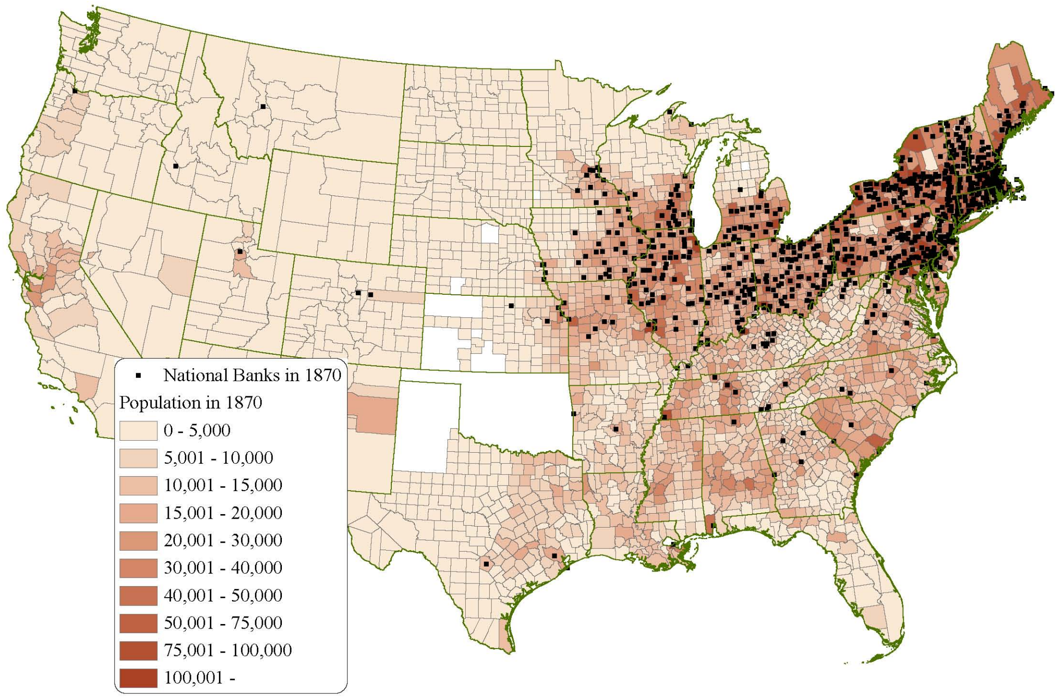


Figure 3: Population and national banks in 1880

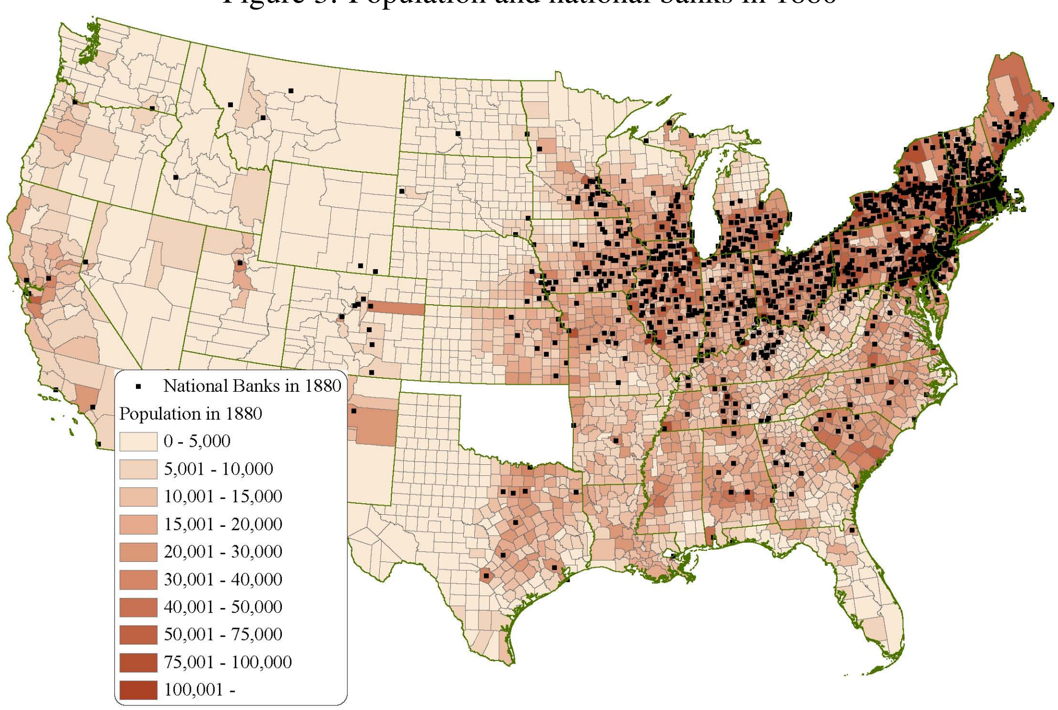


Figure 4: Population and national banks in 1890

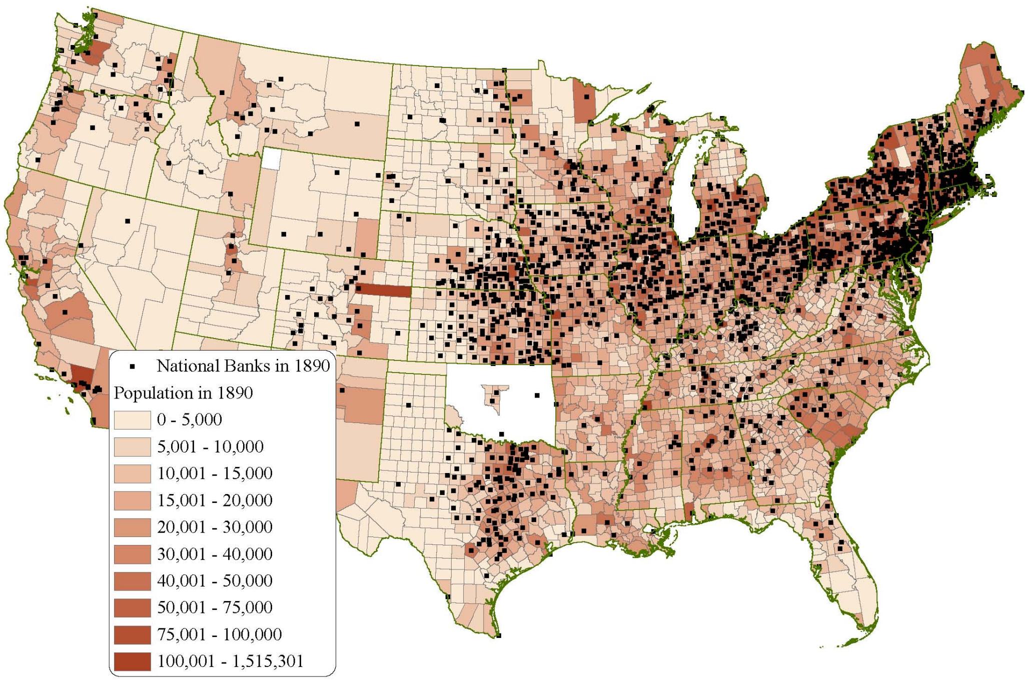
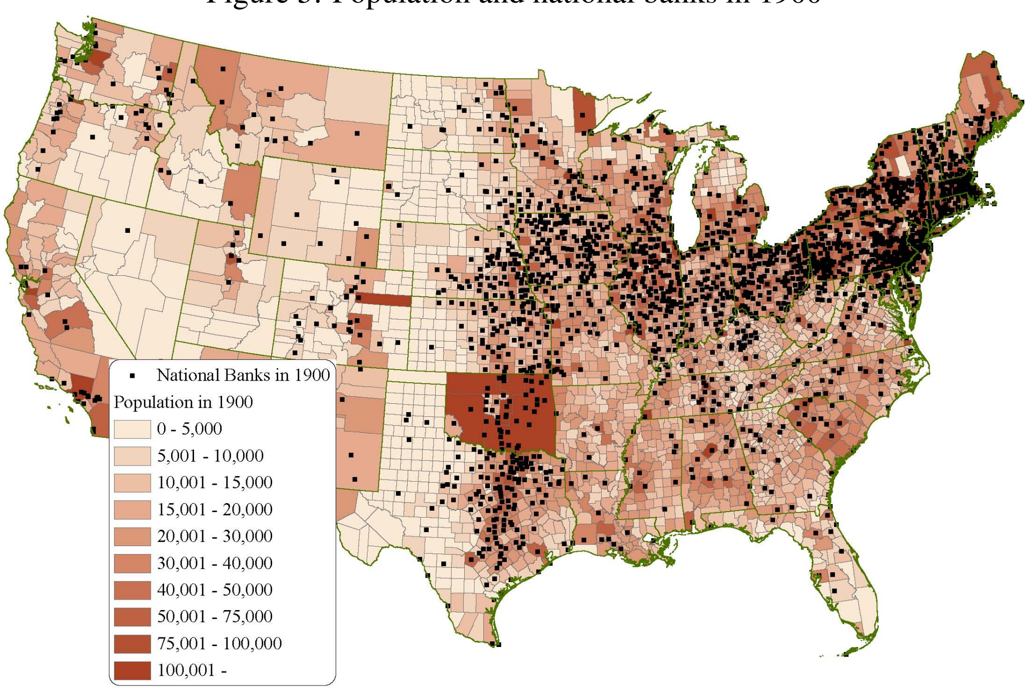


Figure 5: Population and national banks in 1900



Sources for figures 2 through 5: the national bank accounts in 1870, 1880, 1890 and 1900; Minnesota Population Center (2004) for the shapefiles; and Haines and ICPSR (2010) for the census populations. See appendix D.

Figure 6: Possible profit functions, and the observed capital

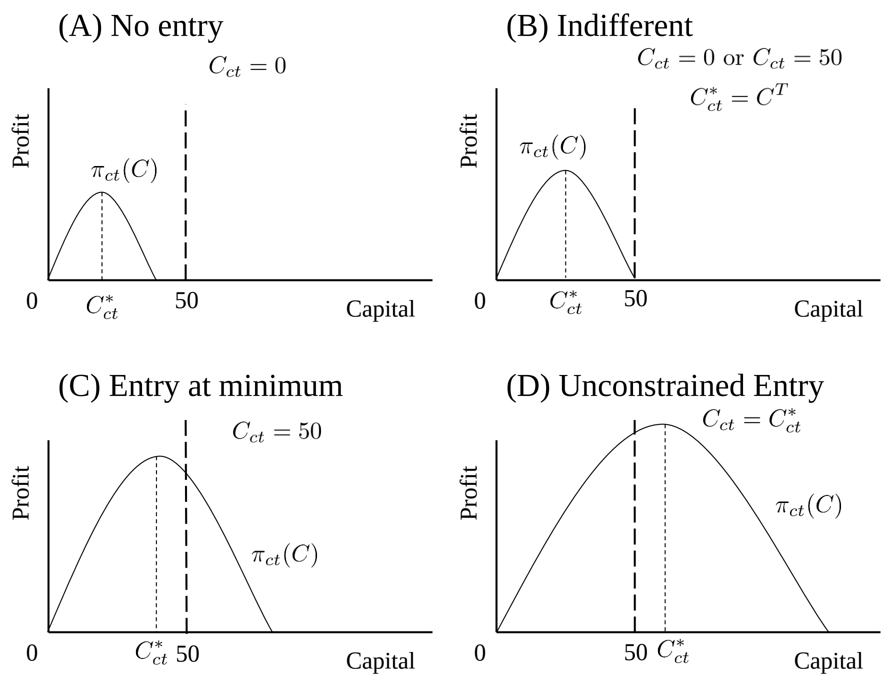


Figure 7: How observed and excess capital change with optimal capital