Journal of Financial Economics 157 (2024) 103868

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/finec

Real effects of supplying safe private money

Chenzi Xu [a,b,c,*], He Yang [d,1]

[a] Stanford Graduate School of Business, Stanford University, USA
[b] National Bureau of Economics Research (NBER), USA
[c] Centre for Economic Policy Research (CEPR), UK
[d] Amazon, USA

ARTICLE INFO

Dataset link: Real Effects of Supplying Safe Private Money (Original data)

JEL classification:
E42
E51
N11
N21

Keywords:
Private money
Bank notes
Market access
Safe assets
Payments
Currency

ABSTRACT

Privately issued money often bears default risk, which creates transaction frictions when used as a medium of exchange. The late 19th century US provides a unique context to evaluate the real effects of supplying a new type of money that is safe from default. We measure the local change in "monetary" transaction frictions with a market access approach derived from general equilibrium trade theory. Consistent with theories hypothesizing that lowering transaction frictions benefits the traded and inputs-intensive sectors, we find an increase in traded goods production, in the share of manufacturing output and employment, and in innovation.

1. Introduction

A predominance of private monies may introduce consumer protection and financial stability risks because of their potential volatility and the risk of run-like behavior. Indeed, the period in the nineteenth century when there was active competition among issuers of private paper banknotes in the United States is now notorious for inefficiency, fraud, and instability in the payments system.

–Lael Brainard, Member of the Federal Reserve Board of Governors (2021)

All forms of money are a liability of the issuer, and their usefulness as mediums of exchange depends on the confidence users have in the value of the assets backing the liability and/or the willingness of others to accept the liability as payment. Privately created monies often vary in their safety and can be prone to large devaluations when users receive better information about, or lose confidence in, the value of the underlying assets.

While private money can provide payment services in certain markets and generate significant seigniorage profits for issuers and early adopters, information frictions make it costly to verify their default risks. These default risks can lead to large "monetary" transaction costs, which can be particularly high at longer distances as it becomes more costly to acquire information. In contrast, fully backed safe monies have no default risk and therefore have a "no questions asked" quality that makes these debt instruments a frictionless medium of exchange (Holmström, 2015), which can foster economic growth by easing transaction costs.

Empirical evidence on the extent to which unsafe private monies hinder econonomic transactions and weigh on economic growth is scarce. This paper provides a novel analysis of the real benefits of reducing monetary transaction frictions in the economy. We study how supplying a new form of safe money changed the geography of monetary transaction frictions in the US from 1870 to 1880, and we show that it had a positive impact on the evolution of trade intensity and local economic development from 1880 to 1890.

This historical context is particularly suitable for our question because the US was a large and developed economy, but unlike other early

* Corresponding author.
E-mail address: chenzixu@stanford.edu (C. Xu).
[1] Work done outside of employment at Amazon.

https://doi.org/10.1016/j.jfineco.2024.103868
Received 25 March 2022; Received in revised form 6 May 2024; Accepted 8 May 2024
Available online 21 May 2024

industrial nations such as the UK, France, and Germany, it operated under a fragmented monetary system without a central bank responsible for maintaining a common, uniform currency. In this setting, the money supply consisted of notes issued by thousands of individual local private banks, backed by bank asset portfolios with little or even no regulatory oversight, and with low recovery rates following bank failures (Calomiris and Gorton, 1991). Private bank notes were not legal tender and they were only redeemable into specie at the issuing location. As a result, they bore both redemption risks (the possibility of not receiving the full face value of a note in legal tender or specie) and redemption costs (the costs of transporting a note to its issuer), which together manifested in the way notes traded at a discount from their face values across town borders and at higher distances (Gorton, 1999; Ales et al., 2008). Both contemporaries and historians have remarked on the frictions created by this fragmented monetary system, which lacked a sufficient supply of safe monies to sustain the transactional needs of the economy (e.g., Knox, 1900; Cagan, 1963).

The National Banking Act of 1864 introduced a new type of federally regulated “national bank” whose notes differed from existing “state bank” notes in two key ways. First, these notes were fully backed by federal government debt, which severed the link between a bank’s idiosyncratic portfolio credit risk and the real value of its notes.[2] Second, national bank notes were regulatorily required to be accepted at par by all other national banks around the country. Transactions inside the network could therefore always occur within a common national bank currency at zero discount, even if the issuing banks differed. Outside the network, even transactions conducted in national bank notes faced redemption costs. These differences made using national bank notes especially advantageous for long-distance transactions. While there were alternative payment mediums, such as specie and checks issued from bank deposit accounts, the former was costly to ship and insure, and the latter were even more informationally-sensitive because they depended on the credit-worthiness of individuals and thus were also more frictional at longer distances.

While the Act itself passed at the federal level, there was significant geographic variation in where national banks were created and hence in the overall network configuration of safe note supply and circulation. A change in the geographic distribution of national banks altered the monetary frictions facing a local economy and could reduce transaction costs via two main channels. First, it eliminated bilateral monetary transaction costs between any two locations in the country that both had national banks.[3] Second, for locations that remained outside the network, it potentially altered the lowest-cost route to a national bank that supplied and redeemed safe notes, which changed the dependence on alternatives. In conjunction, being closer to the national bank network lowered all monetary transaction costs for a location, but particularly so for long-distance transactions that would otherwise be more exposed to the frictions that made alternative transaction mediums costly to use.

In order to proceed with our empirical analysis, we embed monetary transaction costs into a model of multiregional trade that produces a measure of a local economy’s trade potential, which we call “monetary” market access (MMA). This measure captures, in reduced form, the overall change in a location’s trade potential as a function of all of the changes in trade frictions and market sizes that it faces in the entire network. For any given location, there are two sources of changes in its MMA: the first is direct changes in its own access (because a national bank enters the town), and the second is changes elsewhere in the network.

In order to provide causal evidence on the impact of changing monetary transaction costs, we employ two primary approaches. Our first approach instruments for the component of the change in MMA that comes from a location gaining its own national bank and joining the network for the first time. The instrument uses plausibly exogenous variation in the likelihood that a national bank is established locally, coming from regulatory capital requirements that were based on town population cut-offs. Banks established in towns with population below 6,000 needed to raise $50,000 of equity capital while banks in towns with population above 6,000 (and fewer than 50,000) were required to raise twice as much. The discontinuous jump in the implied equity capital per capita meant that towns just below the population cut-off faced significantly lower entry costs for establishing a national bank. These town population requirements were binding since banks were not allowed to branch, the capital had to be raised upfront, and there were strict residency requirements for bank directors.[4] These features of the regulatory environment create a plausibly exogenous entry cost for new national banks in towns near the population cut-off and hence for direct access to a supplier and redeemer of the safe currency in a location.

Our second approach measures the change in MMA in a location based solely on changes elsewhere in the network. By construction, this approach generates variation in MMA that is exogenous to changes in the location itself, including for instance the change in credit supply after the creation of a national bank. We construct this measure by holding fixed the national bank network at the beginning of the sample period for the focal location, and we only allow MMA to change from indirect exposure to changes in the rest of the network. This strategy allows us to remove all unobserved local shocks that might affect the outcomes of interest such as changes to local productivity or credit. We provide several measures at varying distances from a focal location, including within the county, and within radiuses of 25, 50, and 100 miles.

Together, these two approaches provide complementary evidence on the causal impact of changing monetary frictions. In order to maintain comparability between our two sets of empirical approaches, our analysis focuses on the sample of towns where we can identify the effect of gaining a national bank for the first time. First, we filter on towns that have population below 6,000 in 1870 (the first decadal census after the passage of the Act) so that they all face the same low capital requirement initially. Second, we impose that they do not have a national bank as of 1875 and therefore had no prior direct access to a local supply of safe bank notes.[5] Third, we require that their populations were between 4,000 and 8,000 in 1880. Choosing a small population bandwidth of 4,000 to 8,000 in the subsequent decade allows us to limit our sample to towns that are likely to be similar in both observable and unobservable characteristics in 1870. We show that pre-period observable characteristics were not significantly different between the towns with population above 6,000 versus those below 6,000 population in 1880, both conditionally and unconditionally. Within this set of towns, some crossed the 6,000 population cutoff in the 1880 census, which doubled the entry cost for a national bank. We use crossing this threshold as an exogenous shifter for direct access to safe private money.

[2] It did so by mandating that all national bank notes had to be backed 110% in the market value of federal government bonds, and it established structured procedures for insolvency and receivership that were designed to prioritize note holders against any losses, which eliminated default risk on notes.

[3] National banks were also more efficient at providing other forms of payment services that operated through the network because the Act included other regulatory changes such as a pyramidal structure in reserve deposits and a formalized clearing system, which strengthened their interbank connections and benefited check clearing among national banks relative to state banks.

[4] The non-branching rule meant banks could not be created in a low-capital environment while locating its business in more populous environments elsewhere. The upfront capital and residency requirement made it difficult to simply raise the required equity outside of a town’s borders or at a later date.

[5] The analysis is not sensitive to choosing 1875 as the first year in the 1870s that a town has a national bank.

For this first approach, we assess the first stage effect of crossing the population threshold on the change in MMA from 1870 to 1880 in a two-period difference-in-differences estimation. To strengthen the empirical strategy, we control for a town's growth trajectory with the population change from 1870 to 1880. This control variable absorbs differences across towns in 1870 that impact their population change. We also control for a town's financial development and physical trade costs in the pre-period for the same reason. The identifying assumption is that conditional on controls, there were no concurrent shocks to places just below the population cut-off that would have caused their outcomes to be systematically different after 1880. We find a strong first stage: being below the population cut-off in 1880 is associated with a 62% increase in monetary market access (F-stat of 9.6) by the middle of the decade relative to the previous decade.[6]

Our analysis of real variables is guided by the intuition, formalized by trade theory, that reducing the trade costs associated with moving a good from the producer to consumer will lower the good's final price, which will raise its competitiveness in the consumer's market, thereby leading to increased production at its origin. Since monetary transaction frictions outside the national bank network are linked to frictions that increase with distance, reducing these frictions is analogous to reducing cross-border trade costs. We have three main sets of results.

First, we show that monetary transaction frictions have a meaningful impact on the real economy as reducing these frictions leads to large and statistically significant growth in the production of tradable goods overall, measured by agricultural and manufacturing goods per capita. We find that a one standard deviation increase in MMA raises tradable production by 15.1% and 6.2%, using the instrumented change in joining the network and the rest-of-network changes, respectively. These results are robust to a number of alternative specifications including different configurations of control variables and construction of standard errors.

We next assess heterogeneity across goods by focusing on the agricultural sector, where we are able to observe the composition of goods produced. While all agricultural goods can be traded, only some are part of a national commodities market. By definition, the final price of such commodities depends more on the national market, and therefore reductions in local trade frictions pass more fully on to producers in the form of higher profits, which incentivize agricultural producers to tilt their production toward these goods. We find a 2.3% increase in the share of production of commodity goods for a one standard deviation increase in MMA.

Second, we find that the growth in the traded sector is driven by growth in the manufacturing sector, and that the economy becomes more manufacturing-intensive overall. This heterogeneity is consistent with theories pointing to structural transformation arising from relative changes in final output prices leading to sectoral reallocation of factors (e.g., Herrendorf et al., 2014). The share of manufacturing in overall production and employment both grow by 8.4% relative to their pre-shock means from a one standard deviation change in MMA.

We also find that the growth in manufacturing is driven by growth in inputs usage. Reducing frictions for sourcing inputs can increase the quantity of inputs and allow firms to source them more broadly and therefore also improve the match quality of the inputs they use. This "input sourcing" channel can raise output without requiring credit expansion as importing better inputs makes firms more productive (Goldberg et al., 2010). We find that the growth in employment combined with an increase in inputs appear to drive the output effects.

While we are not able to fully rule out the possibility that these effects are affected by the standard bank credit channel in trade (e.g., Paravisini et al., 2015; Xu, 2022), our second approach excludes variation from the endogenous (location's own national bank) component and its correlated effects, including the impact of national bank entry on the local credit supply. In addition, we can control for the volume of lending from national banks as a proxy for this channel in all of our specifications and show that it does not significantly change the estimated coefficients.

In our third set of results, we assess the link between monetary market access and long-run innovation and output. Schmookler (1954) posits that the changes in patenting activity in the US in the seven decades following the Civil War may have been due to expansions in the markets firms could reach (i.e., their market access) such that it was more profitable for firms to invest in productivity-enhancing technologies. We assess this hypothesis with the number of patents granted, and we find that a one standard deviation increase in the change in MMA leads to 1.3 times more the cumulative amount of innovation measured relative to before the Civil War. The growth in innovation is also consistent with our findings that manufacturing output remained elevated until at least 1900.[7]

Our paper primarily relates to the theoretical and empirical work on the optimal design and regulation of digital private monies, much of which has drawn comparisons between stablecoin design and traditional banking (Eichengreen, 2019; Catalini and de Gortari, 2021; Catalini and Shah, 2021; Gorton and Zhang, 2021; Gorton et al., 2022b). These papers highlight various mechanisms by which pegged private digital currencies (stablecoins) can achieve their eponymous stability, most of which involve maintaining sufficient safe reserves like US Treasuries as the backing asset, as in this historical context. To the best of our knowledge, this paper is the first to provide empirical evidence that improving the safety of privately issued money, in this case by regulating the quantity and quality of the assets backing the monetary liability, positively affects real economic activity.

More generally, our analysis is related to the literature in which assets that are sufficiently safe may provide monetary services by acting as a medium of exchange, as in Coppola et al. (2023). There is a separate literature that microfounds the phenomenon of fiat money circulating as the medium of exchange as in Kiyotaki and Wright (1989), and work showing that safe monetary assets can also transition into being a fiat money as in Bolt et al. (2023). In our context, the redeemability of a private bank note for a guaranteed amount is its defining feature, both theoretically and empirically.

Private bank notes are only one form of short-term bank debt, and while the large literature on bank debt illiquidity has mostly focused on their fire-sale externalities and their role in increasing financial fragility (Diamond and Dybvig, 1983; Stein, 2012; Admati and Hellwig, 2014), this paper studies a context in which a change in the regulatory environment ensures that they are truly safe with full liquidity coverage, which creates a Gorton and Pennacchi (1990)-type transactions medium. The national bank notes that are fully backed by government debt in our context take the form of the optimal contract that removes any discount from the debt's face value, thereby making transactions efficient (Dang et al., 2017). Hence, we provide evidence for a subnational monetary channel in the tradition of Friedman and Schwartz (1965) and Romer and Romer (1989).

While the literature on the bank credit channel has demonstrated the importance of credit supply for international trade in both modern and historical contexts (e.g., Amiti and Weinstein, 2011; Paravisini et al., 2015, 2023; Xu, 2022; Matray et al., 2023), our paper relates to work documenting that the presence of local banks impacts growth and labor market outcomes in modern developing countries (e.g., Fonseca and Matray, 2024; Ji et al., 2023). We focus, like Sarkisyan (2023), on the benefits of accessing a safe medium of exchange.

[6] Our instrument also predicts a 52% increase in the direct likelihood of gaining a national bank locally.

[7] The county-level data on manufacturing do not exist for 1910, and the levels are no longer significantly different by 1920. Given the large amount of changes that likely occurred in WWI and after the establishment of the Federal Reserve system, we end our analysis at 1900.

More generally, this paper adds to the literature on the historical determinants of economic growth in the United States and contributes to the better understanding of how the banking sector has shaped the geography of economic activity in the late 19th century (Friedman and Schwartz, 1965; Rousseau and Wachtel, 1998; Rousseau and Sylla, 2005; Landon-Lane and Rockoff, 2007). Our results are consistent with Sylla (1969, 1982), which argue that the National Banking Act's high regulatory capital requirements held back economic development by failing to expand the bank note supply sufficiently, and Cagan (1963) that attributes economic growth in the late 19th century to currency stability. Fulford (2015) uses a non-parametric approach to estimate the average impact of adding a marginal national bank to counties over three decades in this period and similarly finds large positive effects on total production per capita. This paper hypothesizes that monetary transaction frictions are a channel for those effects, and it follows Donaldson and Hornbeck (2016) in conceptualizing and estimating the cost of those frictions through a reduced form market access approach. Our focus on the change in the monetary system shows the empirical relevance of those frictions to the real traded sector.

The rest of the paper is organized as follows. Section 2 discusses the institutional details of the historical banking sector and monetary system. Section 3 describes the data, and Section 4 presents the empirical framework with monetary market access. Section 5 explains the sample construction, measurement, and empirical strategies. Section 6 presents the empirical the effect of national bank entry on real economic outcomes, and Section 7 concludes.

2. Historical context of frictions in private money

Until the establishment of the Federal Reserve system in 1913, the United States had a fragmented monetary system with thousands of different circulating mediums of exchange. One of the predominant types was paper notes printed by individual private banks. These bank notes were liabilities issued against the bank's assets, which could be redeemed at the issuing bank for specie (when the US was on a metallic standard) or for legal tender. Private bank notes were not themselves legal tender and could therefore be refused for payment completely or could be accepted only at a discount from face value. The guaranteed consumption value of a note therefore depended on the amount of specie or legal tender the note's owner could receive for it.

The discount on the face value of a private bank note priced two main frictions: the first was uncertainty about the value of the underlying assets backing the note—i.e., its redemption value once it was presented at the issuing bank ("redemption risks"), and the second was the transportation or time costs of returning the note to its issuer for redemption in the first place ("redemption costs").

2.1. Unsafe bank notes during the Free Banking Era

During the first half of the 19th century, in the period known as the Free Banking Era, there was only one type of private bank note—the "state bank" note—for which both redemption risks and redemption costs were high. These frictions primarily arose from the structure of the banking system in which all regulatory oversight was delegated to individual states, and states often allowed banks to operate with few or no restrictions.

Redemption risks and costs. The lack of regulatory oversight led to a risky banking sector that was prone to failures and panics for three reasons. First, there were few limitations on asset composition and low minimum capital requirements, so banks often had highly levered, risky portfolios with little equity buffer. Second, banks were prevented from branching so they had concentrated geographic exposure, and local shocks easily led to bank failures (Bordo, 1998). Third, there was neither a formal system of interbank lending nor a lender of last resort (White, 1982). In combination, bank runs and failures were frequent.[8] The bank notes issued by failed banks could not be redeemed and were valueless.[9]

Even when the bank itself was solvent, its notes could be risky and could have uncertain redemption value for several reasons. First, each bank designed its own notes, so there were thousands of types of notes of different sizes, styles, and nominal values in circulation, which made it difficult to distinguish genuine notes from counterfeits. Second, conditional on a note being genuine, the lack of regulatory oversight allowed banks to issue notes beyond their redemption capabilities. Therefore, even if the underlying assets were safe and retained their full value, notes might only be partially collateralized.[10] Third, in practice, notes tended to be backed by assets such as the bank's loan portfolio, state government bonds, and equities, all of whose values fluctuated. Significant drops in the collateral led to note redemption runs during which the notes were heavily discounted (Rolnick and Weber, 1982).[11]

Finally, in addition to a note's fundamental redemption risk, there was a cost to returning it back to its point of origin where it could be redeemed, deposited, or traded for consumption. On average, these costs increased with distance (Gorton, 1999).

Bank note discounts. Redemption risks associated with counterfeit and default, and redemption costs associated with returning a note to its issuing bank were priced as a discount on the face value of a note. In practice, assessing a bank note's underlying risk was costly and information-intensive so transactions in a local economy only occurred in local bank notes (Appleton, 1831). Specialized note brokers intermediated the currency market where they charged a commission for their expertise and priced the discount on a note based on its underlying qualities. Publications such as the "Bank Note Reporters and Counterfeit Detectors" helped brokers to ascertain the authenticity of unfamiliar notes.[12] Newspapers published the average discount rates, which varied by bank and over time (Gorton, 1999; Gorton et al., 2022b).

Appendix Figure A.1 plots average discounts of state bank notes in several states relative to banks in Philadelphia. The volatility in the discounts in part reflects significant underlying time-varying idiosyncratic redemption risks.[13] Rockoff (1975) estimates that losses on notes due to bank failures ranged from 7% in Indiana to 63% in Minnesota. They also reflect the variation in redemption costs: while the states in the northeast closer to Philadelphia (Figure A.1a) had discounts up to 10%, the ones more distant from Philadelphia (Figure A.1b) had discounts as high as 80%.

[8] Between 1875 and 1890, 2.5% of state banks failed.

[9] John Jay Knox, the Comptroller of the Currency from 1872 to 1884 wrote, "The cost of exchange was also immensely increased by the lack of uniformity in the credit of the banks in the several states at the commercial centers", (Knox, 1900, p. 315).

[10] Milton Friedman referred to the phenomenon of banks over-inflating their currency to the point of not being able to meet redemption as "wildcat banking", a term that is now frequently applied to the antebellum period in American banking.

[11] "It is possible that the loss to note holders was about five percent per annum...the runs on the banks were not made by the depositors, but by note holders", (Knox, 1900, p. 315).

[12] Figure A.2a displays an example of a private bank note from Massachusetts with face value $20 where the name and location of the issuing bank is prominently displayed. Figure A.2b shows the written description for the same bank's notes in a printed "counterfeit detector", where the $20 bill is described in the bottom left corner.

[13] For example, Illinois banks committed over 5 million dollars between 1836 and 1842 to building a canal that would connect the Illinois River and Lake Michigan. However, this investment completely drained state funds and caused a wave of state bank failures in Illinois. As a result, the relative discount of Illinois state bank notes averaged at around 70% in 1842, compared to about 15% in the previous year.

Despite these costs, bank notes survived and were widely used because there were few alternatives. There was little circulating specie in the payments system, and moreover payment in specie was expensive because of physical transportation and insurance costs.[14] Coins, primarily Spanish and US silver dollars, were scarce, in an array of unwieldy denominations, and were subject to debasement and deterioration (Greenfield and Rockoff, 1995; Ware, 1990). Greenbacks were a federally-issued fiat currency that traded at a discount for much of this period and also in limited supply.[15]

Demand for reducing bank note transaction costs. The limited supply of safe assets that could act as a frictionless medium of exchange created large transaction costs in exchange and trade, and these costs drew the attention of lawmakers and policymakers.[16] In 1863, Senator John Sherman from Ohio argued in Congress for the passage of the National Banking Act explicitly in terms of improving bank notes as a medium of exchange:

> This currency will be uniform. It will be printed by the United States. It will be of uniform size, shape, and form; so that a bank bill issued in the State of Maine will be current in California; a bank bill issued in Ohio will be current wherever our Government currency goes at all; and a bank bill issued in the State of Connecticut will be freely taken in Iowa or anywhere else. There is *no limit to its convertibility*. It will be of uniform value throughout the United States. I have no doubt these United States notes will, in the end, be taken as the Bank of England note now is all over the world, as a medium, and a standard medium of exchange [...] They will be safe; they will be uniform; they will be convertible. Those are all the requisites that are necessary for any system of currency or exchange.[17]

The transactional costs of using state bank notes, together with the need to raise funds for the federal government during the Civil War, eventually led to the passage of the National Banking Act.

2.2. National banking era introduction of safe bank notes

The National Banking Act aimed to stabilize the banking system and to create a network of national banks that were subject to federal regulations. While state bank notes had redemption risks associated with counterfeit and default and redemption costs associated with returning a note to its issuing bank, national bank notes eliminated or significantly reduced both types of frictions.

No redemption risk. National banks were regulated by the newly created Office of the Comptroller of the Currency (OCC), and their notes were designed to be standardized, uniform, and safe. Therefore, although they continued to be privately issued by individual national banks, they bore no idiosyncratic redemption risk.

First, notes were overcollateralized by US Treasury bonds where banks were only allowed to issue 90% of the market value of government bonds on their balance sheets, and losses in the collateral's market value had to be supplemented with additional bonds.[18] This overcollateralization strictly constrained the note issuance.[19] However, backing the currency in this way eliminated transaction frictions arising from asymmetric information about a bank's solvency (Dang et al., 2017).

Second, in the case of a national bank failure, the OCC oversaw the liquidation process that prioritized note holders from losses, and its success was documented in the annual reports of the OCC, which reported that there were no losses from bank note defaults in this period. Third, bank note redemption was a condition of operation, and national banks that refused to redeem their notes were immediately put into receivership.

In addition to structuring national bank notes in a way that guaranteed them to have full recovery rates in the case of a bank failure, the banks were also much less likely to fail in the first place. Both equity capital and the banks' asset portfolios faced stricter regulations that were designed to prevent runs and to improve financial stability. Capital requirements were also much tighter: they were higher; 50% had to be raised before business commenced and the rest completed within five months; and they could only be met in cities with sufficient local capital since a national bank had to have at least 5 directors, and at least 75% of them must have resided locally.

To reduce risk-taking, national banks also faced limits on the long-term loans they could make, were not allowed to take land as collateral, and were subject to detailed annual supervision. These restrictions limited the provision of long-term credit and credit to the agricultural sector where collateral most often took the form of farmland (Knox, 1900). The average national bank failure rate was approximately 0.25% in this period, which was 10% of the state bank failure rate.

Lower redemption costs. The second key component of the legislation was that all national banks were required to accept the notes of all other national banks at par.[20] National bank notes could also immediately be redeemed for specie or legal tender at the US Treasury in addition to at the issuing bank. In combination with the notes' full backing, this second requirement meant that within the national bank network, the real value of a national bank note was guaranteed for both consumption and savings. They could therefore circulate as a medium of exchange or be deposited as savings without a discount on their face value.

[14] Corroborating evidence on the lack of gold in circulation comes from the national "gold" banks that could issue notes redeemable in gold. All of these, except for a few in California, failed because they could not obtain the gold necessary to back their notes (Greenfield and Rockoff, 1990). While the US Mint produced more uniform coins, these were primarily used in international trade rather than domestically (Carothers, 1930).

[15] Greenbacks were issued by the federal government in 1862 after suspending specie redemption for bank notes in 1861. In a demonstration of Gresham's Law, greenbacks immediately pushed specie out of circulation, and gold did not return to circulation in the United States outside of California (Mitchell, 1903; Greenfield and Rockoff, 1995). Only after the US retired a large fraction of greenbacks and returned to the gold standard in 1878 did greenbacks trade at par with their face value. However, there was a nominal cap on their circulation that was not adjusted for inflation or growth so greenbacks represented a diminishing share of the overall money supply each year.

[16] See Appendix B for contemporary examples of how the uncertain value of state bank notes led to legal disputes and inconvenience in exchange.

[17] Senate floor, February 10, 1863; http://www.yamaguchy.com/library/spaulding/sherman63.html.

[18] This form of narrow banking was designed in part to create a demand for US Treasury debt that also helped to finance the federal government (Gorton et al., 2022a). Until 1882, there was an additional constraint that limited banks from issuing notes beyond the value of 70%–90% of their equity capital (the limit was higher at lower levels of capital), regardless of the value of bonds the banks held as assets. The full text of the original legislation related to each of these legislations is in Appendix B.

[19] Indeed, there is well-documented evidence by both contemporaries and historians that there was an overall *shortage* of national bank notes in the decades after the Act was passed (e.g., Bell, 1912; Friedman and Schwartz, 1965; Gorton et al., 2022a), in part because of the low supplies of government debt. In the 1890s, this shortage was so severe that the personal liabilities of firms and individuals were circulated as a medium of exchange (Warner, 1895).

[20] Although national banks were not required to *redeem* other banks' notes in specie or legal tender, in practice it was straightforward to deposit another bank's notes, withdraw the bank's own notes, and redeem those.

While this part of the national bank legislation eliminated redemption costs within the network, it did not do so outside of it. State banks had no obligation to accept national bank notes at par, and indeed all banks had strong incentives not to do so. National bank notes were a less useful asset than legal tender or specie for banks because they could not be used to raise bank capital, pay certain taxes, or buy government bonds, and they did not count toward bank reserve requirements. State banks could embed a preference for legal tender by, for example, extending loans in national bank notes but requiring that the loan be repaid in legal tender.[21]

Other bank liabilities as payments. In addition to notes, both state and national banks provided check clearing services as a transactions medium. However, checks were sensitive to individual depositors' idiosyncratic risks and therefore even more information-sensitive. As in the case of notes, banks were only obligated to redeem checks at par when they were presented at the original bank's office.[22]

The National Banking Act also created a pyramidal system of reserves and a centralized clearing house for national banks that aimed to reduce counterparty risk within the system. As a result, checking accounts exhibited the same asymmetry in monetary transaction frictions as bank notes.

2.3. Geography of monetary transaction frictions

The geographic distribution of private bank note circulation is governed by fundamental restrictions and by the monetary benefits they provide. In terms of fundamental restrictions, all private bank notes could only be created by their issuer. However, national bank notes were redeemable by any other national bank while state bank notes were only redeemable by the original issuing bank.

In terms of their monetary benefits, national and state bank notes differed in their usefulness as both mediums of exchange and as savings vehicles. By regulation, national bank notes were a frictionless medium of exchange within the national bank network because both sellers and buyers could use the par value of the notes for consumption and savings. In contrast, state bank notes were only frictionless within their local economy because they faced redemption risks and costs at a distance.[23] As savings vehicles, national bank notes could be deposited at par in the network, but were disadvantaged outside the network.

These forces imply that national bank notes were most valuable inside the national bank network while state bank notes were most valuable within their local economy. Both types of notes would naturally tend to circulate where they were most valued. In conjunction with the well-documented evidence by both contemporaries and histories that there was an overall shortage of national bank notes, these forces imply that being closer to the national bank network improved access to its superior form of money.

[21] This arrangement was common, and it passed the redemption costs on to the depositor.

[22] Checks could only be cleared when both the status of the bank and of the personal account could be verified. In addition, banks lost reserves as soon as a check cleared, whereas notes could be used to settle transactions without immediate demand for reserve (Briones and Rockoff, 2005). Empirically, they only became common in long-distance wholesale trade at the turn of the twentieth century (Kinley, 1910; Preston, 1920; James and Weiman, 2010).

[23] The Annual Report of the Comptroller of the Currency in 1877 described the following way: "The notes which were returned from the commercial centers for redemption were readily paid out and circulated at home, and the demand for specie, wherever it existed, was almost entirely owing either to an excess of currency or to a want of confidence in the institutions which issued it".

2.4. Period of study

We study the impact of the introduction of national bank notes into local economies in the 1880s relative to the 1870s, and we examine outcomes in the 1890s relative to the 1880s. We focus on these decades for two reasons. First, the US monetary system underwent many adjustments in the decade after the Civil War, including the demonetization of silver in 1873 and a significant amendment to the National Banking Act in 1874. We therefore focus on changes in the national bank network beginning in 1875.

Second, our interest is on the impact of introducing a superior medium of exchange before it became the dominant currency. At the very beginning of this period, there were large changes in the banking sector overall (appendix Figure A.3 shows the evolution of the number and total assets of national banks and state banks from 1863 to 1900), but the decades from 1870 to 1890 feature a relatively steady proportion of state and national banks without large changes in their relative composition (Jaremski, 2014).

3. Data

We combine several newly collected and digitized historical datasets. Table 1 reports summary statistics in the 1870s.

Town populations: We manually collect town and city populations from the original reports of the 10th and 11th Decennial Censuses of Population that report information covering the decades 1860 to 1880 (Figure A.4a). These original PDFs allow us to locate towns with population below 2500 people, which are not in the digitized censuses.

Bank characteristics: We collect and digitize the balance sheets for all national banks from 1870 to 1890 from the Annual Report of the Comptroller of the Currency (Figure A.4b). These balance sheets report locations as well as the main asset and liabilities components such as their loans and bank note circulation. We collect the locations of non-national banks (state and private bankers) from *The Banker's Almanac and Register* of 1876 and 1885 (Figure A.4c).

Physical trade costs: We measure counties' physical trade costs with railroad access in 1875 and 1880 (Atack, 2016) and transportation market access in 1870 and 1880 (Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2021). The latter measure is constructed taking into account the entire transportation network including railroads, wagon roads, and waterways.

County outcomes: Our outcomes are from the Censuses of Agriculture and Manufacturing from 1870 to 1900. These Censuses provide us with population measures, inputs usage, capital, and production for both sectors in each county. For manufacturing, we are also able to observe employment, and in agriculture, we observe disaggregated production by crops. These censuses also report county-level populations by gender and age group. We use the adjusted populations from Hornbeck (2010) and Hornbeck and Rotemberg (2021) that align with county borders defined in 1870.

4. Empirical framework

In this historical setting where each bank note was a debt claim to specie or legal tender, monetary transaction frictions arose from transacting in unsafe debt that has redemption risk and costs. The theoretical framework that ties the supply of national bank notes to transaction frictions is developed in Coppola et al. (2023), which shows that increasing the safe supply of an asset increases its usefulness as a medium of exchange by increasing the probability that a two-sided transaction can occur in that common medium.

Empirically, the monetary transaction friction is the discount paid for using a particular medium of exchange in a bilateral transaction. It therefore acts as a standard trade cost. Consider the example of purchasing goods from New York in Philadelphia. In the case where the transaction is conducted in Philadelphia bank notes, the seller in

Table 1
Summary statistics.

	Mean	Med.	St. Dev.	N
Population 1870	3968.62	4096.00	1098.78	147
Population 1880	5059.28	4866.00	927.76	147
Railroads	4.35	4.00	2.69	147
MMA (log)	10.54	11.04	2.00	147
tMA (log)	15.83	16.00	0.49	147
State banks	0.65	0.00	0.96	147
Tradables production	893.18	808.28	453.28	147
Commodities share	0.89	0.89	0.09	147
Manu. production	662.73	492.02	500.87	147
Manu. inputs	399.47	289.39	325.56	147
Manu. production share	0.67	0.71	0.24	147
Manu. employment share	0.33	0.26	0.23	147

Notes: Table 1 reports the mean, median, and standard deviation of each location characteristics in the pre-entry period. The populations in 1870 and 1880 are at the town level, which is the relevant population for our instrument. The other measures are taken at the county level. The measure of railroads is from Atack (2016) and measured in 1876. State and other non-national bank financial institutions are obtained from the *Banker's Almanac and Register* of 1876, and the locations of national banks are from the *Annual Report of the Comptroller of the Currency* in 1874. We calculate monetary and transport market access using bilateral transport costs and monetary costs in 1870 and 1875, respectively. Manufacturing and agricultural production are from their respective censuses of 1870.

New York would incorporate the costs of converting the "foreign" note into local currency (the discount) in the final price, and the monetary transaction cost acts as a non-revenue generating tax on the final price of the good.[24]

In contrast, if the transaction occurs in a common currency that both the buyer and seller face no frictions in accessing to use for consumption or savings (i.e., they both belong to the national bank network), then there are no monetary transaction costs, and the final price of the good in Philadelphia is lower relative to a transaction conducted in a frictional medium. Following the introduction of the national bank network, transactions—particularly at long distances—experience a large decline in frictions. In aggregate, each location benefits from the total reduction of costs coming from all other possible locations it can transact with.

As in a general class of trade models (e.g., Eaton and Kortum, 2002), reducing the final price of a good in a destination makes it more competitive and raises its consumption. The change in demand then induces more production until firms reach a new optimal equilibrium at a higher level of production.

4.1. Market access with monetary transaction frictions

We quantify the monetary frictions that each location faces by adapting a model of multiregional trade with trade costs where we augment the standard measures of bilateral trade costs so that it combines physical trade costs, which we denote by ϕ, with monetary transaction frictions μ. Locations are denoted by o and d depending on whether they are the origin or destination of a good. Bilateral trade costs are defined in an "iceberg" way such that the final price of a good j that is traded from o to d has the following price: $p_{od}(j) = \tau_{od} p_{oo}(j)$ where p_{oo} refers to a non-traded good that is produced and consumed locally and $\tau_{od} \geq 1$ incorporates all bilateral trade costs.[25] Consumers have a CES utility function with elasticity of substitution σ. In combination with firm production, this set-up delivers a standard gravity equation for the patterns of trade across locations.[26] Reducing τ_{od} therefore reduces the final price $p_{od}(j)$ paid by consumers in d for good j, which makes it more competitive and (all else equal) raises its consumption.

The bilateral gravity equation can be transformed into a location-level "market access" measure that is a reduced-form expression capturing a location's overall propensity to trade, both as a consumer and a producer.[27] Intuitively, market access aggregates the potential value of trade with all destinations, where each destination is weighted by its proximity and size. A location's market access increases with proximity to and with the size of those markets. Changes in market access summarize the total direct and indirect effects on each location from changes in the full bilateral matrix of trade costs.

We construct the matrix of bilateral trade costs in the following way: physical trade costs ϕ_{od} are calculated by finding the lowest cost routes between any two locations from existing road, railroad, and waterway networks at each decade, which we take from Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2021). μ_{od} is the shortest *monetary* distance between two locations. By law, locations o and d inside the national bank network have a common currency and face no monetary transaction costs ($\mu_{od} = 1$). Outside of the network, the common currency is not locally supplied to both the buyer and the seller so one or both parties must pay redemption costs ($\mu_{od} > 1$) to convert into local currency or legal tender/specie. Redemption requires physically sending the note to its origin and therefore entails the physical transport costs along a route. We therefore proxy for μ_{od} with physical transport costs when a transaction occurs outside the network in order to approximate the cost of physically obtaining notes, the discount of converting non-local notes (which itself is a function of distance), and the information frictions from using alternative payment methods like checks or specie.[28] We find the lowest cost route for both trading partners to obtain a common currency, which can be national bank notes (where one or both parties travels to the nearest location with a national bank) or one of the local currencies (where one party covers the distance to access the other). Fig. 1 provides examples of the configurations of distances under each of these examples.

4.2. Measuring monetary market access

These components of physical and monetary trade costs generate a bilateral matrix of trade costs between all od pairs of locations and allow us to calculate monetary market access for any location, which we define as:

$$\text{MMA}_o = \sum_d \mu_{od}^{-\zeta} \phi_{od,1870}^{-\theta} N_d \quad (1)$$

where ζ and θ are the trade elasticities with respect to each cost μ (monetary) and ϕ (physical) respectively, and N is the size of the destination trade partner.

[24] For simplicity, in this example, the supplier pays for the conversion, although the incidence of the discount may be split in other ways with the buyer.

[25] Iceberg trade costs assume that a fraction $\tau - 1$ of goods goes towards the delivery cost. In order to receive one unit of a good, τ units must be shipped. They thus act as a non-revenue generating tax, and in our framework, it does not matter which party nominally pays the tax.

[26] The gravity result is very general across preferences, production functions, and trade costs. Costinot and Rodríguez-Clare (2014) derives the set of sufficient conditions which encompass many ways of modeling production including for instance, both perfect competition and monopolistic competition.

[27] Measuring the exact measure of market access in each location requires simultaneously solving a system of equations such that all prices clear all markets. However, Donaldson and Hornbeck (2016) shows that a reduced form approximation has a 99.99% correlation with the analytical solution. As a result, we also rely on the approximation.

[28] Note that even if physical money is not transported in these transactions, and other payment mediums like checks are used instead, the relative magnitudes of the frictions between being in the network versus outside of it would still hold because national bank notes were still superior to alternatives like checks. Empirically, we use the physical trade costs from Donaldson and Hornbeck (2016) to proxy for these physical transport costs. It is worth noting that those measures were constructed to capture the cost of transporting freight, which may be different from the cost of moving physical bank notes. However, measurement error in this proxy will primarily add noise to the estimated effects.

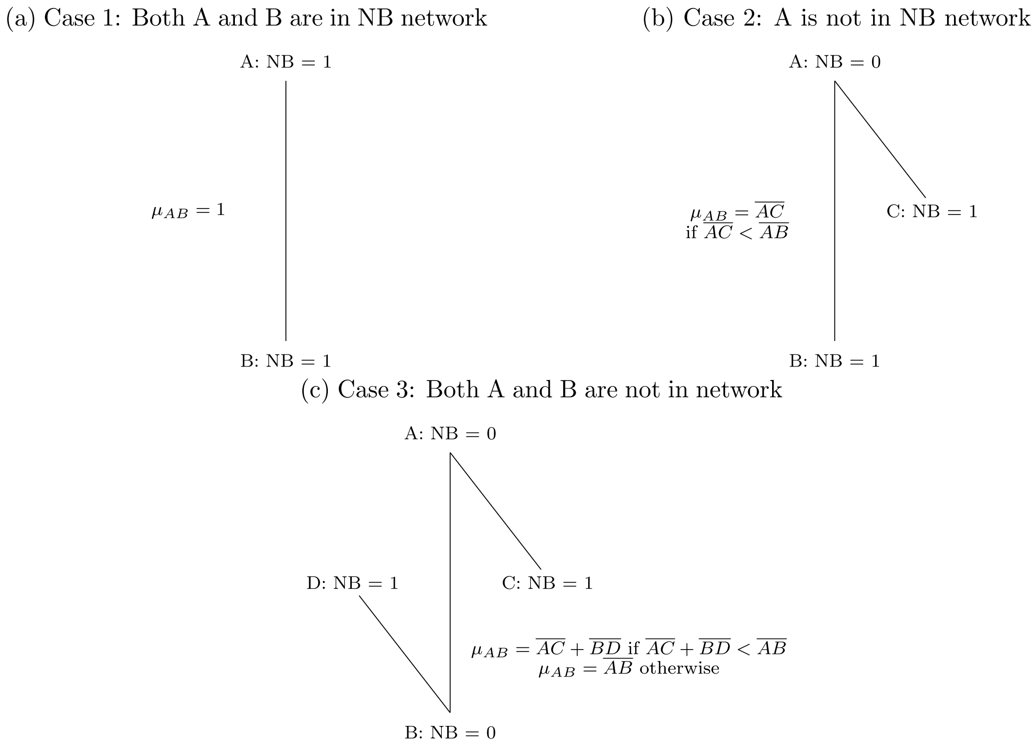

Fig. 1. Minimum Monetary transaction frictions between towns. *Notes:* Fig. 1 provides a visual example of the different cases for calculating the minimum monetary transaction frictions between towns. When both towns each have a national bank (NB = 1), then $\mu = 1$ (Case 1). When only one town is in the network (B: NB = 1) while the other is not (A: NB = 0) as in Case 2, there are two possibilities. If there is a third town (C: NB = 1) near town A with a national bank such that the distance between A and C ($\overline{AC}$) is less than the distance between A and B, then A can access the network via C and the transaction frictions are based on the distance between A and C. Otherwise, the monetary distance is $\overline{AB}$. Finally, in Case 3, both towns are not in the network and so there are two possibilities again: the first is that each town is close to a national bank and can therefore each access the network independently; the second is that one of the towns uses the currency of the other by traveling there.

We calculate MMA_o in two decades, the 1870s and 1880s. For the 1870s, we use the national bank network as of 1875, which we map in Fig. 2(a), and the transportation network and populations as of 1870. For the 1880s, we use the national bank network in 1885 but hold fixed the transportation network to its 1870 appearance (denoted by $\phi_{od,1870}$).[29] Doing so ensures that the changes in MMA across decades come from changes in the locations of national banks rather than the expansion of the physical transportation network. In our baseline, we choose a monetary trade elasticity of $\zeta = 1$ and an overall trade elasticity of $\theta = 4$.[30]

We map our calculated measure of monetary market access in 1875 in Fig. 2(b). As expected, monetary market access is highest in the northeast region where the national banking network is densest. It is also higher in areas with lower physical transport costs, such as those near major waterworks and railways.

In order to estimate the impact of increasing the supply of safe money, we use the changes in the national bank network (Fig. 2(c)) to calculate the changes in monetary market access (ΔMMA) from 1875 to 1885 (Fig. 2(d)). A location's MMA can increase because of a direct reduction in monetary transaction frictions due to gaining its own national bank or because of an indirect change in the national bank network elsewhere. These two components do not combine linearly since they depend on prices and market sizes that evolve jointly in general equilibrium.

The change in MMA in a location (Fig. 2(d)) therefore differs from a binary measure of the introduction of a national bank (Fig. 2(c)) in two key ways. First, it captures all the changes in transaction costs in the entire network that impacts a location's trade potential even if the location itself does not gain a national bank. Second, it captures the location's trade potential to all destinations in aggregate, weighing each destination's importance based on its size and proximity. These differences allow for heterogeneity in the degree of treatment: a location that is already close to a national bank prior to receiving its own national bank will not be impacted to the same degree as a location in a monetary desert. Similarly, a location that is close to many large markets will gain more from reducing transaction costs than a location whose trade potential is low.

For example, Fig. 2(d) shows that New Mexico experienced a very high change in monetary market access even though there was not a single national bank introduced in the state. In contrast, many national banks entered in the northeast region of the country, but the change in MMA in that region is small and very similar to the changes in the southeast where few national banks entered. These examples highlight that it is important to account for how the full structure of the network impacts an individual location's trade potential.

5. Empirical strategy

Our measure of the change in MMA includes two components: the (endogenous) direct component due to a location gaining a national bank and joining the network and the (exogenous) indirect component from changes elsewhere in the network. We therefore adopt two approaches to identify the relationship between changes in monetary market access and real outcomes. Our first approach instruments for the

[29] We map the locations of national banks in 1885 in appendix Figure A.6. Measures of MMA are very similar using 1870 and 1880 national bank locations instead of our mid-decade networks. MMA_{1880} is also very similar holding populations fixed to 1870 levels instead of using 1880 levels.

[30] We provide robustness checks based on a range of θ values, including $\theta = 1$ from recent quasi-experimental empirical work by Boehm et al. (2020), $\theta = 2$ from Hornbeck and Rotemberg (2021), and $\theta = 8$ from Donaldson and Hornbeck (2016).

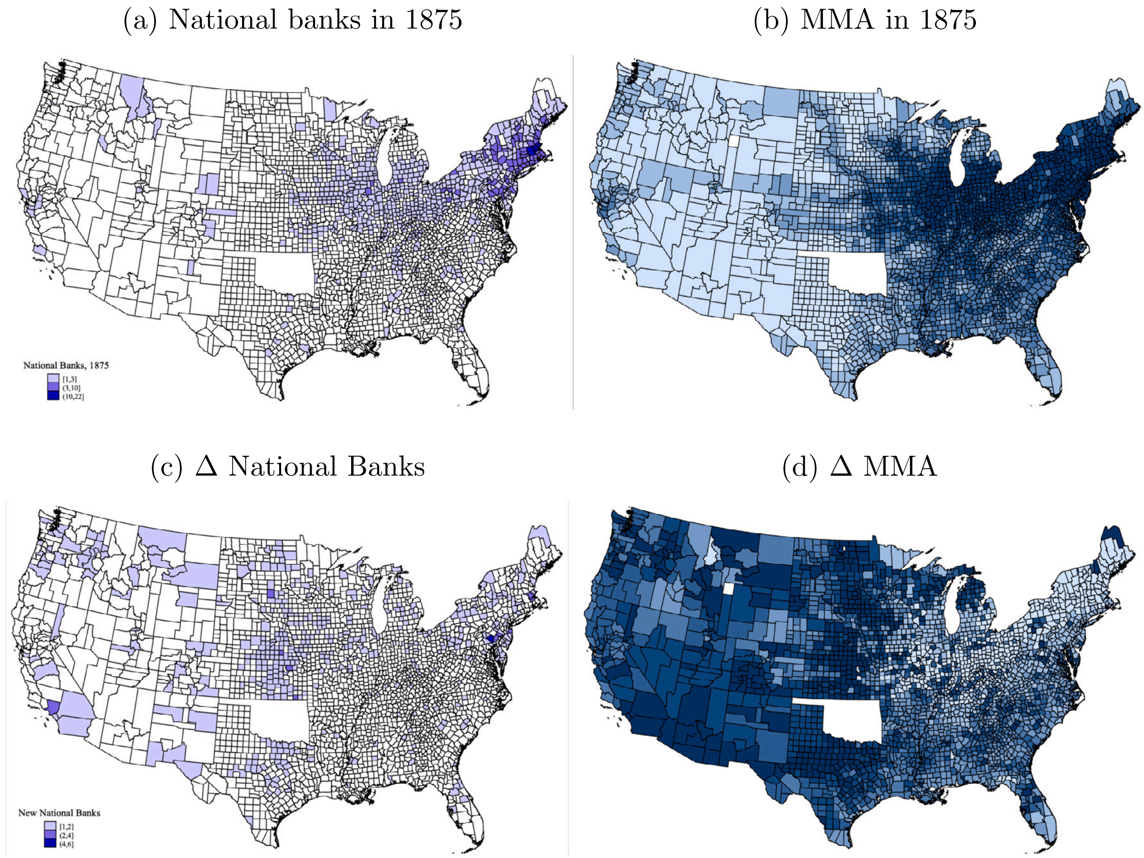

Fig. 2. Geographical distribution of national banks and monetary market access. *Notes:* Fig. 2(a) plots the density of the national banking network at the county level using the *Annual Report of the Comptroller of the Currency* published in 1875. Fig. 2(b) plots our calculated measure of MMA from Eq. (1). Darker shading reflects higher levels of MMA split into seven equally sized bins. Oklahoma was an unincorporated territory that did not have the legal status to allow for national banks and so we do not include it in our data, but distances are calculated allowing for travel through Oklahoma. Fig. 2(c) maps the change in national bank presence from 1875 to 1885. Fig. 2(d) maps the change in MMA between 1875 and 1885 where we calculate MMA in each decade holding the transportation network fixed to its 1870 appearance. Bilateral physical transport costs in 1870 are from Donaldson and Hornbeck (2016). The data for the locations of national banks each year are from the *Annual Reports of the Comptroller of the Currency*.

endogenous component of creating a national bank by using exogenous variation in regulatory capital requirements. Our second approach holds fixed the local national bank network and measures the change in MMA using only distant changes in the network.

5.1. Instrumental variable approach

The direct measure of the change in MMA includes changes to a location's own direct access to national banks, which may be biased by unobserved characteristics that are correlated with future growth.

We approach this endogeneity problem by using the regulatory capital requirement for creating a national bank as an instrument for the change in monetary market access. The capital requirement was based on the size of the town in which the bank was chartered to operate, and it doubled at each population threshold, shown in Appendix Figure A.7. Towns with population less than 6,000 people faced a capital requirement of $50,000 while those between 6,000 and 50,000 faced double the requirement of $100,000.[31]

A priori, the direction of the bias in the OLS is not obvious. For example, assume that larger places have better growth potential and are more likely to be able to generate bank equity. This force makes it more likely that a location joins the national bank network and has higher MMA. However, more growth is also associated with higher population, making it more likely for the town to cross the 6,000 population threshold and to trigger a sharp increase in the required bank equity. As a result, this would reduce the likelihood of creating a bank and directly raising the location's MMA. In this case, the OLS will be downward biased.

In addition, while the concern is that banks enter in areas where they expect to do well, conditional on entering at all, locations choose between establishing a national bank or a state bank. Both types of banks had costs, and the correlation between future growth potential and the relative costs and benefits of each type of bank is not clear. State banks were costly because their bank notes were taxed, but national banks faced higher capital regulations and heavier regulations on their asset allocation. In addition, they were politically unpopular in "silver" areas, which further restricted their adoption.[32] Therefore, while we can expect a positive correlation between the overall *number* of banks and future growth, the correlation between the *composition* of these banks and future growth is a priori unsigned.

Our instrument for national bank entry in 1885 is an indicator variable of a town being below the 6,000 population cut-off in 1880. We focus on the first threshold at the 6,000 population mark because cities near the 50,000 population mark all founded national banks

[31] Locations with more than 50,000 people faced a $200,000 capital requirement.

[32] Specifically, national banks could not take land as collateral, which was the main asset in the agricultural sector. National banks were also part of the postbellum debate on whether to return to a bimetallic standard or to adopt a pure gold standard. "Silver" interests were primarily rural and agricultural while urban areas and financiers (including national banks) advocated for gold.

immediately after the passage of the Act.[33] The validity of the instrument relies on the premise that national bank entry is impacted by the population threshold, but that the threshold was not chosen based on the size of towns that demanded national banks in the 1880s. Given that the legislation was written almost two decades prior to the entry period we study, and that we focus on a small bandwidth around the threshold, we consider it unlikely that policymakers could have so accurately predicted and targeted this demand. In addition, we control for population change in the pre-period to capture potential differences in demand.

In order to implement the instrument, we restrict ourselves to relevant locations by constructing our sample in the following way. First, we select towns with fewer than 6,000 residents in 1870 that did not have a national bank by 1875. These towns all faced the same lower entry cost initially but did not have a national bank and lacked a local supply of liquid bank notes. Second, we require that towns did not grow or shrink beyond a population range of 4,000 to 8,000 in 1880 in order to eliminate outliers that are unlikely to be similar in observable and unobservable ways.

Within this sample, some towns crossed the 6,000 threshold in 1880 and faced double the capital requirement while the rest did not. Remaining below the threshold is our instrument for positive direct changes in MMA. As an example, consider towns A and B, each with 4,000 residents as of the 1870 census. In 1880, town A grew to a population of 5,000, whereas town B grew to 7,000 people. Without the capital requirement, the towns are close enough in size that they are similarly able to sustain a bank. However, the discontinuity in the requirement raises the entry costs significantly for town B relative to A, thereby reducing the probability that it founds a national bank and directly raises its monetary market access.

We use the instrumental approach as opposed to a regression discontinuity design because of the lack of density in the population distribution immediately around the cut-off.[34] This identification strategy follows a number of papers studying bank behavior in the postbellum United States, and in particular those that use the National Banking Act's regulatory capital requirements based on population as a source of exogenous variation (Fulford, 2015; Gou, 2016; Carlson et al., 2022).[35] We combine this identification strategy with the empirical estimations based on changes in MMA.

The identifying assumption for our empirical analysis of real effects is that towns that crossed versus those that did not cross the 6,000 threshold in 1880, despite having the same population change, did not experience a concomitant shock that would have impacted their real outcomes afterward. We consider this unlikely for several reasons. First, there is no evidence of any other regulation at the 6,000 population threshold introduced around this period.[36] Second, focusing on a relatively narrow population bandwidth around the 6,000 cutoff increases the likelihood that towns were comparable in both observable and unobservable ways. Third, we provide evidence that towns in our sample are not observably different during the pre-period.

Fig. 3 shows the coefficients and confidence intervals for both conditional and unconditional covariate balance tests where we regress each characteristic on an indicator variable for being below the 6,000 population threshold in 1880. Each regression is individually estimated, and we normalize all dependent variables so that the magnitudes can be interpreted as standard deviations from the mean. All variables are measured in the decade before national bank entry unless otherwise noted. The "Unconditional" regressions are estimated without controls, and the "Conditional" regressions include state fixed effects and the population change from 1870 to 1880. In terms of population, places that are below the threshold in 1880 are slightly smaller in 1870, unconditionally. After controlling for the population change, they are significantly smaller, as expected.

It is unsurprising that towns that did not cross the population threshold in 1880 had lower population changes. These differences could potentially bias the results if they are correlated with a town's real outcomes through non-bank channels. For example, places that rapidly expanded in the previous decade could continue to grow faster due to agglomeration effects. We therefore control for population changes between 1870 and 1880 to account for a town's overall growth trajectory that could persist into subsequent decades.

First stage. We estimate the following first stage regression:

$$\Delta \text{MMA}_{i,s} = \beta \mathbb{I}(\text{Pop1880} < 6000)_{i,s} + \Gamma' X_{i,s} + \eta_s + \varepsilon_{i,s} \tag{2}$$

for town i in state s. $\Delta\text{MMA}_{i,s}$ is the log change in monetary market access in 1885 relative to 1875, and $\mathbb{I}(\text{Pop1880} < 6000)_{i,s}$ is an indicator variable for having a town population below the 6,000 threshold in 1880 census. $X_{i,s}$ is a vector of control variables such as the population change between 1870 and 1880, the number of railroads in 1875, the number of state banks in 1876, and transportation market access in 1870.[37] η_s denotes state fixed effects, which limits the estimation to using within-state variation and accounts for unobserved state-level differences such as regulatory changes between decades. All specifications cluster standard errors by county.

β is our coefficient of interest, which captures the effect of being below the population threshold in 1880 on the change in MMA between 1875 and 1885. Table 2 presents the first stage relationship as we add control variables. The point estimate in column 6 indicates that being below the population threshold in 1880 is associated with a 62% increase in MMA in 1885 relative to 1875. This positive relationship is robust to including a variety of control variables. Controlling for the population trajectory (comparing column 1 to column 2) improves the precision of the estimate and raises the first stage KP F-stat from 3.3 to 7.4. Comparing the results also shows that the specification that does not include the population control likely suffers from omitted variable bias: places that grow more are likely to have both higher demand for a national bank and to be above the population threshold, which downward biases the effect of the capital constraint. Including the population trajectory improves the comparability of locations in the sample along this observable dimension. We therefore include the change in population in the baseline specification.

We provide several robustness checks for the instrument. First, we show that the instrument also predicts obtaining a national bank directly (appendix Table A.2). Second, we show the effects for a range

[33] Banks in large cities had sufficient capital to immediately convert to a national charter, and so the capital requirements did not bind for them. Larger cities also had less than a quarter of the entry costs per capita compared to those around the 6,000 mark, so the barriers were also lower in real terms. The real magnitudes of the $50,000 difference in required capital at the 6,000 mark was approximately 140 times the average manufacturing wage in 1880.

[34] Figure A.4 shows the distribution of town size for all towns with between 2,000 and 10,000 population in 1880, represented by the uncolored bars. The colored bars represent all towns with fewer than 6,000 people in 1870 that did not have a national bank as of 1875. which is the relevant subsample of towns for our analysis. The black bars represent the towns that we include in our main sample.

[35] Gou (2016) uses the introduction of a new population cutoff in the early 20th century to study the effect of capital requirements on bank stability. Similarly to Fulford (2015) and Carlson et al. (2022), we focus on the 6,000 population cutoff in the 1880 census, and like the latter, we also control for the change in town population following the previous census as a proxy for a town's overall growth trajectory. Carlson et al. (2022) focuses on a different question, as it examines the role of bank competition on lending at a later period with a different sample of towns where there was at least one national bank in the initial period.

[36] State-level regulation changes are absorbed by the fixed effects in the specifications, as are federal regulations that changed over time.

[37] We use transportation market access from Donaldson and Hornbeck (2016): $\text{tMA}_o = \sum_d \phi_{od}^{-\theta} N_d$.

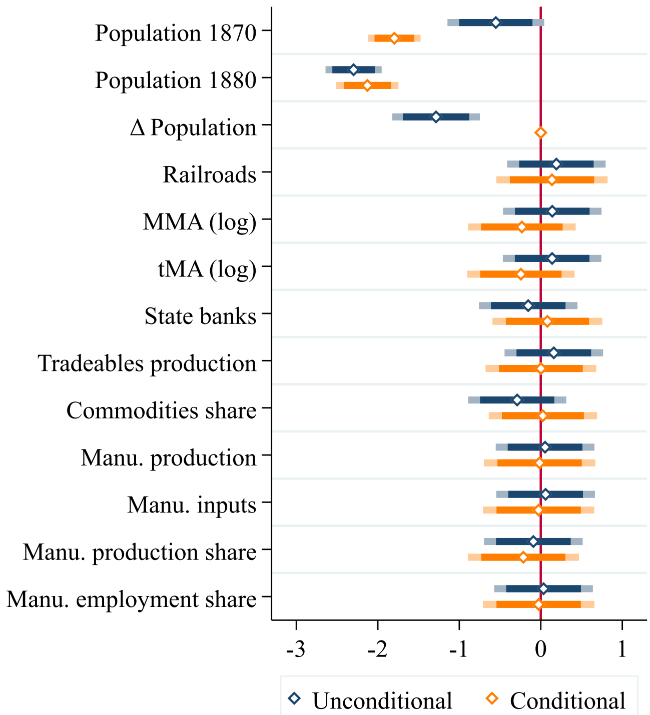

Fig. 3. Covariate balance. *Notes:* Fig. 3 plots the coefficients and 95% (in dark bands) and 99% (in light bands) confidence intervals of individually estimated and normalized regressions of: $X_{i,s} = \beta \mathbb{I}(\text{Pop1880} < 6k) + \eta_s + \epsilon_{i,s}$ where X_i is a location-level characteristic and η_s are state fixed effects. All variables are normalized to have mean zero and standard deviation of one. Population is measured at the town level and all other outcomes are at the county level. The conditional regressions include state fixed effects and the population change from 1870 to 1880 as well. Railroads are from Atack (2016) and measured in 1876. State and other non-national bank financial institutions are obtained from the *Banker's Almanac and Register* of 1876. Monetary market access is calculated in 1875 using the physical transport network in 1870, and transport market access (tMA) is calculated in 1870. Manufacturing and agricultural measures are from their respective censuses of 1870. Standard errors are clustered by county.

Table 2
First stage relationship between population threshold and ΔMMA.

	ΔMMA				
	(1)	(2)	(3)	(4)	(5)
$\mathbb{I}$(pop < 6k)	0.406*	0.715***	0.634***	0.640***	0.619***
	(0.223)	(0.264)	(0.207)	(0.207)	(0.200)
Δ Pop		Y	Y	Y	Y
tMA (1870)			Y	Y	Y
Railroads (1875)				Y	Y
State banks (1876)					Y
State FE	Y	Y	Y	Y	Y
F-stat	3.296	7.353	9.413	9.565	9.583
N	147	147	147	147	147

Notes: Table 2 estimates the first stage relationship: $\Delta\text{MMA} = \beta \mathbb{I}(\text{Pop1880} < 6000)_{i,s} + \Gamma' X_{i,s} + \eta_s + \epsilon_{i,s}$. ΔMMA is the log difference in monetary market access between 1885 and 1875, calculated according to Eq. (1). ΔPop refers to the population change from 1870 to 1880. The log of transport market access cost (tMA) is calculated in 1870, and the number of railroads and state banks are measured in 1875 and 1876, respectively. Regressions are weighted by the share of the town population that is within our relevant sample for the instrument within each county. The F-stat is the first stage KP F. State FEs are included in all specifications, and standard errors are clustered by county. * p < 0.1, ** p < 0.05, *** p < 0.01.

of θ parameters including $\theta = 1$, $\theta = 2$, and $\theta = 8$ (appendix Table A.3). The magnitudes of the estimates range from 0.57 to 0.62, which are very similar to our baseline with $\theta = 4$. Third, we provide the first stage relationship for the change in MMA where we allow the railroad network to change and we calculate MMA in 1885 using the 1880 physical transportation costs, which we denote as $\Delta\text{MMA}_{rail80}$ (appendix Table A.4). This first stage relationship has slightly larger magnitudes and an F-stat of 15.7.

5.2. Network change approach

Our second empirical approach uses the variation in the monetary market access measure that does *not* come directly from the local introduction of a national bank. We denote this measure "MMA_{-O}" to highlight that the origin location's own national bank is not included. We expand the radius to 25, 50, and 100 miles from a location to ensure that this measure does not pick up potential changes in credit supply in the broader area or local productivity shock that might have larger spillovers. We construct our measures of "$\text{MMA}_{-O,r}$" (for each radius r) by holding fixed the national bank network to its 1875 configuration within that radius. Since these measures do not include the endogenous component of a location's own choice to introduce a national bank, we can estimate its relationship to our real outcomes with OLS.

5.3. Specification to estimate real effects

We estimate the impact of entry on real outcomes using difference-in-differences that absorb time trends between periods and location-average differences. Our baseline regressions take the following form:

$$\Delta Y_{i,s} = \beta \Delta\text{MMA}_{i,s} + \Gamma' X_{i,s} + \eta_s + \epsilon_{i,s} \tag{3}$$

where $\Delta Y_{i,s}$ is the change in an outcome in town i in state s in 1890 relative to 1880. $\Delta\text{MMA}_{i,s}$ is the change in monetary market access, which takes the forms in our two approaches: $\widehat{\Delta\text{MMA}}_{i,s}$ is predicted from the instrument, and $\Delta\text{MMA}_{-O,s}$ is based on changes elsewhere in the network. We standardize the units so that all of the results are in terms of a one standard deviation increase in a given measure. β is the main coefficient of interest, which captures the change of the output response to having a national bank. X_{is} is a vector of control variables, and η_s denotes state fixed effects. Time-invariant characteristics of the locations are subsumed by the first differences, and standard errors are clustered by county.

Since national banks are established in towns within counties but our outcomes are at the county level, we weight our regressions with the share of county population within the town. The statistical significance of our results is not sensitive to the weighting, which we show in robustness tables in the appendix. However, we maintain the weighted version as our baseline because they produce more conservative magnitudes.

Our specifications also address the possibility that our results are being driven by changes in physical transport costs. First, our baseline measure of MMA is calculated holding transportation costs fixed to 1870 values so that monetary market access varies due to changes in μ (monetary frictions) across decades rather than changes in ϕ.[38] Second, we include levels of physical trade costs and overall financial development measured in the pre-period as part of our standard set of controls.

5.4. Addressing the credit supply channel

The introduction of safe bank notes occurred through national banks, and so it is important to discuss the possibility that the estimated effects capture a classic credit supply channel as opposed to a transaction friction channel. We address this possibility in three ways, although we fully acknowledge that the nature of the development of

[38] Appendix Table A.4 presents the first stage results for an alternative measure of the change in market access that allows the transport network to change.

the monetary system in the United States precludes us from shutting down this channel completely.

First, in our instrumental approach (Section 5.1), we directly control for the total volume of loans observed on the balance sheets of the national banks in each location. This control variable captures changes in credit that would be reflected in overall volumes but does not capture differences that would appear in other characteristics of the loans, for example its duration or interest rate. We therefore view the robustness of our results to the inclusion of this control as evidence that this specific form of the credit supply channel is not driving our results.

Second, we also directly control for the weighted average change in the total volume of loans in all destinations that a location can trade with. We construct this control in an analogous manner to market access where changes to credit supply in destinations that are physically closer to a location are weighed more than changes in destinations farther away.[39]

For both of these approaches, our first stage estimation also includes these additional control variables. Appendix Table A.5 presents these results where $Loans_o$ and $Loans_d$ are included separately in columns 6–7, and both are included in column 8. The first stage remains highly statistically significant, and the F-stats range from 7 to 10 in all of these specifications.

Third, as explained in Section 5.2, we use the change in the network coming from outside the city and hold fix the city characteristics, such that this source of variation is orthogonal to city level shocks like changes in credit supply and does not rely on the city benefiting from the entry of a new national bank.

6. Results

In this section, we present the results on how access to national banks impacted the local economy, first in overall traded output, and then in the composition of output and the sources of growth.

6.1. Tradables production

The general result in the economic geography literature is that firm production is tied to its distance to the markets in which its output is sold, so we begin by examining the impact of higher monetary market access on real activity in the tradable sector (e.g., Harris, 1954; Redding and Venables, 2004; Donaldson and Hornbeck, 2016). We measure the tradable sector with the total value of agricultural and manufacturing output, which we scale by the population in each county.[40]

Table 3 columns 1 to 3 present our result for both the OLS and the second stage of the IV approach. Column 1 only includes the baseline controls of the pre-period population change and state fixed effects. Column 2 also controls for pre-period financial development (non-national banks), transportation trade costs (railroads), and the level of physical MA in 1870. Column 3 includes the total amount of lending by national banks in each location in 1885 ($Loans_o$) and the aggregate change in lending in all destinations ($Loans_d$). While the magnitudes of the effects are larger after controlling for these two sources of credit, they are within one standard error of the coefficient estimated without these controls and therefore not significantly different from a statistical perspective.[41]

In the bottom panel, our instrumented second stage effects ($\widehat{\Delta \text{MMA}}$) are positive and statistically significant in our preferred specifications with controls. Compared to the pre-period mean values of 775.6, the estimated coefficient of 116.7 implies that a one standard deviation increase in market access raises traded goods production by 15.1%.

The estimated magnitudes range from 115.3 (14.9%) when we set $\theta = 1$ to 125.5 (16.2%) when we set $\theta = 8$ (appendix Table A.6 columns 1 to 5). In the unweighted regression, the estimated coefficient is 202.5 (26.1%) with p-value of 0.04 (appendix Table A.6 column 5). These magnitudes are almost double the baseline magnitudes for the weighted regression since they do not adjust for the share of the county in the treated town. We also estimate the second stage effects when we calculate ΔMMA allowing for physical transport costs to change between decades, and the effects are also statistically significant at the 0.04 level (appendix Table A.6 column 6). In addition, in every specification we provide 95% confidence intervals from the Moreira (2003) conditional likelihood ratio (CLR) test with the p-value of significance in italicized parentheses below. These confidence intervals are constructed from unweighted regressions to also address inference concerns from low first stage F-stats. These confidence intervals imply a p-value of 0.04 in all specifications. In appendix Table A.16 columns 1–2, we include higher order polynomials of the change in population and show that the results have very similar magnitudes and statistical significance. In appendix Table A.17 columns 1–2, we add the direct lending controls at origin and destination individually and show that the magnitudes are similar and statistically significant.

In Table 3 columns 4 and 5, we present the OLS results for the direct relationship between MMA calculated only from changes in national bank entry outside of a given radius (ΔMMA_{-O} and $\Delta \text{MMA}_{-od,25}$, for the county and 25 miles, respectively). We find statistically significant magnitudes of 6.2% and 6.8% percent of the pre-period mean. In appendix Table A.13 columns 1–4, we present the coefficients and standard errors for $\Delta \text{MMA}_{-O,r}$ for the different set of controls and at the other radiuses r. Each coefficient is from an individually estimated regression with the control variables indicated at the bottom of the column. All of these effects are similar in magnitude and statistically significant. At a radius of 100 miles, the estimates are noisier, as would be expected given that there are an increasing number of true changes to the national bank network as the perimeter widens.

Our second set of results presented in Table 3 columns 6 to 10 zooms in on the agricultural sector and measures the change in the intensity of tradables goods production using the share of production in commodities goods in total agricultural production. Commodities are defined as those listed on the Chicago Board of Trade, which were grains including wheat, oats, buckwheat, and corn.[42] We assume that commodities, which have a national market, sell locally at final prices that vary with the national price.[43] As a result, reductions in local trade frictions pass on more fully to producers in the form of higher profits, which incentivize agricultural producers to tilt their production toward these goods.

The magnitude of the instrumented effects with controls in column 8 (0.02) relative to the pre-entry mean of 0.91 indicates that a one standard deviation increase in monetary market access increased the share of traded crop production by approximately 2.3%. As in the case for overall tradables production, the effects are similarly statistically significant for different values of θ, the unweighted regression, and the alternative measure of change in MMA (appendix Table A.7).

In columns 9 and 10, we present the results for our second approach. We find a statistically significant effect of 1.5% relative to the pre-period mean for the average location for a one standard deviation increase in the change in MMA (holding fixed the network in the location's county). These effects are also robust to different combinations of control variables (appendix Table A.17 columns 3 to 5) at all of the different radiuses r (appendix Table A.13 columns 5 to 8).

39 Specifically, we construct the measure the following way (and denote it Loans_d in a slight abuse of notation): $\text{Loans}_d = \sum_d \phi_{od,1870}^{-\theta} \Delta \ln(\text{Loans}_d)$.

40 We use the number of men above the age of 21 as our baseline per capita measure, but the results are robust to other population measures.

41 Appendix Table A.5 reports the first stage relationship with these additional control variables in columns 6–8.

42 The census does not record the values of production in each good, but only the volumes and acres of land used. We calculate the share of acres for these products in total agricultural land.

43 Specifically, we assume that price p for good i trading between locations o and d can be modeled as: $p_{od}^i = \beta p_{national}^i + \gamma_o + \delta_d$ where $\beta > 0$. With these assumptions, it is not necessary for γ and δ to equal 0.

Table 3
Traded output (1880–1890).

	Tradables production					Commodities share				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OLS										
ΔMMA	54.06**	58.10***	81.63***			0.0211***	0.0179***	0.0207***		
	(22.91)	(21.45)	(20.21)			(0.00583)	(0.00538)	(0.00625)		
ΔMMA_{-O}				47.95**					0.0140**	
				(23.49)					(0.00672)	
ΔMMA_{-od25}					52.79**					0.0148**
					(23.99)					(0.00699)
IV: Second stage										
$\Delta\widehat{\text{MMA}}$	67.94	69.94*	116.7**			0.0201***	0.0184***	0.0216**		
	(42.82)	(39.33)	(50.91)			(0.00724)	(0.00684)	(0.00913)		
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls		Y	Y	Y	Y		Y	Y	Y	Y
Loans_o			Y	Y	Y			Y	Y	Y
Loans_d			Y	Y	Y			Y	Y	Y
Mean of Dep. Var.	775.6	775.6	775.6	775.6	775.6	0.905	0.905	0.905	0.905	0.905
N	147	147	147	147	147	147	147	147	147	147

Notes: Table 3 presents the OLS (panel 1) and instrumented second stage of the regression in Eq. (3) (panel 2). The independent variable ΔMMA is the log change in MMA from 1875 to 1885, calculated according to Eq. (1). ΔMMA_{-O} is similarly calculated based on changes outside a location assuming that its own national bank status does not change. ΔMMA_{-od25} is the change outside a location with distance variation assuming that its own national bank status does not change, excluding the changes to national banks in counties within the 25 mile ranges. The inputs to MMA are bilateral transport costs measured in 1870 using the measures from Donaldson and Hornbeck (2016), populations in 1870 and 1880 from the respective censuses of population, and the locations of national banks in 1885 from the *Annual Reports of the Comptroller of the Currency*. There are no national banks in the locations in our sample in 1875 by construction. We use $\theta = 4$ in our baseline measure. All regressions include state fixed effects. The vector of control variables includes the number of railroads and state banks measured in 1875 and 1876, respectively, and the log value of transport market access in 1870. The control variable for "Loans_o" is the log value of lending by national banks in 1885. The control variable "Loans_d" is the average change in log lending by national banks at destinations, with destinations weighted by physical transport proximity. The dependent variable in columns 1 to 5 is the change in per capita (male population over age 21) value of total manufacturing and agricultural production from 1880 to 1890. The dependent variable in columns 6 to 10 is the share of agricultural production in commodity goods measured in terms of acreage of farmland used. We report the mean of the dependent variable measured in 1880. Regressions are weighted by the share of town population within a county, and standard errors are clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.2. *Manufacturing sector growth and structural transformation*

Having shown that national bank entry led to overall growth and more specialization in trade-cost sensitive goods, we next turn to the effect of national banks on manufacturing sector growth. First, there is an increase in overall manufacturing production, which we show in Table 4, columns 1 to 5. Using our IV approach, we find a statistically significant increase of 13.1% manufacturing production per capita from a one standard deviation increase in monetary market access (column 3). Our second approach using changes outside the county (column 4), or a radius of 25 miles (column 5), yields magnitudes of 7.5% and 8.5% for the average location respectively.

Appendix Tables A.8 and A.9 show that these results are robust to different values of θ, to being unweighted, and to the alternative measure of MMA that allows for changes in physical trade costs. As before, the other parameterizations of θ are very similar, and the effects are larger in magnitude in the unweighted regressions. In all specifications, the Conditional Likelihood Ratio (CLR) confidence intervals imply a p-value significance of 0.02 for total production and inputs, while they are not significant for capital. Appendix Tables A.17 and A.14 show that the second approach is robust to different configurations of controls and additional radiuses respectively.

We next directly assess the input-sourcing market access channel where being able to reduce the effective distance to the markets where inputs are sourced also benefits firms (e.g., Redding and Venables, 2004). In particular, manufacturing is dependent on the price, quantity, and quality of inputs, and higher MMA allows local manufacturers to import more and better inputs in addition to exporting more transformed goods (e.g., Goldberg et al., 2010).

In columns 6 to 10 of Table 4 we assess the impact on inputs usage. We find that there is a statistically significant increase in manufacturing inputs per capita of 11.4% using our first approach (column 8) and 6.7%–7.7% using our second approach (columns 9 to 10). In appendix Table A.10, we show that the magnitudes of the effects for manufacturing capital are almost an order of magnitude smaller (1.6%) and not statistically significant.

There are several possible explanations for the smaller effects on capital. First, regulations encouraged national banks to make short-term loans rather than long-term loans (White, 1998). These loans provided working capital to meet short-term liquidity needs rather than long-term investment needs. The short-term credit may have also facilitated inputs sourcing since trade financing mostly provides working capital (e.g., Feenstra et al., 2014; Matray et al., 2023), and the lack of long-term credit provision limited manufacturers' ability to acquire physical capital. Second, since the national banks were unit banks, their ability to diversify their loan portfolio with borrowers across different places was limited, which could discourage them from expanding their balance sheets. Third, large values of firm investment could not be easily accommodated as a national bank could lend no more than 10% of its capital stock to one entity.[44] These numerous requirements could impede long-term investment in the manufacturing sector even while it directly benefited from lower monetary transaction frictions.

Having shown that the manufacturing sector grows overall, and that this growth partly comes from better sourcing of inputs, we next show in Table 5 that the share of manufacturing production and employment in the economy also increases, consistent with a process of structural transformation. In our preferred specification with the full set of controls, a one standard deviation increase in monetary market access leads to a 11.1% increase in the share of manufacturing. Our second approach in columns 4 and 5 yield effects of 8.4% and 9.4% respectively.

The growth in the share of manufacturing output overall is reflected in the increased share of employment in the sector. In Table 5 column 9, we find an 8.4% increase in the manufacturing labor share. As in all other cases, these effects are statistically significant and robust to different values of θ, the unweighted versions of the regressions, and the change in MMA allowing for growth in the physical transport network (Appendix Tables A.11 and A.12). Coefficients are very similar

[44] That is to say, a bank with $50,000 of capital stock could lend no more than $5,000 to a firm. Source: National Banking Act of 64, Sec. 29.

Table 4
Manufacturing sector production (1880–1890).

	Manu production					Manu inputs				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OLS										
ΔMMA	50.97**	56.74***	81.75***			31.59***	32.41***	43.55***		
	(19.66)	(18.63)	(18.99)			(11.47)	(10.78)	(11.87)		
ΔMMA_{-O}				46.87*					25.68*	
				(23.66)					(14.69)	
ΔMMA_{-od25}					53.51**					29.16*
					(23.82)					(14.94)
IV: Second stage										
$\widehat{\Delta\text{MMA}}$	70.74*	72.43**	119.5**			46.44**	47.04***	69.23***		
	(40.68)	(36.35)	(49.77)			(19.46)	(18.04)	(25.44)		
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls		Y	Y	Y	Y		Y	Y	Y	Y
Loans_o			Y	Y	Y			Y	Y	Y
Loans_d			Y	Y	Y			Y	Y	Y
Mean of Dep. Var.	626.4	626.4	626.4	626.4	626.4	380.9	380.9	380.9	380.9	380.9
N	147	147	147	147	147	147	147	147	147	147

Notes: Table 4 presents the OLS (panel 1) and instrumented second stage of the regression in Eq. (3) (panel 2). The independent variable ΔMMA is the log change in MMA from 1875 to 1885, calculated according to Eq. (1). ΔMMA_{-O} is similarly calculated based on changes outside a location assuming that its own national bank status does not change. ΔMMA_{-od25} is the change outside a location with distance variation assuming that its own national bank status does not change, excluding the changes to national banks in counties within the 25 mile ranges. The inputs to MMA are bilateral transport costs measured in 1870 using the measures from Donaldson and Hornbeck (2016), populations in 1870 and 1880 from the respective censuses of population, and the locations of national banks in 1885 from the *Annual Reports of the Comptroller of the Currency*. There are no national banks in the locations in our sample in 1875 by construction. We use $\theta = 4$ in our baseline measure. All regressions include state fixed effects. The vector of control variables includes the number of railroads and state banks measured in 1875 and 1876, respectively, and the log value of transport market access in 1870. The control variable for "$Loans_o$" is the log value of lending by national banks in 1885. The control variable "$Loans_d$" is the average change in log lending by national banks at destinations, with destinations weighted by physical transport proximity. The dependent variables are the change in per capita (male population over age 21) manufacturing production (columns 1 to 5) and manufacturing inputs (columns 6 to 10). We report the mean of the dependent variable measured in 1880. Regressions are weighted by the share of town population within a county, and standard errors are clustered at the county level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 5
Structural transformation (1880–1890).

	Manu production share					Manu employment share				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OLS										
ΔMMA	0.0335**	0.0592***	0.0804***			0.0388***	0.0393***	0.0519***		
	(0.0137)	(0.0128)	(0.0162)			(0.0121)	(0.0131)	(0.0133)		
ΔMMA_{-O}				0.0606***					0.0275**	
				(0.0219)					(0.0136)	
ΔMMA_{-od25}					0.0679***					0.0302**
					(0.0218)					(0.0141)
IV: Second stage										
$\widehat{\Delta\text{MMA}}$	0.0333	0.0485**	0.0691**			0.0621***	0.0658***	0.0992***		
	(0.0234)	(0.0232)	(0.0311)			(0.0211)	(0.0197)	(0.0326)		
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls		Y	Y	Y	Y		Y	Y	Y	Y
Loans_o			Y	Y	Y			Y	Y	Y
Loans_d			Y	Y	Y			Y	Y	Y
Mean of Dep. Var.	0.721	0.721	0.721	0.721	0.721	0.329	0.329	0.329	0.329	0.329
N	147	147	147	147	147	147	147	147	147	147

Notes: Table 5 presents the OLS (panel 1) and instrumented second stage of the regression in Eq. (3) (panel 2). The independent variable ΔMMA is the log change in MMA from 1875 to 1885, calculated according to Eq. (1). ΔMMA_{-O} is similarly calculated based on changes outside a location assuming that its own national bank status does not change. ΔMMA_{-od25} is the change outside a location with distance variation assuming that its own national bank status does not change, excluding the changes to national banks in counties within the 25 mile ranges. The inputs to MMA are bilateral transport costs measured in 1870 using the measures from Donaldson and Hornbeck (2016), populations in 1870 and 1880 from the respective censuses of population, and the locations of national banks in 1885 from the *Annual Reports of the Comptroller of the Currency*. There are no national banks in the locations in our sample in 1875 by construction. We use $\theta = 4$ in our baseline measure. All regressions include state fixed effects. The vector of control variables includes the number of railroads and state banks measured in 1875 and 1876, respectively, and the log value of transport market access in 1870. The control variable for "Loans_o" is the log value of lending by national banks in 1885. The control variable "Loans_d" is the average change in log lending by national banks at destinations, with destinations weighted by physical transport proximity. The dependent variable in columns 1 to 5 is the change in the share of production in the manufacturing sector from 1880 to 1890. The dependent variable in columns 6 to 10 is the change in the share of employment in the manufacturing sector (calculated as manufacturing employment divided by male population over age 21) from 1880 to 1890. We report the mean of the dependent variable measured in 1880. Regressions are weighted by the share of town population within a county, and standard errors are clustered at the county level. * p < 0.1, ** p < 0.05, *** p < 0.01.

across the specifications and are larger in the unweighted specification. Appendix Tables A.15 and A.17 show that these effects are all robust to different configurations of control variables and distance radiuses.

Discussion of the role of direct investment. Our results indicate that a change in market access expands the traded sector overall and the manufacturing sector in particular. While it is ultimately not possible to eliminate the possibility that some of these effects are due to a direct increase in credit supply from national banks, we have several reasons to believe this is unlikely in our case.

First, in all of our specifications, directly controlling for the amount of lending locally does not change our results. Second, we find similar and statistically significant effects when we use our measure of the change in monetary market access that does not use any variation in the changes to national bank presence near a location. Third, within the manufacturing sector, our results are not driven by changes to capital investment. Fourth, the credit supply channel does not generate predictions about the changes in the composition of agricultural production that we find.

6.3. Long-term effects

In our final set of results, we turn to examining the long-run impact of these decadal changes in monetary market access by analyzing the long-run impact on innovation and manufacturing output. The reduction in trade costs can generate innovation through two channels: first, the market access channel increases the number of potential consumers that producers can reach, which increases the incentives to innovate; second, improvements in sourcing inputs can lead to quality upgrading (e.g, Schmookler, 1954; Lileeva and Trefler, 2010; Melitz and Redding, 2021).

The historical innovation data are from Petralia et al. (2016), which records the number of patents obtained by residents within a county each year. Since innovation rates are lumpy, we use the cumulative growth in patenting as our main dependent ratio and winsorize at 5%. Fig. 4(a) plots the cumulative growth (normalized to equal 1 in 1880) relative to the base years (1860–1865) for each group of locations: those with above median changes in their monetary market access (dark squares) and those with below median changes (light crosses). The aggregate pattern is clear: as of 1880, there is no difference in the cumulative growth in innovation, but after 1880 there is a growing divergence such that by 1910, the ratio of accumulated patents is approximately 1.3 times larger in the above-median group.

Fig. 4(b) plots the annual event study difference-in-difference with the network changes in monetary market access as the main independent variable, which exhibits similar patterns: there are no differential pre-trends before 1880, but by 1885 there is a significantly positive gap between the two groups. These effects continue to accumulate before plateauing at a statistically significant average post-period coefficient value of 1.3.

Next, we examine the long-run impact on the manufacturing sector, which during this period was the main producer and beneficiary of innovation. Fig. 4(c) plots the magnitudes and 95% confidence intervals of the estimated effects and shows that the heightened levels of manufacturing remained persistently high as well. The figure also indicates that there were no differential pre-trends between the places that received national banks versus those that did not.[45]

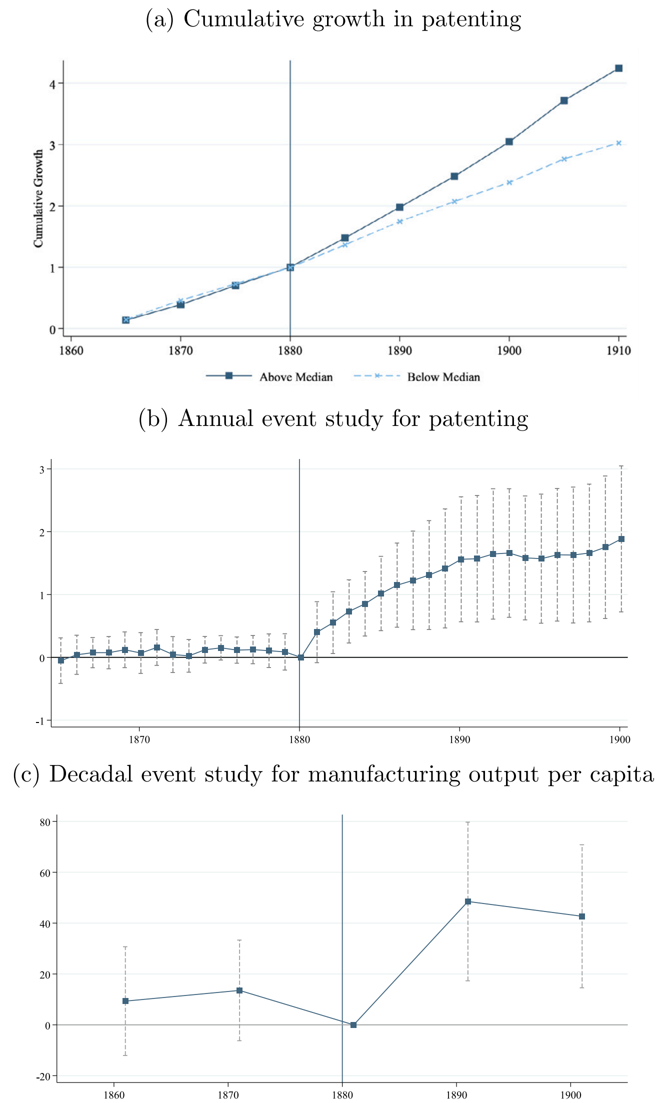

Fig. 4. Persistent positive effect on patenting and manufacturing activity. *Notes:* Fig. 4(a) plots the cumulative growth of patent number (normalized to equal 1 in 1880) relative to the base years (1860–1865) for each group of locations: those with above median changes in their monetary market access (dark squares) and those with below median changes (light crosses). Fig. 4(b) shows the annual diff-in-diff coefficients for the estimated effect of standardized values of ΔMMA_{-O} on cumulative patent growth each year. Fig. 4(c) shows the decadal diff-in-diff coefficients for manufacturing output per capita. We control for location fixed effects, state-by-year fixed effects, and the full vector of control variables interacted with year fixed effects. Regressions are weighted by the share of town population within a county measured in 1880, and standard errors are two-way clustered at the county and year (decade) level. 1880 is the omitted year (decade), and the vertical bars represent the 95% confidence intervals.

7. Conclusion

This paper studies the real effects of introducing a new currency that was safe from redemption risks and costs on different sectors of the economy in the historical context of the late 19th century United States. We use a market access approach to quantify the impact of a less frictional medium of exchange on the aggregate economy. Intuitively, this approach allows us to assess how much the overall "distance" to potential markets (monetary market access) changes as a function of the changes in the national banking network.

We find that increasing a location's monetary market access leads to more production of traded goods, a shift in production towards the most trade-cost sensitive goods, and a larger manufacturing share. These effects, along with higher rates of innovation, persisted for several decades.

Overall, our results indicate that stabilizing the value of assets used as a medium of exchange had real economic benefits, especially for sectors that were likely more exposed to payments frictions. As financial technology progresses and new digital currencies present alternative payment methods, these lessons from the historical national banking era can provide policymakers additional guidance on the costs of transactions frictions arising from the lack of full safety of monetary instruments.

[45] Data from the 1910 Census of Manufacturers is not available, so we stop our analysis before 1920 given the disruptions due to WWI and the establishment of the Federal Reserve System.

CRediT authorship contribution statement

Chenzi Xu: Writing - review & editing, Writing - original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **He Yang:** Writing - review & editing, Writing - original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Real Effects of Supplying Safe Private Money (Original data) (Mendeley Data)

.

Acknowledgments

Philipp Schnabl was the editor for this article, and the authors thank him for valuable guidance. The authors also thank Adrien Matray for numerous discussions. Philip Ager, Rick Hornbeck, and Martin Rotemberg generously shared data. We also thank Anat Admati, Brian Beach, Asaf Bernstein, Michael Bordo, John Campbell, Lauren Cohen, William Diamond, James Feigenbaum, Carola Frydman, Andrew Garin, Michela Giorcelli, Claudia Goldin, Gary Gorton, Walker Hanlon, Arvind Krishnamurthy, Christopher M. Meissner, Nathan Nunn, Siddharth George, Robin Greenwood, Alberto Manconi, Santiago Perez, Sarah Quincy, Tzachi Raz, Hugh Rockoff, Amit Seru, Jann Spiess, Jeremy Stein, Adi Sunderam, as well as seminar and conference participants at Harvard, Yale, UC Davis, Einaudi Institute, HKUST, NBER DAE, EHA, the AEAs, and Stanford SITE for helpful comments. Wenyi Yin, Jiayi Cheng, William Halverson, and Peilin Yang provided excellent research assistance. We are grateful for financial support from The Lab for Economic Applications and Policy (LEAP) at Harvard, the Harvard Department of Economics Summer Research Fellowship, the King Center Junior Faculty Research Grant at Stanford, and National Science Foundation (NSF) Grant 2117003.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfineco.2024.103868.

References

Admati, Anat, Hellwig, Martin, 2014. The Bankers' New Clothes: What's Wrong with Banking and What to Do About It. Princeton University Press.

Ales, Laurence, Carapella, Francesca, Maziero, Pricila, Weber, Warren E., 2008. A model of banknote discounts. J. Econom. Theory 142 (1), 5–27.

Amiti, Mary, Weinstein, David E., 2011. Exports and financial shocks. Q. J. Econ. 126 (4), 1841–1877.

Appleton, Nathan, 1831. An Examination of the Banking System of Massachusetts: In Reference to the Renewal of the Bank Charters, vol. 137, Stimpson and Clapp.

Atack, Jeremy, 2016. Historical geographic information systems (GIS) database of U.S. railroads for 1870 to 1890.

Bell, Spurgeon, 1912. Profit on national bank notes. Am. Econ. Rev. 2 (1), 38–60.

Boehm, Christoph E., Levchenko, Andrei A., Pandalai-Nayar, Nitya, 2020. The Long and Short (Run) of Trade Elasticities. National Bureau of Economic Research.

Bolt, Wilko, Frost, Jon, Shin, Hyun Song, Wierts, Peter, 2023. The Bank of Amsterdam and the limits of fiat money.

Bordo, Michael D., 1998. Currency Crises (And Banking Crises) in Historical Perspective. Stockholm School of Economics Research Report, (10).

Briones, Ignacio, Rockoff, Hugh, 2005. Do economists reach a conclusion on free-banking episodes? Econ. J. Watch 2 (2), 279.

Cagan, Phillip, 1963. The first fifty years of the national banking system. In: Carson, Deane (Ed.), Banking and Monetary Studies. Vol. 15, p. 42.

Calomiris, Charles W., Gorton, Gary, 1991. The origins of banking panics: models, facts, and bank regulation. In: Financial Markets and Financial Crises. University of Chicago Press, pp. 109–174.

Carlson, Mark, Correia, Sergio, Luck, Stephan, 2022. The effects of banking competition on growth and financial stability: Evidence from the national banking era. J. Polit. Econ. 130 (2), 462–520.

Carothers, Neil, 1930. Fractional Money: A History of the Small Coins and Fractional Paper Currency of the United States. J. Wiley & Sons, Incorporated.

Catalini, Christian, de Gortari, Alonso, 2021. On the economic design of stablecoins. Available at SSRN 3899499.

Catalini, Christian, Shah, Nihar, 2021. Setting standards for stablecoin reserves. Available at SSRN.

Coppola, Antonio, Krishnamurthy, Arvind, Xu, Chenzi, 2023. Liquidity, Debt Denomination, and Currency Dominance. National Bureau of Economics Research.

Costinot, Arnaud, Rodríguez-Clare, Andrés, 2014. Trade theory with numbers: Quantifying the consequences of globalization. In: Handbook of International Economics, vol. 4, Elsevier, pp. 197–261.

Dang, Tri Vi, Gorton, Gary, Holmström, Bengt, Ordonez, Guillermo, 2017. Banks as secret keepers. Amer. Econ. Rev. 107 (4), 1005–1029.

Diamond, Douglas W., Dybvig, Philip H., 1983. Bank runs, deposit insurance, and liquidity. J. Polit. Econ. 91 (3), 401–419.

Donaldson, Dave, Hornbeck, Richard, 2016. Railroads and American economic growth: A "market access" approach. Q. J. Econ. 131 (2), 799–858.

Eaton, Jonathan, Kortum, Samuel, 2002. Technology, geography, and trade. Econometrica 70 (5), 1741–1779.

Eichengreen, Barry, 2019. From Commodity to Fiat and Now to Crypto: What Does History Tell Us?. National Bureau of Economic Research.

Feenstra, Robert C., Li, Zhiyuan, Yu, Miaojie, 2014. Exports and credit constraints under incomplete information: Theory and evidence from China. Rev. Econ. Stat. 96 (4), 729–744.

Fonseca, Julia, Matray, Adrien, 2024. Financial inclusion, economic development, and inequality: Evidence from Brazil. J. Financ. Econom. 156.

Friedman, Milton, Schwartz, Anna Jacobson, 1965. A Monetary History of the United States, 1867-1960. Princeton University Press.

Fulford, Scott L., 2015. How important are banks for development? National banks in the united states, 1870–1900. Rev. Econ. Stat. 97 (5), 921–938.

Goldberg, Pinelopi Koujianou, Khandelwal, Amit Kumar, Pavcnik, Nina, Topalova, Petia, 2010. Imported intermediate inputs and domestic product growth: Evidence from India. Q. J. Econ. 125 (4), 1727–1767.

Gorton, Gary, 1999. Pricing free bank notes. J. Monetary Econ. 44, 33–64.

Gorton, Gary, Laarits, Toomas, Muir, Tyler, 2022a. Mobile collateral versus immobile collateral. J. Money Credit Bank..

Gorton, Gary, Pennacchi, George, 1990. Financial intermediaries and liquidity creation. J. Finance 45 (1), 49–71.

Gorton, Gary B., Ross, Chase P., Ross, Sharon Y., 2022b. Making Money. National Bureau of Economic Research.

Gorton, Gary B., Zhang, Jeffery, 2021. Taming wildcat stablecoins. Available at SSRN 3888752.

Gou, Michael, 2016. Did capital requirements promote bank stability in the early 20th century United States?.

Greenfield, Robert L., Rockoff, Hugh, 1990. A tale of two dollars: currency competition and the return to gold, 1865–1879.

Greenfield, Robert L., Rockoff, Hugh, 1995. Gresham's law in nineteenth-century America. J. Money Credit Bank. 27 (4), 1086–1098.

Harris, Chauncy D., 1954. The market as a factor in the localization of industry in the United States. Ann. Assoc. Am. Geogr. 44 (4), 315–348.

Herrendorf, Berthold, Rogerson, Richard, Valentinyi, Akos, 2014. Growth and structural transformation. Handb. Econ. Growth 2, 855–941.

Holmström, Bengt, 2015. Understanding the role of debt in the financial system.

Hornbeck, Richard, 2010. Barbed wire: Property rights and agricultural development. Q. J. Econ. 125 (2), 767–810.

Hornbeck, Richard, Rotemberg, Martin, 2021. Railroads, Market Access, and Aggregate Productivity Growth. University of Chicago Booth School of Business, mimeo.

James, John A., Weiman, David F., 2010. From drafts to checks: the evolution of correspondent banking networks and the formation of the modern US payments system, 1850–1914. J. Money Credit Bank. 42 (2–3), 237–265.

Jaremski, Matthew, 2014. National banking's role in U.S. industrialization, 1850-1900. J. Econ. Hist. 74, 109–140.

Ji, Yan, Teng, Songyuan, Townsend, Robert M., 2023. Dynamic bank expansion: spatial growth, financial access, and inequality. J. Polit. Econ. 131 (8), 2209–2275.

Kinley, David, 1910. The Use of Credit Payments in Payments in the United States, vol. 339, US Government Printing Office.

Kiyotaki, Nobuhiro, Wright, Randall, 1989. On money as a medium of exchange. J. Polit. Econ. 97 (4), 927–954.

Knox, John Jay, 1900. A History of Banking in the United States. BradFord Rhodes and Company.

Landon-Lane, John, Rockoff, Hugh, 2007. The origin and diffusion of shocks to regional interest rates in the United States, 1800-2002. Explor. Econ. Hist. 44 (3), 487–500.

Lileeva, Alla, Trefler, Daniel, 2010. Improved access to foreign markets raises plant-level productivity... for some plants. Q. J. Econ. 125 (3), 1051–1099.

Matray, Adrien, Mueller, Karsten, Xu, Chenzi, Kabir, Poorya, 2023. EXIM's Exit: The Real Effects of Trade Financing by Export Credit Agencies. National Bureau of Economics Research.

Melitz, Marc J., Redding, Stephen J., 2021. Trade and innovation. National bureau of economic research.
Mitchell, Wesley Clair, 1903. A History of the Greenbacks: With Special Reference to the Economic Consequences of Their Issue: 1862-65. University of Chicago Press.
Moreira, Marcelo J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71 (4), 1027–1048.
Paravisini, Daniel, Rappoport, Veronica, Schnabl, Philipp, 2023. Specialization in bank lending: Evidence from exporting firms. J. Finance 78 (4), 2049–2085.
Paravisini, Daniel, Rappoport, Veronica, Schnabl, Philipp, Wolfenzon, Daniel, 2015. Dissecting the effect of credit supply on trade: Evidence from matched credit-export data. Rev. Econ. Stud. 82 (1), 333–359.
Petralia, Sergio, Balland, Pierre-Alexandre, Rigby, David, 2016. Unveiling the geography of historical patents in the United States from 1836 to 1975. Sci. Data.
Preston, Howard H., 1920. The federal reserve banks' system of par collections. J. Polit. Econ. 28 (7), 565–590.
Redding, Stephen, Venables, Anthony J., 2004. Economic geography and international inequality. J. Int. Econ. 62 (1), 53–82.
Rockoff, Hugh, 1975. The Free Banking Era: A Reconsideration. Arno Press, New York.
Rolnick, Arthur J., Weber, Warren E., 1982. Free banking, wildcat banking, and shinplasters. Fed. Reserve Bank Minneapolis Q. Rev..
Romer, Christina D., Romer, David H., 1989. Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. NBER Macroecon. Annu. 4, 121–170.
Rousseau, Peter L., Sylla, Richard, 2005. Emerging financial markets and early US growth. Explor. Econ. Hist. 42 (1), 1–26.
Rousseau, Peter L., Wachtel, Paul, 1998. Financial intermediation and economic performance: Historical evidence from five industrialized countries. J. Money Credit Bank. 30 (4), 657–678.
Sarkisyan, Sergey, 2023. Instant Payment Systems and Competition for Deposits. Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.
Schmookler, Jacob, 1954. The level of inventive activity. Rev. Econ. Stat. 183–190.
Stein, Jeremy C., 2012. Monetary policy as financial stability regulation. Q. J. Econ. 127 (1), 57–95.
Sylla, Richard, 1969. Federal policy, banking market structure, and capital mobilization in the United States, 1863–1913. J. Econ. Hist. 29 (4), 657–686.
Sylla, Richard, 1982. Monetary innovation in America. J. Econ. Hist. 42 (1), 21–30.
Ware, Lynn Willoughby, 1990. Cotton money: Antebellum currency conditions in the Apalachicola/Chattahoochee River Valley. Georgia Hist. Q. 74 (2), 215–233.
Warner, John DeWitt, 1895. The currency famine of 1893. Sound Curr. 2 (6), 337–356.
White, Eugene N., 1982. The membership problem of the National Banking System. Explor. Econ. Hist. 19 (2), 110–127.
White, Eugene N., 1998. Were banks special intermediaries in late nineteenth century America. Fed. Reserve Bank St. Louis Rev..
Xu, Chenzi, 2022. Reshaping global trade: The immediate and long-run effects of bank failures. Q. J. Econ. 137 (4), 2107–2161.